

Dissertation zur Erlangung des Doktorgrades  
der Fakultät für Chemie und Pharmazie  
der Ludwig-Maximilians-Universität München

Beschreibung dynamischer molekularer Systeme  
mit effizienten linear-skalierenden  
QM- und QM/MM-Methoden

Denis Benjamin Flaig

aus

Stuttgart

2012

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. C. Ochsenfeld betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, 24.09.2012

---

(Denis Flaig)

Dissertation eingereicht am 24.09.2012

1. Gutachter: Prof. Dr. C. Ochsenfeld
2. Gutachter: Prof. Dr. H. Ebert

Mündliche Prüfung am 13.11.2012

## **Danksagung**

Herrn Prof. Ochsenfeld danke ich herzlich für die Möglichkeit, in seiner Arbeitsgruppe die Dissertation anfertigen zu können. Vielen Dank für die Bereitstellung des Forschungsthemas, die freundliche Arbeitsatmosphäre und beständige Unterstützung.

Prof. Ebert danke ich herzlich für die Anfertigung des Zweitgutachtens.

Außerdem bedanke ich mich bei den Kolleginnen und Kollegen in der Arbeitsgruppe für die nette, aufbauende Zusammenarbeit in den letzten Jahren und insbesondere den vielfältigen Hilfen von allen Seiten. Danke, Dr. M. Beer, Dr. T. Berezniak, I. Blank, Dr. L. Clin, Dr. B. Doser, Dr. M. Hanni, T. Kessel, J. Kluge, Dr. J. Kussmann, Dr. D. Lambrecht, Dr. M. Löffler, A. Lünser, Dr. B. Maryasin, M. Maurer, S. Maurer, S. Roßbach, Dr. K. Sadeghian, J. Schäffer, Dr. B. Schmitt, Dr. S. Schweizer, H. Senft, P. Snajberk, Dr. C. Sumowski und Dr. J. Zienau.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>7</b>
<b>2</b>	<b>Theoretische Grundlagen</b>	<b>11</b>
2.1	Schrödinger-Gleichung und Kern-Elektron-Separation . . . . .	11
2.2	Molekulardynamische Methoden . . . . .	14
2.3	Quantenchemische Elektronenstruktur-Methoden . . . . .	16
2.4	Berechnung statischer Moleküleigenschaften . . . . .	24
2.5	Molekularmechanische Methoden . . . . .	30
2.6	QM/MM-Methoden . . . . .	32
<b>3</b>	<b>Effiziente linear-skalierende SCF-Verfahren</b>	<b>37</b>
3.1	Einführung . . . . .	37
3.2	Linear-skalierendes quasi-indirektes bzw. semidirektes SCF-Verfahren . . . . .	38
3.3	Abstands-abhängiges Schwarz-Screening . . . . .	51
3.4	Zusammenfassung und Ausblick . . . . .	53
<b>4</b>	<b>QM-Ansätze zur effizienten Beschreibung der Molekulardynamik</b>	<b>55</b>
4.1	Einführung . . . . .	55
4.2	Extrapolationsmethode für CPSCF-Anfangsschätzungen . . . . .	57
4.3	Integralwiederverwendung für veränderte Konformationen . . . . .	62
4.4	QM/MM-Implementierung für Molekulardynamik . . . . .	70
4.5	Zusammenfassung und Ausblick . . . . .	71
<b>5</b>	<b>QM- und QM/MM-Berechnung von NMR-Verschiebungen</b>	<b>73</b>
5.1	Einführung . . . . .	73
5.2	Genauigkeiten von NMR-Verschiebungen auf GIAO-SCF- und MP2-Niveau . . . . .	75

5.3	Intermediäre Referenzen bei der Berechnung von NMR-Verschiebungen . . . . .	82
5.4	Konvergenz von QM/MM-Abschirmungen mit der Größe der QM-Region . . . . .	88
5.5	Zusammenfassung und Ausblick . . . . .	105
<b>6</b>	<b>Enzymatische Reparatur oxidativer DNA-Schäden</b>	<b>109</b>
6.1	Einführung . . . . .	109
6.2	Einzelheiten zur Strukturauswahl und -vorbereitung . . . . .	111
6.3	Gesamte Schadensstabilisierung . . . . .	112
6.4	Individuelle Beiträge der Enzymresiduen zur Schadensstabilisierung . . . . .	116
6.5	Zusammenfassung und Ausblick . . . . .	121
<b>7</b>	<b>Allgemeine Zusammenfassung</b>	<b>123</b>
	<b>Anhang A Q-Chem-ChemShell-Interface</b>	<b>127</b>
	<b>Anhang B Molekularer Testsatz</b>	<b>129</b>
	<b>Anhang C Details zur QM-Größenkonvergenz von NMR-Abschirmungen</b>	<b>131</b>
	<b>Anhang D Details zur Untersuchung der DNA-Reparatur</b>	<b>143</b>
	<b>Tabellen- und Abbildungsverzeichnis</b>	<b>145</b>
	<b>Literaturverzeichnis</b>	<b>151</b>
	<b>Veröffentlichungen und Lebenslauf</b>	<b>163</b>

# Kapitel 1

## Einleitung

Die Theoretische Chemie kann heute auf eine Vielzahl an Methoden zurückgreifen, um thermodynamische und kinetische Eigenschaften oder die Dynamik molekularer Systeme zu beschreiben. Die Bandbreite theoretischer Methoden erstreckt sich beispielsweise von einfachen molekularmechanischen Verfahren über semi-empirische bis hin zu quantenchemischen Methoden (für eine Übersicht siehe z.B. Referenz [1]). Die verschiedenen methodischen Ansätze gründen dabei auf sehr unterschiedlich genauen physikalischen Beschreibungen der Materie. Mit steigender Genauigkeit der zugrunde liegenden Beschreibung erreichen die Methoden typischerweise eine höhere Vorhersagekraft, allerdings zumeist auf Kosten eines maßgeblich höheren Rechenaufwands. Um molekulare Systeme der aktuellen chemischen Forschung sehr genau theoretisch zu beschreiben, wird der Rechenaufwand trotz der erheblichen methodischen und technologischen Fortentwicklungen der letzten Jahrzehnte schnell zum eigentlich beschränkenden Faktor. Mit den immens gestiegenen Möglichkeiten hat sich nämlich gleichzeitig auch der Anspruch an die Theoretische Chemie massiv erweitert: Galt die Aufmerksamkeit früher oft besonders einzelnen und eher kleinen Molekülen, hat sich der theoretische Blickpunkt spätestens mit den Fortschritten der supramolekularen Chemie in den 80er und 90er Jahren darauf verlagert, weitaus größere molekulare Einheiten zu untersuchen. Heute strebt die Theorie an, sowohl synthetische supra- und makromolekulare als auch große biomolekulare Systeme zuverlässig zu beschreiben. Dabei spielt zugleich eine große Rolle, weitreichende Umgebungseffekte (z.B. Lösungsmittelleffekte) möglichst genau zu berücksichtigen.

Eine entscheidende Kenngröße der theoretischen Methoden ist in diesem Hinblick das Skalenverhalten des Rechenaufwands mit der Systemgröße  $M$ . So steigt beispielsweise für konventionelle quantenchemische Methoden der Rechenaufwand typischerweise nicht linear mit der

Molekülgröße an, sondern mit höheren Potenzen von  $M$ . Um das Skalenverhalten wesentlich zu reduzieren, konnten allerdings in den letzten Jahrzehnten wichtige methodische Fortschritte erzielt werden. Die methodischen Fortentwicklungen senken beispielsweise das Skalenverhalten für viele quantenchemische Rechnungen auf Basis der Hartree-Fock- oder Dichtefunktional-Theorie mit zunehmender Systemgröße auf linear und auch für Post-Hartree-Fock-Theorien konnten inzwischen linear-skalierende Methoden entwickelt werden (für einen Überblick siehe z.B. Referenzen [2, 3]).

Neben dem Skalenverhalten ist für den Rechenaufwand der theoretischen Methoden zudem sehr bedeutsam, wie groß der gegenüber dem Anstieg mit der Systemgröße konstante Vorfaktor der Rechenzeit ist. So begründet beispielsweise der kleinere Vorfaktor der molekularmechanischen Methoden deren typischen Geschwindigkeitsvorteil auch gegenüber linear-skalierenden quantenchemischen Methoden— allerdings um den Preis einer Beschreibung auf niedrigerem theoretischen Niveau und daher einer geringeren Anwendbarkeit und Aussagekraft. Um die Vorzüge der linear-skalierenden quantenchemischen Methoden optimal ausnutzen zu können, besteht ein wichtiges Ziel darin, den Vorfaktor für linear-skalierende QM-Methoden ohne signifikante Genauigkeitseinbußen weiter zu senken.

Diesem Ziel entsprechend beschreibt ein wesentlicher Teil der Dissertation Weiterentwicklungen, die zu einer Reduktion des Vorfaktors für linear-skalierende quantenchemische Methoden auf Basis der Hartree-Fock- oder Kohn-Sham-Dichtefunktional-Theorie führen. Das Kapitel 3 geht dabei auf methodische Fortentwicklungen zur Reduktion des Vorfaktors für eine gegebene Molekülgeometrie ein, während das Kapitel 4 ergänzende Ansätze für Rechnungen mehrerer Molekülgeometrien im Rahmen der Born-Oppenheimer-Molekulardynamik umfasst.

So stellt das Kapitel 3 ein linear-skalierendes Integralspeicher-Verfahren dar, das die Vorteile der traditionell indirekten oder semidirekten Ansätze mit denen direkter SCF-Methoden kombiniert. Darüber hinaus geht das Kapitel auf die wesentlichen Beiträge für ein neues Abstands-abhängiges Schwarz-Screening ein, das zu einer weiteren Effizienzsteigerung von SCF-Methoden führt.

Das Kapitel 4 beschreibt zwei methodische Ansätze, um bei dynamischen molekularen Systemen die Born-Oppenheimer-Hyperfläche effizient abzutasten: einen Extrapolationsansatz für CPSCF-Anfangsschätzungen und einen Ansatz zur numerisch kontrollierten Wiederverwen-

---

dung von Zweielektronen-Integralen aus vergangenen Zeitschritten bei der Born-Oppenheimer-Molekulardynamik. Außerdem werden die Eckpunkte einer neu implementierten Programmschnittstelle für gemischte QM/MM-Energie- und Gradientenrechnungen dargestellt.

Das Kapitel 5 widmet sich anschließend der Kombination und Untersuchung bestehender Methoden zur genauen Berechnung kernmagnetischer Verschiebungen. Ein besonderer Schwerpunkt liegt hierbei darauf, QM/MM-Ansätze [4] mit linear-skalierender NMR-Methoden [5–7] zu kombinieren und zu untersuchen, wie groß die QM-Regionen gewählt werden müssen, um NMR-Verschiebungen von zentralen Molekülregionen zuverlässig zu berechnen. Außerdem umfasst das Kapitel eine breite Untersuchung der Methodengenauigkeit von GIAO-SCF-, und -MP2-Methoden für NMR-Verschiebungsrechnungen. Zudem wird ein weiterführender Ansatz zur Steigerung der Methodengenauigkeiten am Beispiel von Polypeptiden untersucht.

Im letzten Teil der Arbeit (Kapitel 6) werden schließlich Ergebnisse einer quantenchemischen Untersuchung zur enzymatischen Reparatur oxidativer DNA-Schäden dargestellt. Die Untersuchung quantifiziert die unterschiedliche Stabilisierung geschädigter DNA im Vergleich zu intakter DNA durch das Reparaturenzym und identifiziert wichtige Proteinresiduen bei der Schadensstabilisierung. Ein besonderes Augenmerk der Untersuchung liegt ebenfalls auf Aspekten der theoretischen Beschreibung, wie Basissatz- und Methodenfehler sowie der Konvergenz der Ergebnisse mit der Größe der QM-Region.



## Kapitel 2

### Theoretische Grundlagen

Das folgende Kapitel umfasst die theoretischen Grundlagen der quantenmechanischen (QM) und molekularmechanischen (MM) Methoden, die im Rahmen der Dissertation angewendet, untersucht und weiterentwickelt werden. Die Darstellung beginnt ausgehend von den Postulaten der (nicht-relativistischen) Quantenmechanik und entwickelt von diesem Standpunkt aus die grundlegende Theorie der

- molekulardynamischen Methoden (Abschnitt 2.2)
- quantenchemischen Elektronenstruktur-Methoden (Abschnitt 2.3)
- Berechnung statischer Moleküleigenschaften (Abschnitt 2.4)
- molekularmechanischen Mechanik (Abschnitt 2.5)
- kombinierten quanten-/molekularmechanischen (QM/MM) Methoden (Abschnitt 2.6)

#### 2.1 Schrödinger-Gleichung und Kern-Elektron-Separation

Die Quantenmechanik postuliert, dass ein molekulares System vollständig durch eine orts- und zeitabhängige Wellenfunktion  $\Phi(\mathbf{R}, \boldsymbol{\tau}, t)$  beschrieben werden kann. Die Wellenfunktion erfüllt (im nicht-relativistischen Fall) die für die Quantenchemie zentrale Gleichung,

$$\hat{H}\Phi(\mathbf{R}, \boldsymbol{\tau}, t) = i\hbar \frac{\partial}{\partial t} \Phi(\mathbf{R}, \boldsymbol{\tau}, t), \quad (2.1)$$

die zeitabhängige Schrödinger-Gleichung, mit der Zeit  $t$ , dem Satz der Koordinaten der Atomkerne  $\mathbf{R} = \{\mathbf{R}_1, \mathbf{R}_2 \cdots \mathbf{R}_{N_K}\}$  und Elektronen  $\boldsymbol{\tau} = \{\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \cdots \boldsymbol{\tau}_{N_e}\}$ . Der Hamilton-Operator

$\hat{H}$  bestimmt die Form der Schrödinger-Gleichung und ist für ein abgeschlossenes molekulares System gegeben durch

$$\hat{H} = \overbrace{-\frac{1}{2} \sum_i \nabla_i^2 - \sum_{i,A} \frac{Z_A}{r_{iA}} + \sum_{i,j>i} \frac{1}{r_{ij}} + \underbrace{\sum_{A,B>A} \frac{Z_A Z_B}{R_{AB}}}_{V_{\text{NN}}}}^{\hat{H}_e} \overbrace{-\frac{1}{2} \sum_A \frac{1}{M_A} \nabla_A^2}_{\hat{T}_N} \quad (2.2)$$

in atomaren Einheiten mit den Kernladungen  $Z_A$ , Kernmassen  $M_A$  sowie den Elektronenindizes  $i, j$  und Kernindizes  $A, B$ . Lösungen der zeitabhängigen Schrödinger-Gleichung basieren im Allgemeinen auf Separationsansätzen für die Wellenfunktion. Dabei werden im Wesentlichen zwei sich ergänzende Ansätze verfolgt, die sich im Weg der Separation unterscheiden und in der Literatur als Ehrenfest- bzw. Born-Oppenheimer-Separationsweg besprochen werden (siehe Abbildung 2.1 und Referenz [8]).

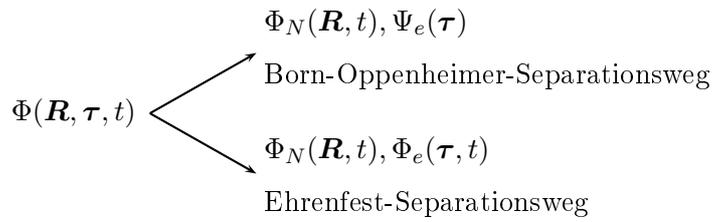


Abb. 2.1: Schematische Darstellung der Ehrenfest- bzw. Born-Oppenheimer-Wege zur Ort-Zeit- und Kern-Elektron-Separation.

Im Folgenden wird der Born-Oppenheimer-Separationsweg als Grundlage der Born-Oppenheimer-Moleküldynamik und den quantenchemischen Elektronenstruktur-Methoden ausführlicher dargestellt. Der Hamilton-Operator hängt für ein abgeschlossenes molekulares System nicht explizit von der Zeit  $t$  ab, so dass sich mit dem Separationsansatz

$$\Phi(\mathbf{R}, \boldsymbol{\tau}, t) = \Psi(\mathbf{R}, \boldsymbol{\tau}) \cdot f(t) \quad (2.3)$$

die zeitabhängige Schrödinger-Gleichung in eine nur zeitabhängige Gleichung

$$i\hbar \frac{\partial}{\partial t} f(t) = \mathcal{E} f(t) \quad (2.4)$$

und nur ortsabhängige Gleichung separieren lässt, die zeitunabhängige Schrödinger-Gleichung,

$$\hat{H}\Psi(\mathbf{R}, \boldsymbol{\tau}) = \mathcal{E}\Psi(\mathbf{R}, \boldsymbol{\tau}), \quad (2.5)$$

mit dem stationären Energieeigenwert  $\mathcal{E}$ . Verwendet man den Born'schen Separationsansatz [9, 10]

$$\Psi(\mathbf{R}, \tau) = \sum_n \Psi_{N,n}(\mathbf{R}) \cdot \Psi_{e,n}(\mathbf{R}, \tau) \quad (2.6)$$

über verschiedene Zustände  $n$ , ergibt sich für die zeitabhängige Gesamtwellenfunktion,

$$\Phi(\mathbf{R}, \tau, t) = f(t) \underbrace{\sum_n \Psi_{N,n}(\mathbf{R}) \cdot \Psi_{e,n}(\mathbf{R}, \tau)}_{\sum_n \Phi_{N,n}(\mathbf{R}, t)}, \quad (2.7)$$

Dabei zeigen sich die zeitabhängigen Kernwellenfunktionen  $\Phi_{N,n}(\mathbf{R}, t)$  als zeitabhängige Entwicklungskoeffizienten der zeitunabhängigen elektronischen Wellenfunktion  $\Psi_{e,n}(\mathbf{R}, \tau)$ . Indem man den Separationsansatz in die zeitabhängige Schrödinger-Gleichung einsetzt, mit einer zu  $\Psi_{e,n}(\mathbf{R}, \tau)$  orthonormalen elektronischen Wellenfunktion  $\Psi_{e,m}(\mathbf{R}, \tau)$  multipliziert und über die elektronischen Koordinaten integriert, erhält man die nicht-adiabatische Kern-Schrödinger-Gleichung:

$$\left[ \hat{T}_N + E_m(\mathbf{R}) \right] \Phi_{N,m}(\mathbf{R}, t) + \sum_n \hat{T}_{mn}(\mathbf{R}, \tau) \Phi_{N,n}(\mathbf{R}, t) = i\hbar \frac{\partial}{\partial t} \Phi_{N,m}(\mathbf{R}, t). \quad (2.8)$$

Darin sind die Terme  $E_m(\mathbf{R})$  die Energieeigenwerte des elektronischen Hamilton-Operators  $\hat{H}_e$ , d.h. Lösungen der elektronischen Schrödinger-Gleichung

$$\hat{H}_e \Psi_{e,m}(\mathbf{R}, \tau) = E_m(\mathbf{R}) \Psi_{e,m}(\mathbf{R}, \tau). \quad (2.9)$$

Die nicht-adiabatischen Kopplungsterme  $\hat{T}_{mn}(\mathbf{R}, \tau)$  hängen explizit von den Kern- und Elektronenkoordinaten  $\mathbf{R}$  und  $\tau$  ab (siehe z.B. Referenz [8]). Um eine Entkopplung der Kern- und Elektronenkoordinaten zu erreichen, werden bei der adiabatischen Näherung die Außerdiagonalelemente von  $\hat{T}_{mn}(\mathbf{R}, \tau)$  vernachlässigt:

$$\left[ \hat{T}_N + E_m(\mathbf{R}) \right] \Phi_{N,m}(\mathbf{R}, t) + \hat{T}_{mm}(\mathbf{R}) \Phi_{N,m}(\mathbf{R}, t) = i\hbar \frac{\partial}{\partial t} \Phi_{N,m}(\mathbf{R}, t) \quad (2.10)$$

Die adiabatische Näherung impliziert, dass sich während der Bewegung der Kerne der elektronische Zustand nicht ändert und ist gerechtfertigt, solange die Energiedifferenzen der elektronischen Zustände für alle relevanten Molekülkonformationen  $\mathbf{R}$  groß sind und keine elektronischen Anregungen während der Kernbewegung beschrieben werden sollen (für Methoden zur Beschreibung gekoppelter Kern- und Elektronendynamik siehe z.B. Referenzen [8, 11, 12]).

Bei der Born-Oppenheimer Näherung, die auf dem sehr großen Verhältnis von Kern- und Elektronenmasse basiert, vernachlässigt man zudem die Diagonalelemente  $\hat{T}_{mm}(\mathbf{R})$  [13]:

$$\left[\hat{T}_N + E_m(\mathbf{R})\right] \Phi_{N,m}(\mathbf{R}, t) = i\hbar \frac{\partial}{\partial t} \Phi_{N,m}(\mathbf{R}, t) \quad (2.11)$$

In der verbliebenen Kern-Schrödinger-Gleichung bilden die Eigenwerte  $E_m(\mathbf{R})$  des elektronischen Hamilton-Operators also den potentiellen Beitrag zur Beschreibung der Kernbewegung.

## 2.2 Molekulardynamische Methoden

Die Tabelle 2.1 gibt eine Übersicht über verschiedene grundlegende molekulardynamische Methoden, die auf unterschiedlichen theoretischen Niveaus aufbauen (für weiterführende MD-Techniken siehe z.B. Referenz [8]). Als Grundlage der Dissertation wird im Folgenden ausgehend vom vorangehenden Abschnitt die grundlegende Theorie der Born-Oppenheimer-Molekulardynamik (BOMD) dargestellt. Außerdem wird der Ansatz der klassischen MD erläutert. Die zugrunde liegenden quantenchemischen Elektronenstrukturmethoden bzw. molekularmechanischen Methoden werden in den darauf folgenden Abschnitten beschrieben.

Tab. 2.1: Bezeichnung und Übersicht über verschiedene molekulardynamische Methoden.

Bezeichnung	Behandlung der Atomkerne	Beschreibung
Quanten-MD	quantenmech.	Quantenmech. Rechnung jenseits der klass. Behandlung der Kernbewegung
Ab-initio-MD	klassisch	Ab-initio-Elektronenstruktur-Rechnung während der MD
a) Ehrenfest-MD		basiert auf Ehrenfest-Separation
b) Born-Oppenheimer-MD		basiert auf Born-Oppenheimer-Separation
c) Car-Parinello-MD		Propagation der elektr. Wellenfunktion
klassische MD	klassisch	vorparametrisierte Kraftfelder

Bei der Born-Oppenheimer-Molekulardynamik (BOMD) geht man von der quantenmechanischen Beschreibung der Kerndynamik (siehe Gleichung 2.11) über zu einer klassischen Behandlung, die sich in der Newton'schen Darstellung formulieren lässt durch das System von

$N_K$  Gleichungen

$$M_A \frac{\partial^2}{\partial t^2} \mathbf{R}_A = \underbrace{-\nabla_A E_0(\mathbf{R})}_{\mathbf{F}_A(\mathbf{R})}, \quad (2.12)$$

wobei  $E_0(\mathbf{R})$  der Eigenwert des elektronischen Hamiltonoperators für den Grundzustand ist, also

$$E_0(\mathbf{R}) = \langle \Psi_{e,0}(\mathbf{R}, \boldsymbol{\tau}) | \hat{H}_e | \Psi_{e,0}(\mathbf{R}, \boldsymbol{\tau}) \rangle, \quad (2.13)$$

für die normierte elektronische Wellenfunktion  $\Psi_{e,0}(\mathbf{R}, \boldsymbol{\tau})$ . Der große Quotient aus Kern- und Elektronenmasse bildet die Grundlage dafür, die Kerne klassisch zu beschreiben. Im Zuge (*on the fly*) der MD werden der Eigenwert des elektronischen Hamiltonoperators  $E_0(\mathbf{R})$  und die Kräfte  $\mathbf{F}_A(\mathbf{R})$  an den Kernpositionen für jede Molekülkonformation  $\mathbf{R}$  mittels einer quantenchemischen Elektronenstruktur-Rechnung neu bestimmt (siehe Abschnitte 2.3 und 2.4). Die Gleichung 2.12 bildet zugleich die Grundlage zu den einfacheren klassischen MD-Methoden, in denen die Kräfte  $\mathbf{F}_A(\mathbf{R})$  nicht aus quantenchemischen Elektronenstruktur-Rechnungen, sondern aus einfachen molekularmechanischen Energieausdrücken berechnet werden (siehe Abschnitt 2.5). Eine weitere Möglichkeit besteht darin, die Kräfte  $\mathbf{F}_A(\mathbf{R})$  aus QM/MM-Methoden zu berechnen (siehe Abschnitt 2.6).

Die klassischen Kernbewegungs-Gleichungen werden im Allgemeinen durch numerische Integrationsverfahren gelöst, wie die Gleichung 2.14 am Beispiel eines Verlet-Integrators [14] zeigt:

$$\begin{aligned} \mathbf{R}_A(t_0 + h) &\approx \mathbf{R}_A(t_0) + h \left. \frac{\partial}{\partial t} \mathbf{R}_A \right|_{t_0} + \frac{1}{2} h^2 \frac{\mathbf{F}_A(t_0)}{M_A} \\ \mathbf{R}_A(t_0 - h) &\approx \mathbf{R}_A(t_0) - h \left. \frac{\partial}{\partial t} \mathbf{R}_A \right|_{t_0} + \frac{1}{2} h^2 \frac{\mathbf{F}_A(t_0)}{M_A} \\ \mathbf{R}_A(t_0 + h) &\approx 2\mathbf{R}_A(t_0) - \mathbf{R}_A(t_0 - h) + h^2 \frac{\mathbf{F}_A(t_0)}{M_A} \end{aligned} \quad (2.14)$$

Der Zeitschritt  $h$  muss bei den numerischen Integrationsverfahren möglichst klein gewählt werden, um den numerischen Fehler bei der Integration der Kernbewegungs-Gleichungen möglichst klein zu halten. Gleichzeitig wächst natürlich mit kleinem  $h$  die Anzahl benötigter Schritte um eine bestimmte Simulationsdauer zu erreichen. Üblicherweise bemisst man  $h$  an typischen Zeiten der schnellsten Molekülschwingung im System (im Bereich von  $1 \text{ fs} \approx 41.3 \text{ a.u.}$  [15, 16]). Für das Beispiel eines Verlet-Integrators [14] lässt sich für  $t_0 = 0$  die benötigte Molekülkonformation  $\mathbf{R}_A(-h)$  z.B. bestimmen nach

$$\mathbf{R}_A(-h) = \mathbf{R}_A(0) - h \underbrace{\left. \frac{\partial}{\partial t} \mathbf{R}_A \right|_0}_{\mathbf{v}_A(0)} + \frac{1}{2} h^2 \frac{\mathbf{F}_A(0)}{M_A}. \quad (2.15)$$

Die Geschwindigkeiten  $\mathbf{v}_A(0)$  werden typischerweise als Maxwell-Boltzmann-Verteilung bei einer bestimmten Temperatur erzeugt und legen zusammen mit der Molekülkonformation  $\mathbf{R}_A(0)$  die Anfangsbedingungen einer einzelnen Trajektorie fest.

Molekulardynamik-Simulationen erlauben den durch die Ortsvektoren aller Atomkerne aufgespannten Konformationsraum abzutasten, um beispielsweise

- verschiedene lokale Minimumstrukturen oder prinzipiell auch die globale Minimumstruktur zu finden
- chemische Umwandlungen (Strukturänderungen, chemische Reaktionen) zu verfolgen
- thermodynamischer Größen mittels der statistischen Thermodynamik zu berechnen.

### 2.3 Quantenchemische Elektronenstruktur-Methoden

Im Allgemeinen ist die elektronische Schrödingergleichung (Gleichung 2.9) für ein Mehrelektronensystem nicht geschlossen analytisch lösbar. Allerdings steht eine Hierarchie an numerischen, Wellenfunktions-basierten ab-initio Methoden zur Verfügung, mit denen im Prinzip beliebige Genauigkeiten erreicht werden können. Das Problem der Wellenfunktions-basierten ab-initio Methoden besteht daher maßgeblich in ihrem Rechenaufwand, der von der Größe des molekularen Systems abhängt: Selbst die Hartree-Fock-(HF)-Methode, als einfachste Näherung innerhalb der Methodenhierarchie, skaliert konventionell mit der dritten Potenz der Systemgröße (für Details siehe Abschnitt 2.3.1). Die Abbildung 2.2 zeigt das Skalenverhalten für einige weitere Wellenfunktions-basierte ab-initio Methoden. Als Ausgangspunkt der weiterführenden Kapitel der Dissertation wird die Hartree-Fock-Methode (siehe Abschnitt 2.3.1) vorgestellt. Außerdem wird in Abschnitt 2.3.3 der Ansatz der Dichtefunktional-Theorie erläutert, die nicht auf einer elektronischen Wellenfunktion aufbaut, sondern auf der Berechnung der elektronischen Grundzustandsenergie als Funktional der Elektronendichte basiert. Die Darstellung geht dabei auch auf die methodischen Schlüsselschritte ein, mit denen man ein lineares Skalenverhalten mit der Molekülgröße erzielen kann. Der Abschnitt 2.3.2 skizziert den Ansatz von aufwändigeren Wellenfunktions-basierten Post-Hartree-Fock-Methoden.

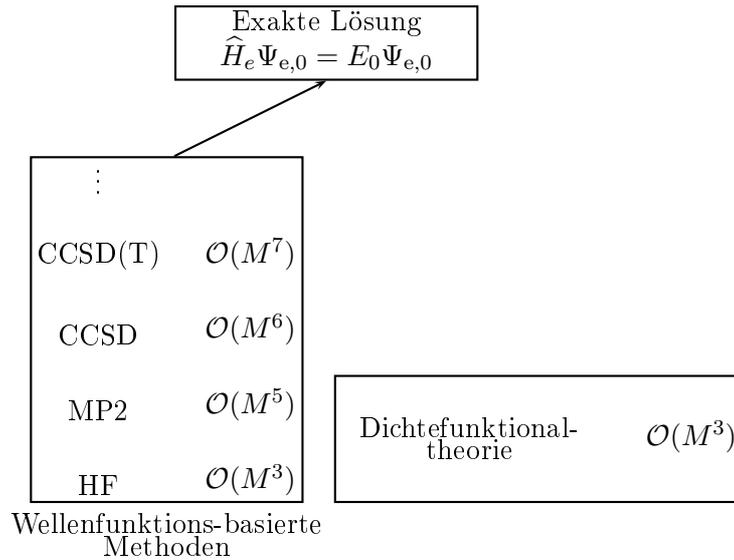


Abb. 2.2: Skalenverhalten  $\mathcal{O}(M^x)$  mit der Größe  $M$  des molekularen Systems Wellenfunktions-basierter Elektronenstruktur-Methoden (HF: Hartree-Fock, MP2: Møller-Plesset-Störungstheorie 2. Ordnung, CCSD: *Coupled-Cluster-Singles-Doubles*, CCSD(T): *Coupled-Cluster-Singles-Doubles* und störungstheoretische Korrektur der *Triples*) sowie Einordnung der Dichtefunktionaltheorie.

### 2.3.1 Die Hartree-Fock-Methode

Um den Satz an elektronischen Koordinaten  $\boldsymbol{\tau} = \{\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_{N_e}\}$  der elektronischen Wellenfunktion zu separieren, setzt man bei der HF-Methode eine einzelne Slater-Determinante der Form

$$\Psi_{e,0}(\boldsymbol{\tau}) = \frac{1}{\sqrt{N_e!}} \cdot \begin{vmatrix} \varphi_1(\boldsymbol{\tau}_1) & \varphi_2(\boldsymbol{\tau}_1) & \cdots & \varphi_{N_e}(\boldsymbol{\tau}_1) \\ \varphi_1(\boldsymbol{\tau}_2) & \varphi_2(\boldsymbol{\tau}_2) & \cdots & \varphi_{N_e}(\boldsymbol{\tau}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_1(\boldsymbol{\tau}_{N_e}) & \varphi_2(\boldsymbol{\tau}_{N_e}) & \cdots & \varphi_{N_e}(\boldsymbol{\tau}_{N_e}) \end{vmatrix} \equiv |\varphi_1 \varphi_2 \cdots \varphi_{N_e}\rangle \quad (2.16)$$

an. Die Spinfunktionen  $\varphi_i$  hängen jeweils nur von den elektronischen Koordinaten  $\boldsymbol{\tau}_j$  (kombinierte Raum-/Spinkoordinaten) eines Elektrons ab ( $j = 1, 2, \dots, N_e$ ) und können orthonormiert gewählt werden. Für geschlossenschalige Systeme kann man Paare von Spinfunktionen als Produkt einer Raumfunktion  $\phi_i(\boldsymbol{r}_j)$  mit den orthonormierten Spinfunktionen  $\alpha(\sigma_j)$  bzw.

$\beta(\sigma_j)$  erhalten (*Restricted-Hartree-Fock*, RHF):

$$\begin{aligned}\varphi_{2i-1}(\boldsymbol{\tau}_j) &= \phi_i(\mathbf{r}_j) \cdot \alpha(\sigma_j) & i = 1, 2, \dots, \frac{1}{2}N_e \\ \varphi_{2i}(\boldsymbol{\tau}_j) &= \phi_i(\mathbf{r}_j) \cdot \beta(\sigma_j)\end{aligned}\quad (2.17)$$

Innerhalb des Slater-Determinantenansatzes ergibt sich der Energieerwartungswert als

$$E_0 = \langle \Psi_{e,0} | \hat{H}_e | \Psi_{e,0} \rangle = \sum_i \langle \varphi_i(\boldsymbol{\tau}_j) | \hat{h}_j | \varphi_i(\boldsymbol{\tau}_j) \rangle + \frac{1}{2} \sum_{i,j \neq i} [\langle \varphi_i \varphi_j | \varphi_i \varphi_j \rangle - \langle \varphi_i \varphi_j | \varphi_j \varphi_i \rangle] + V_{NN} \quad (2.18)$$

mit

$$\hat{h}_j = -\frac{1}{2} \nabla_j^2 - \sum_A \frac{Z_A}{r_{jA}} \quad (2.19)$$

und

$$\langle \varphi_i \varphi_j | \varphi_i \varphi_j \rangle = \int d\boldsymbol{\tau}_1 d\boldsymbol{\tau}_2 \varphi_i^*(\boldsymbol{\tau}_1) \varphi_j^*(\boldsymbol{\tau}_2) \frac{1}{r_{12}} \varphi_i(\boldsymbol{\tau}_1) \varphi_j(\boldsymbol{\tau}_2). \quad (2.20)$$

Die Minimierung des Energie-Erwartungswerts auf der Grundlage des Variationsprinzips und unter der Nebenbedingung der Orthonormalität der Molekülorbitale liefert die allgemeinen Hartree-Fock-Gleichungen.

$$\hat{F} \varphi_i = \sum_j \epsilon_{ji} \varphi_j \quad (2.21)$$

mit den Lagrange-Multiplikatoren  $\epsilon_{ji}$ . Sie lassen sich unitär transformieren, wodurch man die kanonischen Hartree-Fock-Gleichungen erhält (siehe z.B. Referenz [17]):

$$\hat{F} \varphi_i = \epsilon_i \varphi_i \quad (2.22)$$

Der Fock-Operator  $\hat{F}$  repräsentiert als effektiver Einelektronen-Operator die kinetische Energie eines Elektrons, dessen Wechselwirkung mit den Kernen (Einelektronen-Teil) und ein gemittelttes Potential, das ein Elektron durch die Anwesenheit der anderen Elektronen erfährt (Zweielektronen-Teil). Die kanonischen Hartree-Fock-Gleichungen können im Allgemeinen nicht in einem Schritt gelöst werden, vielmehr hängt  $\hat{F}$  von allen (zunächst unbekannt)en Molekülorbitalen  $\varphi_i$  ab, so dass die Gleichungen nach dem Verfahren des selbst-konsistenten Feldes (*self-consistent field*, SCF) iterativ gelöst werden müssen. Für das algebraisch-iterative Vorgehen werden die den Spinorbitalen  $\varphi_i$  zugrunde liegenden Raumorbitale  $\phi_i(\mathbf{r}_j)$  (siehe Gleichung 2.17) üblicherweise als Linearkombination eines endlichen Satzes  $\{\chi_\mu\}$  an atomzentrierten (nicht-orthogonalen) Basisfunktionen (Atomorbitale, AO) genähert:

$$\phi_i(\mathbf{r}_j) = \sum_{\mu}^{N_{\text{bas}}} C_{\mu i} \cdot \chi_{\mu}(\mathbf{r}_j, \mathbf{R}_{\mu}). \quad (2.23)$$

Für die Atomorbitale verwendet man bei molekularen Berechnungen heute überwiegend Gaußtyp-Funktionen der Art

$$\chi_{\mu p}(\mathbf{r}_j, \mathbf{R}_\mu) = (x - R_{\mu,x})^{l_{\mu,x}} (y - R_{\mu,y})^{l_{\mu,y}} (z - R_{\mu,z})^{l_{\mu,z}} e^{-\zeta_p(\mathbf{r}_j - \mathbf{R}_\mu)^2} \quad (2.24)$$

bzw. feste Linearkombinationen mehrerer primitiver Gaußtyp-Funktionen,

$$\chi_\mu(\mathbf{r}_j, \mathbf{R}_\mu) = \sum_{p=1}^{K_\mu} D_{\mu p} \cdot \chi_{\mu p}(\mathbf{r}_j, \mathbf{R}_\mu), \quad (2.25)$$

mit dem Kontraktionskoeffizient  $D_{\mu p}$ , dem Atomzentrum  $\mathbf{R}_\mu$ , dem Exponent  $\zeta_p$ , den Drehimpulsen  $l_{\mu,x}$ ,  $l_{\mu,y}$  und  $l_{\mu,z}$ , und dem Kontraktionsgrad  $K_\mu$  (hier gezeigt für kartesische Basisfunktionen). Durch die Basissatzentwicklung gehen die kanonischen Hartree-Fock-Gleichungen für den RHF-Fall in die algebraische Form der Roothaan-Hall-Gleichung über:

$$\mathbf{F}\mathbf{C} = \mathbf{S}\mathbf{C}\epsilon \quad (2.26)$$

Die Matrixgleichung verknüpft die Fock-Matrix  $\mathbf{F}$ , die MO-Koeffizienten-Matrix  $\mathbf{C}$ , die Überlappungsmatrix  $\mathbf{S}$  und die Matrix der Orbitalenergien  $\epsilon$  miteinander. Die Fock-Matrix setzt sich aus dem Einelektronen-Teil  $\mathbf{h}$  und dem Zweielektronen-Teil  $\mathbf{G}$  zusammen,

$$\mathbf{F} = \mathbf{h} + \mathbf{G}(\mathbf{P}) \quad (2.27)$$

Für ein Element des Zweielektronen-Teils  $G_{\mu\nu}$  in der AO-Basis gilt für den RHF-Fall

$$G_{\mu\nu}(\mathbf{P}) = \underbrace{\sum_{\kappa\lambda} 2P_{\kappa\lambda} (\mu\nu|\kappa\lambda)}_{J_{\mu\nu}} - \underbrace{\sum_{\kappa\lambda} P_{\kappa\lambda} (\mu\lambda|\kappa\nu)}_{K_{\mu\nu}}, \quad (2.28)$$

mit den Matrixelementen  $P_{\kappa\lambda}$  der Einteilchen-Dichtematrix  $\mathbf{P}$

$$P_{\kappa\lambda} = \sum_i^{\frac{1}{2}N_e} C_{\kappa i} C_{\lambda i}^*, \quad (2.29)$$

den Elementen  $J_{\mu\nu}$  des Coulombteils bzw.  $K_{\mu\nu}$  des Austauschteils und den Zweielektronen-Integralen  $(\mu\nu|\kappa\lambda)$  (*electron repulsion integral, ERI*). Der Coulombteil lässt sich als klassische Wechselwirkung der Ladungsverteilungen

$$\Omega_{\mu\nu}(\mathbf{r}_1) = \chi_\mu(\mathbf{r}_1, \mathbf{R}_\mu) \cdot \chi_\nu(\mathbf{r}_1, \mathbf{R}_\nu) \quad (2.30)$$

und

$$\Omega_{\kappa\lambda}(\mathbf{r}_2) = \chi_\kappa(\mathbf{r}_2, \mathbf{R}_\kappa) \cdot \chi_\lambda(\mathbf{r}_2, \mathbf{R}_\lambda) \quad (2.31)$$

interpretieren, während der Austauschanteil als Folge des quantenmechanischen Ansatzes entsteht und im klassischen Bild allein nicht gedeutet werden kann. Jeder Iterationsschritt des SCF umfasst die Berechnung der Fock-Matrix, die Orthogonalisierung der Basis und einen Diagonalisierungsschritt, der eine neue MO-Koeffizienten-Matrix  $\mathbf{C}$  bzw. Einteilchen-Dichtematrix  $\mathbf{P}$  erzeugt. Ist die Konvergenz gemäß eines festgelegten Grenzwerts  $10^{-\vartheta_{\text{SCF}}}$  erreicht, wird der Zyklus abgebrochen, ansonsten erneut durchlaufen. Als Konvergenzkriterium kann beispielsweise die Norm (bzw. das Maximalelement) der DIIS-Fehlermatrix berechnet werden,

$$\boldsymbol{\varepsilon} \equiv \mathbf{F}\mathbf{P}\mathbf{S} - \mathbf{S}\mathbf{P}\mathbf{F}, \quad (2.32)$$

die im Falle der Konvergenz  $\mathbf{0}$  wird. „DIIS“ (*direct inversion of the iterative subspace*, [18]) bezeichnet dabei einen häufig verwendeten Algorithmus zur Beschleunigung der SCF-Konvergenz. Die erste Iteration beginnt notwendigerweise mit einer Anfangsschätzung, z.B. der Superposition atomarer Dichten (SAD), zu der man durch die Summation sphärisch gemittelter, atomarer Dichten gelangt. In Abhängigkeit des Einelektronen-Teils  $\mathbf{h}$ , der Einteilchen-Dichtematrix  $\mathbf{P}$  und der Fock-Matrix  $\mathbf{F}$  lässt sich der RHF-Energieerwartungswert berechnen als

$$E_0 = \text{tr}\{\mathbf{P}[2 \cdot \underbrace{\mathbf{h} + \mathbf{G}(\mathbf{P})}_{\mathbf{F}}]\} + V_{\text{NN}}. \quad (2.33)$$

Der Aufwand, um die Fock-Matrix  $\mathbf{F}$  zu berechnen, steigt formal mit der vierten Potenz der Anzahl an Basisfunktionen  $N_{\text{bas}}$  an, da zur Bildung von  $\mathbf{F}$  in einem konventionellen SCF-Verfahren  $N_{\text{bas}}^4$  Zweielektronen-Integrale  $(\mu\nu|\kappa\lambda)$  berechnet und mit der Einteilchen-Dichtematrix  $\mathbf{P}$  kontrahiert werden müssen. Jedoch klingen die atomzentrierten Basisfunktionen mit zunehmendem Abstand vom jeweiligen Zentrum  $\mathbf{R}_\mu$  exponentiell ab. Zu jeder Basisfunktion  $\chi_\mu(\mathbf{r}_j, \mathbf{R}_\mu)$  gibt es daher nur eine konstante Zahl an Basisfunktionen  $\chi_\nu(\mathbf{r}_j, \mathbf{R}_\nu)$ , die mit ihr eine numerisch signifikante Ladungsverteilung  $\Omega_{\mu\nu}$  bilden. Das Integralscreening, basierend auf der Cauchy-Schwarz'schen Ungleichung, wurde 1989 von Häser und Ahlrichs [19] eingeführt und berücksichtigt dieses Verhalten:

$$|(\mu\nu|\kappa\lambda)| \leq \underbrace{(\mu\nu|\mu\nu)}_{Q_{\mu\nu}}^{\frac{1}{2}} \underbrace{(\kappa\lambda|\kappa\lambda)}_{Q_{\kappa\lambda}}^{\frac{1}{2}} \quad (2.34)$$

Zweielektronen-Integrale werden vernachlässigt, wenn die mit  $\mathcal{O}(N_{\text{bas}}^2)$ -Aufwand berechnete obere Schwarz-Schranke  $Q_{\mu\nu}Q_{\kappa\lambda}$  kleiner als ein festgelegter Grenzwert  $10^{-\vartheta_{\text{Int}}}$  ist, wodurch insgesamt das Skalenverhalten zur Bildung des Zweielektronen-Teils  $\mathbf{G}$  auf  $\mathcal{O}(N_{\text{bas}}^2)$  sinkt.

Darüber hinaus ist für ein molekulares, nicht-metallisches System die Einteilchen-Dichtematrix typischerweise dünn besetzt (*sparse*), d.h. die Elemente  $P_{\kappa\lambda}$  klingen für zunehmende Abstände der Zentren  $\mathbf{R}_\kappa$  und  $\mathbf{R}_\lambda$  der Basisfunktion  $\chi_\kappa$  und  $\chi_\lambda$  rasch ab [20]. Deshalb ist es besonders vorteilhaft, die Einteilchen-Dichtematrix  $\mathbf{P}^n$  der aktuellen Iteration  $n$  in das Integralscreening miteinzubeziehen,

$$|P_{\kappa\lambda}^n(\mu\kappa|\lambda\nu)| \leq |P_{\kappa\lambda}^n|Q_{\mu\kappa}Q_{\lambda\nu}, \quad (2.35)$$

bzw. die aktuelle Differenzdichtematrix  $\Delta\mathbf{P}^n = \mathbf{P}^{n+1} - \mathbf{P}^n$

$$|\Delta P_{\kappa\lambda}^n(\mu\kappa|\lambda\nu)| \leq |\Delta P_{\kappa\lambda}^n|Q_{\mu\kappa}Q_{\lambda\nu}, \quad (2.36)$$

für den Fall einer inkrementellen Bildung der Fock-Matrix [21] gemäß

$$\mathbf{F}^{n+1} = \mathbf{F}^n + \mathbf{G}(\Delta\mathbf{P}^n). \quad (2.37)$$

Für den Austauschteil, in denen die Einteilchen-Dichtematrix den Bra- und Ketteil der Zweielektronen-Integrale miteinander koppelt, kann man dann in einem weiten Bereich an Molekülgrößen ein lineares Skalenverhalten erzielen, mittels eines linear-skalierenden Screenings (LinK, [22,23]). Das Skalenverhalten des Coulombteils lässt sich mithilfe der von White et al. entwickelten *Continuous fast multipole method* (CFMM, [24]) auf linear reduzieren.

Asymptotisch wird das Skalenverhalten eines konventionellen SCF-Verfahrens durch den Aufwand zur Diagonalisierung der Fock-Matrix bestimmt ( $\mathcal{O}(N_{\text{bas}}^3)$ ). Jedoch ist es inzwischen möglich mit diagonalisierungsfreien Algorithmen (siehe beispielsweise Referenz [2] und darin enthaltene Referenzen) den  $\mathcal{O}(N_{\text{bas}}^3)$ -Schritt zu umgehen. Dabei sind die diagonalisierungsfreien Algorithmen insbesondere für sehr große molekulare Systeme (mehrere tausend Atome) wichtig, da der  $\mathcal{O}(N_{\text{bas}}^3)$ -Schritt typischerweise einen kleinen Vorfaktor hat und daher verglichen mit der Bildung der Fock-Matrix erst spät zum zeitbestimmenden Schritt wird.

### 2.3.2 Post-Hartree-Fock-Methoden

In der Hartree-Fock-Theorie wird die Mehrelektronen-Wellenfunktion als eine einzelne, aus Spinorbitalen zusammengesetzte Slater-Determinante (siehe Gleichung 2.16) genähert. Als Konsequenz beschreibt die Hartree-Fock-Theorie die Wechselwirkung eines jeden Elektrons näherungsweise als Wechselwirkung mit einem gemittelten Feld der anderen Elektronen; die Bewegung der Elektronen sind nicht korreliert, d.h. die Wahrscheinlichkeit ein Elektron an

einem Ort  $\mathbf{r}_1$  zu finden und ein anderes gleichzeitig am Ort  $\mathbf{r}_2$  ergibt sich als einfaches Produkt der jeweils separaten Aufenthaltswahrscheinlichkeiten. Der Ausdruck „Korrelationsenergie“ ist definiert als Differenz zwischen der exakten Energie und der Hartree-Fock-Energie im Basissatz-Limit.

Bei der HF-Methode sind typischerweise mehr Molekülorbitale zugänglich als für die Besetzung mit Elektronen benötigt. Während in die HF-Slater-Determinante nur die Spinorbitale mit den niedrigsten Orbitalenergien eingehen, können weitere Slater-Determinanten durch andere Orbital-Kombinationen (Konfigurationen) erzeugt werden. Prinzipiell lässt sich die im Rahmen der Basis optimale Wellenfunktion als Linearkombination aller möglichen Slater-Determinanten ansetzen (*Full configuration interaction*, FCI). Aufgrund des Rechenaufwands muss man sich für die allermeisten molekularen Systeme auf bestimmte Kombinationen einschränken, sei es im Sinne von abgebrochenen CI- oder *Coupled-Cluster*-(CC)-Methoden (siehe z.B. Referenzen [25–27]) mit unterschiedlichen Vor- und Nachteilen. Eine weitere Elektronenkorrelations-Methode, die Møller-Plesset-Störungstheorie, geht ebenfalls von der Hartree-Fock-Methode aus und leitet eine Korrektur für die Elektronenkorrelation auf Basis der zeitunabhängigen Störungstheorie her [17,28]. Die CCSD(T)-Methode basiert sowohl auf einer abgebrochenen CC-Entwicklung und als auch einer störungstheoretischen Korrektur (siehe z.B. Referenzen [29,30]).

### 2.3.3 Dichtefunktionaltheorie

Die Grundlage der Dichtefunktionaltheorie (DFT) ist der Beweis von Hohenberg und Kohn [31], dass die elektronische Energie des Grundzustands  $E_0$  nicht nur als Funktional der elektronischen Wellenfunktion (siehe Gleichung 2.13) bestimmt ist, sondern auch vollständig als Funktional  $E[\rho]$  der Elektronendichte  $\rho(x, y, z)$ . Anstelle die elektronische Schrödingergleichung (näherungsweise) durch Wellenfunktions-basierte Methoden zu lösen, bietet sich damit prinzipiell die Möglichkeit, eine gültige Beschreibung des im Allgemeinen unbekanntes Energiefunktionals  $E[\rho]$  herzuleiten. Der formale Vorteil in der Formulierung von  $E_0$  als Funktional der Elektronendichte  $\rho(x, y, z)$  liegt dabei darin, dass die Elektronendichte nur von drei Raumkoordinaten abhängt, wohingegen die elektronische Wellenfunktion eine Funktion der Koordinaten aller  $N_e$  Elektronen des Systems ist. Die gegenwärtig angewendeten DFT-Methoden gründen jedoch im Wesentlichen wiederum auf einem Orbital-basiertem Verfahren,

das von Kohn und Sham 1965 eingeführt wurde, um die Schwierigkeiten bei der direkten Bestimmung des kinetischen Anteils von  $E[\rho]$  zu überwinden [32]. Der Kohn-Sham-Ansatz (KS) lässt sich analog zur Hartree-Fock-Theorie formulieren, wobei nur der Austauschanteil  $\mathbf{K}$  im Zweielektronen-Teil der Fock-Matrix (siehe Gleichung 2.28) ersetzt werden muss, durch einen Austausch-Korrelationsanteil  $\mathbf{V}_{\text{XC}}$  mit den Elementen

$$\mathbf{V}_{\mu\nu}^{\text{XC}} = \langle \mu | \frac{\partial E_{\text{XC}}}{\partial \rho} | \nu \rangle \quad (2.38)$$

und der Austausch-Korrelationsenergie

$$E_{\text{XC}} = \int f_{\text{XC}} dx dy dz. \quad (2.39)$$

Die Gleichung 2.39 muss man in der Regel durch numerische Verfahren integrieren (für ein linear-skalierendes Verfahren siehe z.B. Referenz [33]). Die exakte Form des Funktionals  $f_{\text{XC}}$  ist im Allgemeinen nicht bekannt und die heute verwendeten Funktionale lassen sich gemäß ihrer grundlegenden Abhängigkeiten von der Elektronendichte in drei Klassen einteilen [34]:

- L(S)DA-Funktionale
- GGA-Funktionale
- Meta-GGA-Funktionale

Die einfachsten Näherungen innerhalb der vorgeschlagenen heuristischen Rangfolge [35, 36], die L(S)DA-Funktionale, basieren auf der *local (spin) density approximation* [37, 38] und hängen nur von der Elektronendichte selbst ab. Die GGA-Funktionale (z.B. PBE [39]) bauen auf der *generalized gradient approximation* auf, d.h. hängen zudem von der ersten Ableitung der Dichte ab. Als Meta-GGA bezeichnet man Funktionale, die zusätzlich entweder von höheren Ableitungen oder der kinetischen Orbitalenergiedichte  $\tau$  [40] abhängen. Eine Verbesserung über die reinen Funktionale hinaus erreicht man typischerweise, indem man in Hybridfunktionalen (z.B. B3LYP [41, 42], PBE0 [43], B97-2 [44]) den Austausch-Korrelationsanteil  $\mathbf{V}_{\text{XC}}$  aus einem Anteil  $\gamma$  exaktem Hartree-Fock-Austausch  $\mathbf{K}$  und einem reinen DFT-Anteil  $\mathbf{V}'_{\text{XC}}$  zusammensetzt:

$$\mathbf{V}_{\text{XC}} = \gamma \mathbf{K} + \mathbf{V}'_{\text{XC}} \quad (2.40)$$

## 2.4 Berechnung statischer Moleküleigenschaften

Die vorausgehenden beiden Abschnitte stellen Methoden vor, mit denen man die elektronische Wellenfunktion bzw. Energie des Grundzustands für eine gegebene Molekülkonformation und in Abwesenheit externer elektrischer oder magnetischer Felder (näherungsweise) bestimmen kann. Ausgehend von der elektronischen Grundzustandsenergie lassen sich viele weitere molekulare Eigenschaften als Antwort des molekularen Systems auf eine Störung berechnen. Die Störung kann dabei sowohl einen externen Ursprung haben (z.B. ein elektrisches oder magnetisches Feld) oder auch einen internen Ursprung (z.B. eine Kernverrückung oder kernmagnetisches Moment) und kann prinzipiell zeitabhängig (dynamische Eigenschaft) oder zeitunabhängig (statische Eigenschaft) sein. Als Voraussetzung der nachstehenden Kapitel wird hier nur auf statische Moleküleigenschaften (z.B. molekulare Kräfte, NMR-Abschirmungen) eingegangen. Bei der auf Energieableitungen basierenden Störungsmethode setzt man die elektronische Energie  $E(\xi)$  für den Fall einer schwachen Störung  $\xi$  als Taylor-Entwicklung um den ungestörten Fall (0) an:

$$E(\xi) = E(\xi = 0) + \left. \frac{\partial E}{\partial \xi} \right|_{\xi=0} \xi + \frac{1}{2} \left. \frac{\partial^2 E}{\partial \xi^2} \right|_{\xi=0} \xi^2 + \frac{1}{6} \left. \frac{\partial^3 E}{\partial \xi^3} \right|_{\xi=0} \xi^3 + \dots = \sum_{k=0}^{\infty} \frac{1}{k!} \left. \frac{\partial^k E}{\partial \xi^k} \right|_{\xi=0} \cdot \xi^k \quad (2.41)$$

Allgemeiner gilt dann für die Störungen durch ein externes elektrisches Feld  $\mathbf{Q}$ , magnetisches Feld  $\mathbf{B}$ , einer Kernverrückung  $\mathbf{R}_A$  oder einem kernmagnetischen Moment  $\boldsymbol{\mu}_A$ :

$$E(\mathbf{Q}, \mathbf{B}, \mathbf{R}_A, \boldsymbol{\mu}_A) = \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \frac{1}{k! l! m! n!} \left. \frac{\partial^k \partial^l \partial^m \partial^n E}{\partial \mathbf{Q}^k \partial \mathbf{B}^l \partial \mathbf{R}_A^m \partial \boldsymbol{\mu}_A^n} \right|_{\substack{\mathbf{Q}=0, \mathbf{B}=0, \\ \mathbf{R}_A=0, \boldsymbol{\mu}_A=0}} \cdot \mathbf{Q}^k \cdot \mathbf{B}^l \cdot \mathbf{R}_A^m \cdot \boldsymbol{\mu}_A^n. \quad (2.42)$$

Indem man die grundlegenden physikalischen Abhängigkeiten der Wechselwirkung berücksichtigt, lassen sich die molekularen Eigenschaften als partielle Energieableitung

$$\left. \frac{\partial^k \partial^l \partial^m \partial^n E}{\partial \mathbf{Q}^k \partial \mathbf{B}^l \partial \mathbf{R}_A^m \partial \boldsymbol{\mu}_A^n} \right|_{\substack{\mathbf{Q}=0, \mathbf{B}=0, \\ \mathbf{R}_A=0, \boldsymbol{\mu}_A=0}}$$

in der Taylor-Entwicklung identifizieren (siehe z.B. Referenzen [1, 45]). So liefert beispielsweise der negative Gradient der elektronischen Energie nach der Kernkoordinate  $\mathbf{R}_A$  die Kraft auf den Atomkern A:

$$\mathbf{F}_A = - \underbrace{\frac{\partial}{\partial \mathbf{R}_A}}_{\nabla_A} E_0 \quad (2.43)$$

Die gemischte zweite Ableitung nach äußerem Magnetfeld  $\mathbf{B}$  und kernmagnetischen Moment  $\boldsymbol{\mu}_A$  ergibt den kernmagnetischen Abschirmtensor  $\boldsymbol{\sigma}_A$  am Atom A:

$$\boldsymbol{\sigma}_A = \frac{\partial^2 E_0}{\partial \mathbf{B} \partial \boldsymbol{\mu}_A} = \begin{pmatrix} \frac{\partial^2 E_0}{\partial B^x \partial \mu_A^x} & \frac{\partial^2 E_0}{\partial B^x \partial \mu_A^y} & \frac{\partial^2 E_0}{\partial B^x \partial \mu_A^z} \\ \frac{\partial^2 E_0}{\partial B^y \partial \mu_A^x} & \frac{\partial^2 E_0}{\partial B^y \partial \mu_A^y} & \frac{\partial^2 E_0}{\partial B^y \partial \mu_A^z} \\ \frac{\partial^2 E_0}{\partial B^z \partial \mu_A^x} & \frac{\partial^2 E_0}{\partial B^z \partial \mu_A^y} & \frac{\partial^2 E_0}{\partial B^z \partial \mu_A^z} \end{pmatrix} \quad (2.44)$$

Die isotrope Abschirmung ergibt sich aus dem Abschirmtensor als

$$\sigma_A = \frac{1}{3} \cdot \text{tr}(\boldsymbol{\sigma}_A) \quad (2.45)$$

und NMR-Verschiebungen gegenüber einer Referenz (Ref.) als

$$\delta_A = \frac{\nu_A - \nu_{\text{Ref.}}}{\nu_{\text{Ref.}}} = \frac{(1 - \sigma_A) - (1 - \sigma_{\text{Ref.}})}{1 - \sigma_{\text{Ref.}}} \simeq \frac{\sigma_{\text{Ref.}} - \sigma_A}{1}, \quad (2.46)$$

mit den Larmorfrequenzen  $\nu_A$  und  $\nu_{\text{Ref.}}$ .

Die Tabelle 2.2 gibt eine Übersicht über weitere ausgewählte molekulare Eigenschaften bis zur zweiten Ordnung.

Tab. 2.2: Übersicht über die Berechnung ausgewählter molekularer Eigenschaften als  $k$ -te Ableitung nach dem elektrischen Feld  $\mathbf{Q}$ ,  $l$ -te Ableitung nach dem magnetischen Feld  $\mathbf{B}$ ,  $m$ -te Ableitung nach einer Kernkoordinate  $\mathbf{R}_A$   $n$ -te Ableitung nach einem kernmagnetischem Moment  $\boldsymbol{\mu}_A$  der elektronischen Energie

Molekulare Eigenschaft	$k$	$l$	$m$	$n$
Elektrisches Dipolmoment	1	0	0	0
Magnetisches Dipolmoment	0	1	0	0
Hyperfeinkopplungskonstante	0	0	0	1
Molekulare Kräfte	0	0	1	0
Elektrische Polarisierbarkeit	2	0	0	0
Magnetisierbarkeit	0	2	0	0
Kernspin-Kopplung	0	0	0	2
Harmonische Schwingungsfrequenzen	0	0	2	0
Kernmagnetische Abschirmung	0	1	0	1

### 2.4.1 Störungen erster Ordnung

Ausgehend von der Gleichung des RHF-Energieerwartungswerts 2.33 (oder dem analogen Kohn-Sham-Ausdruck) ergibt sich für die erste Ableitung nach einer Störung  $\xi$

$$\begin{aligned}
 \frac{\partial}{\partial \xi} E_0 &\equiv E_0^\xi = \text{tr}\{\mathbf{P}[2 \cdot \mathbf{h} + \mathbf{G}(\mathbf{P})]\}^\xi + V_{\text{NN}}^\xi & (2.47) \\
 &= \text{tr}\{\mathbf{P}^\xi[2 \cdot \mathbf{h} + \mathbf{G}(\mathbf{P})] + \mathbf{P}[2 \cdot \mathbf{h}^\xi + \mathbf{G}^\xi(\mathbf{P}) + \mathbf{G}(\mathbf{P}^\xi)]\} + V_{\text{NN}}^\xi \\
 &= 2 \cdot \text{tr}\{\underbrace{\mathbf{P}^\xi[\mathbf{h} + \mathbf{G}(\mathbf{P})]}_F\} + 2 \cdot \text{tr}\{\mathbf{P}\mathbf{h}^\xi\} + \text{tr}\{\mathbf{P}\mathbf{G}^\xi(\mathbf{P})\} + V_{\text{NN}}^\xi,
 \end{aligned}$$

wobei  $\mathbf{G}^\xi(\mathbf{P})$  definiert ist als

$$\mathbf{G}_{\mu\nu}^\xi(\mathbf{P}) = \sum_{\kappa\lambda} P_{\kappa\lambda} \left\{ 2(\mu\nu|\kappa\lambda)^\xi - (\mu\lambda|\kappa\nu)^\xi \right\} \quad (2.48)$$

und ausgenutzt wurde, dass

$$\text{tr}\{\mathbf{P}\mathbf{G}(\mathbf{P}^\xi)\} = \text{tr}\{\mathbf{P}^\xi\mathbf{G}(\mathbf{P})\}. \quad (2.49)$$

Während sich in der Gleichung 2.47 die Ableitungen des Eielektronen-Teils  $\mathbf{h}^\xi$ , der Zweielektronen-Integrale  $(\mu\nu|\kappa\lambda)^\xi$  und der Kern-Kern-Wechselwirkungsenergie  $V_{\text{NN}}^\xi$  geschlossen berechnen lassen, kann man die gestörte Einteilchen-Dichtematrix  $\mathbf{P}^\xi$  in der Regel nur iterativ bestimmen (siehe Abschnitt 2.4.2). Allerdings lassen sich die Energieableitungen erster Ordnung so umformulieren, dass die Abhängigkeit von  $\mathbf{P}^\xi$  entfällt. Dazu kann man von der Zerlegung in die Subraumprojektionen,

$$\mathbf{P}^\xi = \underbrace{\mathbf{P}\mathbf{S}\mathbf{P}^\xi\mathbf{S}\mathbf{P}}_{\mathbf{P}_{oo}^\xi} + \underbrace{\mathbf{P}\mathbf{S}\mathbf{P}^\xi(\mathbf{1} - \mathbf{S}\mathbf{P})}_{\mathbf{P}_{ov}^\xi} + \underbrace{(\mathbf{1} - \mathbf{P}\mathbf{S})\mathbf{P}^\xi\mathbf{S}\mathbf{P}}_{\mathbf{P}_{vo}^\xi} + \underbrace{(\mathbf{1} - \mathbf{P}\mathbf{S})\mathbf{P}^\xi(\mathbf{1} - \mathbf{S}\mathbf{P})}_{\mathbf{P}_{vv}^\xi}, \quad (2.50)$$

ausgehen und die Idempotenzbedingung

$$\mathbf{P} = \mathbf{P}\mathbf{S}\mathbf{P}, \quad (2.51)$$

deren abgeleitete Form

$$\mathbf{P}^\xi = \mathbf{P}^\xi\mathbf{S}\mathbf{P} + \mathbf{P}\mathbf{S}^\xi\mathbf{P} + \mathbf{P}\mathbf{S}\mathbf{P}^\xi \quad (2.52)$$

und die SCF-Konvergenzbedingung (siehe Gleichung 2.32)

$$\mathbf{F}\mathbf{P}\mathbf{S} - \mathbf{S}\mathbf{P}\mathbf{F} = \mathbf{0} \quad (2.53)$$

berücksichtigen. Für die Projektion auf den besetzt-besetzten (*occupied-occupied*) Raum ergibt sich

$$\begin{aligned}
 P_{oo}^\xi &= P S P^\xi S P = P S \left\{ P^\xi S P + P S^\xi P + P S P^\xi \right\} S P & (2.54) \\
 &= P S P^\xi S \underbrace{P S P}_P + \underbrace{P S P S^\xi P S P}_P + \underbrace{P S P S P^\xi S P}_P \\
 &= \underbrace{P S P^\xi S P}_{P_{oo}^\xi} + P S^\xi P + \underbrace{P S P^\xi S P}_{P_{oo}^\xi}
 \end{aligned}$$

und folglich

$$P_{oo}^\xi = -P S^\xi P. \quad (2.55)$$

Der unbesetzt-unbesetzte (*virtual-virtual*) Block von  $P^\xi$  verschwindet:

$$\begin{aligned}
 P_{vv} &= (1 - P S) P^\xi (1 - S P) = (P^\xi - P S P^\xi) (1 - S P) & (2.56) \\
 &= \underbrace{P^\xi - P^\xi S P - P S P^\xi}_{P S^\xi P} + \underbrace{P S P^\xi S P}_{-P S^\xi P} \\
 &= 0
 \end{aligned}$$

Berücksichtigt man die Gleichungen 2.55 und 2.56, lässt sich der von  $P^\xi$  abhängige Term aus der Gleichung 2.47 nun schreiben als

$$\begin{aligned}
 \text{tr}\{P^\xi F\} &= \text{tr}\{P_{vo}^\xi F + P_{ov}^\xi F - (P S^\xi P) F\} & (2.57) \\
 &= \text{tr}\{(1 - P S) P^\xi S P F + P S P^\xi (1 - S P) F - P S^\xi P F\} \\
 &= \text{tr}\{P^\xi \underbrace{S P F (1 - P S)}_{F_{ov}} + P^\xi \underbrace{(1 - S P) F P S}_{F_{vo}} - P S^\xi P F\} \\
 &= -\text{tr}\{P S^\xi P F\},
 \end{aligned}$$

denn innerhalb der Spur lassen sich Matrixmultiplikationen zyklisch permutieren und es gilt als Folge des Brillouin-Theorems [46]:

$$F_{vo} = F_{ov} = \mathbf{0}. \quad (2.58)$$

Insgesamt ergibt sich damit für die Energieableitung erster Ordnung aus der Gleichung 2.47

$$E_0^\xi = -2 \cdot \text{tr}\{P S^\xi P F\} + 2 \cdot \text{tr}\{P h^\xi\} + \text{tr}\{P G^\xi(P)\} + V_{\text{NN}}^\xi, \quad (2.59)$$

so dass Eigenschaften erster Ordnung (z.B. Dipolmomente, Kräfte) berechnet werden können ohne dass die gestörte Einteilchen-Dichtematrix  $P^\xi$  berechnet werden muss.

Während die Gleichung 2.59 allgemein für eine beliebige Störung gilt, hängt der Aufbau der Terme  $\mathbf{S}^\xi$ ,  $\mathbf{h}^\xi$ ,  $\mathbf{G}^\xi(\mathbf{P})$  und  $V_{\text{NN}}^\xi$  von der konkreten Störung  $\xi$  ab. Für Kraftberechnungen sind die Ableitungen  $\mathbf{S}^\xi$ ,  $\mathbf{h}^\xi$ ,  $\mathbf{G}^\xi(\mathbf{P})$  und  $V_{\text{NN}}^\xi$  ungleich null, denn alle vier Größen  $\mathbf{S}$ ,  $\mathbf{h}$ ,  $\mathbf{G}(\mathbf{P})$  und  $V_{\text{NN}}$  hängen von den Kernkoordinaten ab, für die Komponente eines Kraftvektors  $F_A^j$  (mit  $i = x, y, z$  am Atom A) gilt also:

$$F_A^i = -E_0^{R_A^i} = 2 \cdot \text{tr}\{\mathbf{P}\mathbf{S}^{R_A^i}\mathbf{P}\mathbf{F}\} - 2 \cdot \text{tr}\{\mathbf{P}\mathbf{h}^{R_A^i}\} - \text{tr}\{\mathbf{P}\mathbf{G}^{R_A^i}(\mathbf{P})\} - V_{\text{NN}}^{R_A^i} \quad (2.60)$$

Im Falle der Ableitung nach der Komponente eines kernmagnetischen Moments  $\mu_A^j$  (mit  $j = x, y, z$ ) ist hingegen der Term  $V_{\text{NN}}^\xi$  null. Ebenso fallen die Terme  $\mathbf{S}^\xi$  und  $\mathbf{G}^\xi(\mathbf{P})$  weg, denn die Atomorbitale  $\chi_\mu$  hängen, bei der typischen Wahl im Rahmen von GIAO-Methoden, nicht von den kernmagnetischen Momenten ab; die Abhängigkeit von  $\mu_A^j$  wird also nur im Eielektronenteil  $\mathbf{h}$  (als Operatorableitung) beschrieben:

$$E_0^{\mu_A^j} = 2 \cdot \text{tr}\{\mathbf{P}\mathbf{h}^{\mu_A^j}\} \quad (2.61)$$

Dabei versteht man unter „GIAO“ Eichsprungs-einschließende Atomorbitale (*gauge including atomic orbitals*, [47–49]), die standardmäßig die Magnetfeldabhängigkeit einbeziehen (nicht aber die Abhängigkeit von kernmagnetischen Momenten), um so den Fehler durch Eichsprungsvarianz für abgebrochene Basisentwicklungen zu reduzieren (für eine ausführliche Diskussion siehe z.B. Referenz [50]).

#### 2.4.2 Störungen zweiter Ordnung, CPSCF-Verfahren

Um Störungen zweiter Ordnungen zu berechnen, benötigt man neben der ungestörten Einteilchen-Dichtematrix  $\mathbf{P}$  (aus dem SCF-Verfahren) die gestörte Einteilchen-Dichtematrix  $\mathbf{P}^\xi$ . So folgt beispielsweise ausgehend von der Gleichung 2.61 für die Berechnung eines Elements des kernmagnetischen Abschirmungstensors  $\sigma_A^{ij}$ :

$$\sigma_A^{ij} = \frac{\partial^2 E_0}{\partial B^i \partial \mu_A^j} \equiv E_0^{B^i, \mu_A^j} = \frac{\partial}{\partial B^i} \left[ 2 \cdot \text{tr}\{\mathbf{P}\mathbf{h}^{\mu_A^j}\} \right] = 2 \cdot \text{tr}\{\mathbf{P}^{B^i} \mathbf{h}^{\mu_A^j}\} + 2 \cdot \text{tr}\{\mathbf{P}\mathbf{h}^{B^i, \mu_A^j}\} \quad (2.62)$$

Die gestörte Einteilchen-Dichtematrix  $\mathbf{P}^\xi$  kann für HF (und DFT-Funktionalen mit exaktem HF-Austausch) nicht in einem Schritt bestimmt werden, sondern nur iterativ, üblicherweise mithilfe des *Coupled-perturbed-self-consistent-field*-(CPSCF)-Verfahrens. Die dem CPSCF-Verfahren zugrunde liegenden CPSCF-Gleichungen können in der Dichte-basierten Formulierung ausgehend von der SCF-Konvergenzbedingung (Gleichung 2.53) hergeleitet werden. Die

SCF-Konvergenzbedingung muss für eine beliebige Störung  $\xi$  bestehen bleiben, es muss also gelten

$$\frac{\partial}{\partial \xi}(\mathbf{F}P\mathbf{S} - \mathbf{S}P\mathbf{F}) = \mathbf{0} \quad (2.63)$$

und daraus folgt durch Umstellen die Dichte-basierte Formulierung der CPSCF-Gleichungen (für eine Molekülorbital-basierte Darstellung siehe z.B. Referenz [45] und darin enthaltene Referenzen):

$$\underbrace{\mathbf{F}P^\xi\mathbf{S} - \mathbf{S}P^\xi\mathbf{F}}_{\mathbf{A}_1^\xi} + \underbrace{\mathbf{G}(P^\xi)P\mathbf{S} - \mathbf{S}P\mathbf{G}(P^\xi)}_{\mathbf{A}_2^\xi} = \underbrace{\mathbf{S}P\{\mathbf{h}^\xi + \mathbf{G}^\xi(P)\}}_{\mathbf{F}^{(\xi)}} - \underbrace{\mathbf{F}^{(\xi)}P\mathbf{S} + \mathbf{S}^\xi P\mathbf{F} - \mathbf{F}P\mathbf{S}^\xi}_{\mathbf{b}^\xi} \quad (2.64)$$

Auf der rechten Seite der CPSCF-Gleichung stehen die von  $P^\xi$  unabhängigen, d.h. im CPSCF-Verfahren konstanten, Terme ( $\mathbf{b}^\xi$ -Teil). Die linke Seite hängt von  $P^\xi$  ab und muss in jeder CPSCF-Iteration neu berechnet werden. Der  $\mathbf{A}_1^\xi$ -Teil ist dabei als Matrixprodukt der aktuellen gestörten Einteilchen-Dichtematrix  $P^\xi$  mit der Fockmatrix und Überlappmatrix zugänglich. Für den  $\mathbf{A}_2^\xi$ -Teil muss der Zweielektronen-Teil  $\mathbf{G}(P^\xi)$  als Dichtekontraktion der Zweielektronen-Integrale gebildet werden (siehe Gleichung 2.48). Nach der Berechnung des  $\mathbf{A}_2^\xi$ -Teils wird in jeder CPSCF-Iteration eine neue gestörte Einteilchen-Dichtematrix gebildet. Dazu kann man das Gleichungssystem

$$\underbrace{\mathbf{F}P^\xi\mathbf{S} - \mathbf{S}P^\xi\mathbf{F}}_{\mathbf{A}_1^\xi} = \mathbf{b}^\xi - \mathbf{A}_2^\xi \quad (2.65)$$

beispielsweise mittels eines konjugierten Gradientenverfahrens [6] lösen. Alternativ kann die neue gestörte Einteilchen-Dichtematrix nach dem Dichte-basierten-Laplace-transformierten-CPSCF-Verfahren (DL-CPSCF) auch direkt bestimmt werden (siehe Abschnitt 5.4.1 und Referenz [7]). Die erste CPSCF-Iteration gründet auf einer Anfangsschätzung, typischerweise beginnt das CPSCF-Verfahren ausgehend nur vom  $\mathbf{b}^\xi$ -Teil (der  $\mathbf{A}_2^\xi$ -Teil auf der rechten Seite von der Gleichung 2.65 ist null). Die Konvergenz des CPSCF-Verfahren kann wie beim SCF durch die DIIS-Methode [18] beschleunigt werden. Die CPSCF-DIIS-Fehlermatrix in der n-ten Iteration kann analog zum SCF-Kriterium als

$$\boldsymbol{\varepsilon}^{(n)} \equiv \mathbf{A}_2^{\xi(n)} - \overbrace{\left( \mathbf{b}^\xi - \mathbf{A}_1^{\xi(n)} \right)}^{\mathbf{A}_2^{\xi(n-1)}} \quad (2.66)$$

berechnet werden.

Die Dichte-basierten Formulierungen ermöglichen für nichtmetallische Systeme ein lineares Skalungsverhalten der Rechenzeit, wenn sie mit Sparse-Algebra-Algorithmen kombiniert werden, d.h. mit Algorithmen [51], die dünne Besetzungen (*sparsity*) von Matrizen (z.B. von  $\mathbf{P}$ ) ausnutzen können [52].

## 2.5 Molekularmechanische Methoden

Neben den quantenmechanischen (QM) Methoden (siehe Abschnitt 2.3) bildet die Molekularmechanik (MM) die Grundlage der vorgestellten QM/MM-Methoden und -Berechnungen der folgenden Kapitel. Daher sollen hier die Kernaspekte der molekularmechanischen Methoden vorgestellt werden. Anders als die quantenchemischen Methoden (z.B. Hartree-Fock oder KS-DFT) basieren die molekularmechanischen Methoden (oder auch Kraftfeldmethoden) nicht auf einer Beschreibung der Elektronenstruktur, sondern verbleiben auf einer atomaren Ebene: Die elektronische Energie wird nicht als Funktional der elektronischen Wellenfunktion oder Elektronendichte bestimmt, sondern als eine vorparametrisierte Funktion  $E_{\text{MM}}$  der Atomkoordinaten  $\mathbf{R}$ . Die Funktion  $E_{\text{MM}}$  setzt sich typischerweise zusammen aus Energiebeiträgen für die Dehnung von Bindungen  $E^{1\leftrightarrow 2}$ , der Beugung von Bindungswinkeln  $E^{1\leftrightarrow 3}$ , der Rotation von Bindungsebenen  $E^{1\leftrightarrow 4}$  (Bindungswechselwirkungen), sowie aus Van-der-Waals-Termen  $E^{\text{vdW}}$  und Coulomb-Wechselwirkungs-Termen  $E^{\text{qq}}$  (Nicht-Bindungswechselwirkungen):

$$E_{\text{MM}} = \underbrace{\sum_n^{N_{1\leftrightarrow 2}} E_n^{1\leftrightarrow 2}}_{E_{1\leftrightarrow 2}} + \underbrace{\sum_n^{N_{1\leftrightarrow 3}} E_n^{1\leftrightarrow 3}}_{E_{1\leftrightarrow 3}} + \underbrace{\sum_n^{N_{1\leftrightarrow 4}} E_n^{1\leftrightarrow 4}}_{E_{1\leftrightarrow 4}} + \underbrace{\sum_n^{N_{\text{vdW}}} E_n^{\text{vdW}}}_{E_{\text{vdW}}} + \underbrace{\sum_n^{N_{\text{qq}}} E_n^{\text{qq}}}_{E_{\text{qq}}} \quad (2.67)$$

Die Schreibweise „1 ↔ 2“, „1 ↔ 3“ bzw. „1 ↔ 4“ gibt dabei an, ob sich die Terme über zwei, drei bzw. vier miteinander gebundene Atome erstrecken. Die Formen der Energierterme variieren für verschiedene Kraftfelder. Typischerweise werden die Terme  $E_n^{1\leftrightarrow 2}$  und  $E_n^{1\leftrightarrow 3}$  durch einfache Potenzreihen-Entwicklungen um einen festgelegten Gleichgewichts-Bindungsabstand  $R_n^{\text{equ}}$  bzw. -winkel  $\Theta_n^{\text{equ}}$  angesetzt, für das Beispiel des Amber-Kraftfeldes [53] als einfache quadratische Funktionen

$$\begin{aligned} E_n^{1\leftrightarrow 2} &= k_n (R_n - R_n^{\text{equ}})^2, \\ E_n^{1\leftrightarrow 3} &= k'_n (\Theta_n - \Theta_n^{\text{equ}})^2, \end{aligned} \quad (2.68)$$

mit dem Bindungsabstand  $R_n$ , dem Bindungswinkel  $\Theta_n$  und den Kraftfeldparametern  $k_n$ ,  $k'_n$ , etc. Die mit Änderung von Bindungsebenen verbundenen Energien werden häufig mit

periodischen Funktionen beschrieben,

$$E_n^{1\leftrightarrow 4} = k_n'' (1 + \cos[\omega_n]) + k_n''' (1 - \cos[2\omega_n]) + \dots \quad (2.69)$$

mit den Diederwinkeln  $\omega_n$ , teilweise zusätzlich mit Korrekturen für Änderungen von Außer-Ebenen-Winkeln. Van-der-Waals-Wechselwirkungen zwischen Atompaaaren im Abstand  $R_n$  sind häufig als Lennard-Jones-Potential [54] wiedergegeben

$$E_n^{\text{vdW}} = \frac{c_n}{R_n^{12}} - \frac{c_n'}{R_n^6} \quad (2.70)$$

und die langreichweitigen Coulomb-Wechselwirkungen im einfachsten Fall als

$$E_n^{\text{qq}} = \frac{q_{n,1} \cdot q_{n,2}}{R_n} \quad (2.71)$$

mit den partiellen Punktladungsparametern  $q_{n,1}$  und  $q_{n,2}$ . Die Kraftfeldparameter werden durch Fitten an experimentelle Daten oder mithilfe von ab-initio Rechnungen Atomtyp-basiert oder für Einzelfragmente (z.B. Aminosäuren, Nukleotide etc.) bestimmt (für eine Übersicht und Einteilung der Kraftfelder siehe z.B. Referenz [1]).

Die Kräfte für Molekulardynamik-Rechnungen (Abschnitt 2.2) oder Geometrieoptimierungen berechnet man analog zu der Gleichung 2.12 dann näherungsweise als

$$\mathbf{F}_A = -\nabla_A E_{\text{MM}}. \quad (2.72)$$

Neben der prinzipiellen Frage nach der Verfügbarkeit und Qualität der Kraftfeldparameter für ein bestimmtes System lassen sich einige besondere Problematiken einer einfachen Kraftfeldbeschreibung festmachen:

- Die Bindungstypen müssen vor der Rechnung festgelegt werden und können sich nicht ändern
- Kopplungen zwischen den Energietermen sind vernachlässigt (z.B. insbesondere Polarisierungseffekte bei den elektrostatischen Wechselwirkungen)
- Gleichgewichtsferne Strukturen werden in der Regel schlecht beschrieben
- Chemische Reaktionen (z.B. schon einfache Protonenübertragungen) können im Allgemeinen nicht beschrieben werden

## 2.6 QM/MM-Methoden

Der Abschnitt 2.3 führt quantenchemische Methoden ein, mit deren Hilfe molekulare Systeme auf hohem theoretischen Niveau beschrieben werden können. Neben den quantenchemischen Methoden ermöglichen molekularmechanische Methoden eine effiziente Beschreibung auf einem niedrigeren theoretischen Niveau, wobei allerdings teilweise die Genauigkeiten nicht ausreichen (siehe Abschnitt 2.5). Um eine innere Molekülregion (z.B. ein gelöster Stoff oder ein aktives Enzymzentrum) möglichst genau zu beschreiben und gleichzeitig große Umgebungen möglichst effizient zu berücksichtigen, kombinieren QM/MM-Ansätze die quantenchemische mit der molekularmechanischen Beschreibung.

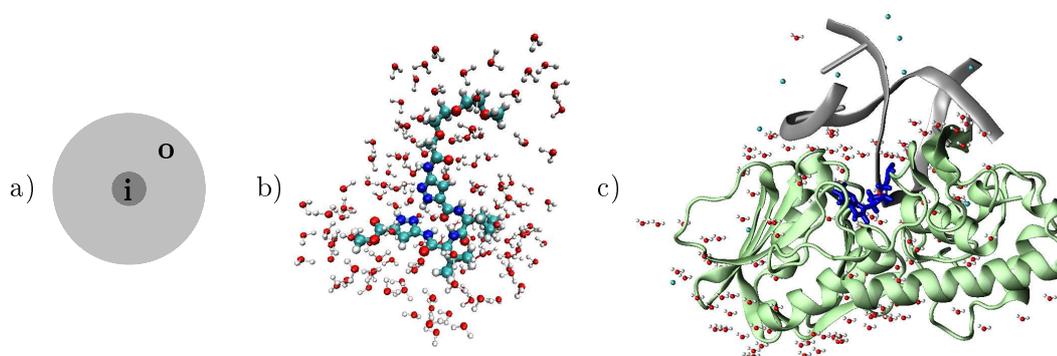


Abb. 2.3: a) Schema einer inneren (*inner*) und äußeren (*outer*) Molekülregion als Grundlage des QM/MM-Ansatzes; b) Beispiel eines gelösten Moleküls [55] (innere Region) in Wasser (äußere Region); c) Beispiel eines DNA-Enzym-Komplexes [56]; das aktive Zentrum in der Nähe des blau gezeigten Residuums bildet die innere Region, die DNA-Proteinumgebung die äußere Molekülregion.

Prinzipiell lässt sich der QM/MM-Ansatz sowohl in einer additiven als auch subtraktiven Weise formulieren [57]. Die Berechnungen und Untersuchungen der nachfolgenden Kapitel basieren auf der additiven Formulierung. Ausgehend vom quantenmechanischen Standpunkt setzt man für die additive Formulierung den Hamilton-Operator des Gesamtsystems formal zusammen als

$$\hat{H} = \hat{H}^i + \hat{H}^o + \hat{H}^{i,o}. \quad (2.73)$$

Die Operatoren  $\hat{H}^i$  und  $\hat{H}^o$  beschreiben jeweils die Hamilton-Operatoren der isolierten inneren bzw. äußeren Molekülregion und  $\hat{H}^{i,o}$  repräsentiert die Wechselwirkung zwischen der inneren

und äußeren Molekülregion. Der Energieerwartungswert setzt sich dann ebenfalls additiv zusammen:

$$\underbrace{\langle \Psi | \hat{H} | \Psi \rangle}_E = \underbrace{\langle \Psi | \hat{H}^i | \Psi \rangle}_{E^i} + \underbrace{\langle \Psi | \hat{H}^o | \Psi \rangle}_{E^o} + \underbrace{\langle \Psi | \hat{H}^{i,o} | \Psi \rangle}_{E^{i,o}} \quad (2.74)$$

Die (vom QM-Standpunkt aus gravierende) Näherung des QM/MM-Ansatzes ersetzt den Energieerwartungswert für die äußere Region  $E^o$  durch den entsprechenden Kraftfeldausdruck  $E_{\text{MM}}^o$  (siehe vorangehender Abschnitt):

$$E \approx E^i + E_{\text{MM}}^o + E^{i,o} \quad (2.75)$$

Ein Kernaspekt der verschiedenen QM/MM-Schemata liegt darin, wie die Wechselwirkung  $E^{i,o}$  zwischen QM- und MM-Teilsystem beschrieben wird. Im Allgemeinen unterscheidet man dazu eine Hierarchie von drei Modellen: die mechanische, elektrostatische und polarisierbare Einbettung [58]. Bei der mechanischen Einbettung wird die QM/MM-Wechselwirkung nur molekularmechanisch behandelt und die äußeren Atome werden für die QM-Rechnung vollständig vernachlässigt. Der Term  $E^{i,o}$  beinhaltet also nur konventionelle MM-artige Beiträge für Bindungen, van-der-Waals und klassische elektrostatische Wechselwirkungen (für die Notation und prinzipielle Form der einzelnen Terme siehe den vorangehenden Abschnitt):

$$E^{i,o} \approx E_{\text{MM}}^{i,o} = E_{1 \leftrightarrow 2}^{i,o} + E_{1 \leftrightarrow 3}^{i,o} + E_{1 \leftrightarrow 4}^{i,o} + E_{\text{VdW}}^{i,o} + E_{\text{qq}}^{i,o} \quad (2.76)$$

Die nächste Stufe, die elektrostatische Einbettung, berücksichtigt hingegen den elektrostatischen Beitrag der QM/MM-Wechselwirkung im QM-Teil und nur die Bindungs- und Van-der-Waals-Wechselwirkungen werden weiterhin molekularmechanisch beschrieben:

$$E^{i,o} \approx \underbrace{E_{1 \leftrightarrow 2}^{i,o} + E_{1 \leftrightarrow 3}^{i,o} + E_{1 \leftrightarrow 4}^{i,o} + E_{\text{VdW}}^{i,o}}_{E_{\text{Bind./VdW}}^{i,o}} + \langle \Psi | \hat{H}_{\text{ES}}^{i,o} | \Psi \rangle \quad (2.77)$$

Der Operator  $\hat{H}_{\text{ES}}^{i,o}$  umfasst die elektrostatische Wechselwirkung der  $N_e$  inneren Elektronen und  $N_N$  Kerne mit den  $N_q$  äußeren Punktladungen  $q_n$  (in atomaren Einheiten)

$$\hat{H}_{\text{ES}}^{i,o} = - \sum_i^{N_e} \sum_n^{N_q} \frac{q_n}{r_{in}} + \sum_A^{N_N} \sum_n^{N_q} \frac{z_A q_n}{R_{An}} \quad (2.78)$$

und kann zum Hamilton-Operator des QM-Systems  $\hat{H}^i$  addiert werden. Für den Energieerwartungswert insgesamt gilt dann bei der elektrostatischen Einbettung:

$$E \approx \underbrace{\langle \Psi | \hat{H}^i + \hat{H}_{\text{ES}}^{i,o} | \Psi \rangle}_{\tilde{E}^{i,o}} + E_{\text{MM}}^o + E_{\text{Bind./VdW}}^{i,o} \quad (2.79)$$

Die Einbeziehung von  $\widehat{H}_{\text{ES}}^{i,o}$  im QM-Teil verändert die Lösung des SCF-Verfahrens (und darauf aufbauender Korrelationsmethoden) und ermöglicht so, dass die MM-Region die elektronische Struktur der QM-Region beeinflusst (polarisiert). Eine weitere Verbesserung der QM/MM-Kopplung kann erreicht werden, wenn polarisierbare Kraftfelder verwendet werden, um auch eine Polarisierung des MM-Teils durch den QM-Teil zuzulassen. Typischerweise steigt der Rechenaufwand für polarisierbare Einbettungsschemata allerdings bereits erheblich an. Die folgenden Kapitel der Arbeit basieren auf der elektrostatischen Einbettung als das derzeit am häufigsten verwendete Kopplungsschema.

Ein weiterer zentraler Aspekt verschiedener QM/MM-Kopplungsschemata ist, wie kovalente Bindungsbrüchen an der QM/MM-Grenzfläche beschrieben werden: So müssen als ein Beispiel kovalente Bindungen gebrochen werden, wenn das aktive Zentrum eines Enzyms die innere QM/MM-Region und das restliche Protein die äußere QM/MM-Region bilden soll. Typischerweise ist die Elektronenverteilung im realen System weder durch die entsprechenden Anionen-Kationen-Paare ( $o^-/i^+$  oder  $o^+/i^-$ ), noch durch das Radikalpaar ( $o^{\cdot}/i^{\cdot}$ ) gut repräsentiert. Daher wurden verschiedene QM/MM-Ansätze entwickelt, die das QM-Teilsystem in verschiedener Weise absättigen, im einfachsten Fall durch Wasserstoffatome. Die Gleichung (2.75) muss dann so angepasst werden, dass die Energiebeiträge der Bindungsatome (*link atoms*) oder auch einer Grenzregion (*boundary region*) einbezogen sind. Grenzregion-basierte Schemata behandeln die Grenzregion dabei sowohl im QM- als auch MM-Teil, während Bindungsatome nur im QM-Teil behandelt werden [57].

Die Gleichungen 2.76 und 2.77 formulieren den QM/MM-Ansatz für den Energieerwartungswert für den Fall der mechanischen bzw. elektrostatischen Einbettung. Davon ausgehend sind molekulare Eigenschaften auf QM/MM-Niveau als Energieableitung zugänglich (siehe Abschnitt 2.4). So kann z.B. eine Kraft auf QM/MM-Niveau als negativer Gradient der QM/MM-Energie an der Stelle eines Atomkerns A berechnet werden, also im Falle der mechanischen Einbettung als

$$\mathbf{F}_A = -\nabla_A \{E^i + E_{\text{MM}}^o + E_{\text{MM}}^{i,o}\} \quad (2.80)$$

oder im Falle der elektrostatischen Einbettung als

$$\mathbf{F}_A = -\nabla_A \{\widetilde{E}^{i,o} + E_{\text{MM}}^o + E_{\text{Bind./VdW}}^{i,o}\}. \quad (2.81)$$

Für die Ableitung nach einem inneren Kern  $A_i$  vereinfacht sich die Gleichung 2.80 zu

$$\mathbf{F}_{A_i} = -\nabla_{A_i}\{E^i + E_{\text{MM}}^{i,o}\} \quad (2.82)$$

und die Gleichung 2.81 zu

$$\mathbf{F}_{A_i} = -\nabla_{A_i}\{\tilde{E}^{i,o} + E_{\text{Bind./VdW}}^{i,o}\}. \quad (2.83)$$

Analog wird für eine Ableitung nach einem äußeren Kern  $A_o$  aus der Gleichung 2.80

$$\mathbf{F}_{A_o} = -\nabla_{A_o}\{E_{\text{MM}}^o + E_{\text{MM}}^{i,o}\}, \quad (2.84)$$

während in der Gleichung 2.81 kein Term wegfällt, also ein Kraftbeitrag für die äußeren Kerne quantenmechanisch berechnet werden muss.

Nach Gleichung 2.44 berechnet man den kernmagnetischen Abschirmtensor als gemischte, zweite Energieableitung nach dem äußeren Magnetfeld und kernmagnetischem Moment, für die QM/MM-Energie also im Falle der mechanischen Einbettung als

$$\sigma_{A_i} = \frac{\partial^2}{\partial \mathbf{B} \partial \mu_{A_i}}\{E^i + E_{\text{MM}}^o + E_{\text{MM}}^{i,o}\} = \frac{\partial^2}{\partial \mathbf{B} \partial \mu_{A_i}}E^i. \quad (2.85)$$

Da die Kraftfeldenergien üblicherweise nicht vom äußeren Magnetfeld oder kernmagnetischen Moment abhängen reduziert sich die Gleichung 2.85 zum konventionellen, rein quantenmechanischen Ausdruck. Einen Einfluss auf die Abschirmungen an den Stellen der QM-Atome durch den MM-Teil erhält man also erst für die elektrostatische Einbettung (siehe auch Kapitel 5):

$$\sigma_{A_i} = \frac{\partial^2}{\partial \mathbf{B} \partial \mu_{A_i}}\{\tilde{E}^{i,o} + E_{\text{MM}}^o + E_{\text{Bind./VdW}}^{i,o}\} = \frac{\partial^2}{\partial \mathbf{B} \partial \mu_{A_i}}\tilde{E}^{i,o} \quad (2.86)$$



## Kapitel 3

### Effiziente linear-skalierende SCF-Verfahren

#### 3.1 Einführung

Eine der Schlüsselgrößen bei quantenchemischen Berechnungen auf Hartree-Fock- oder Kohn-Sham-DFT-Niveau sind die Zweielektronen-Vierzentren-Integrale zur Beschreibung der Elektron-Elektron-Wechselwirkung, deren Berechnung formal mit der vierten Potenz der Systemgröße skaliert. Die in den letzten Jahrzehnten entwickelten Integral-Screeningverfahren zur Reduktion des formalen Skalenverhaltens (siehe Abschnitt 2.3.1) bestimmen wesentlich den Ablauf effizienter SCF-Verfahren. So ist das direkte SCF seit seiner Einführung in den 80er Jahren [19,21,59] zum Standardverfahren geworden, um die Hartree-Fock- bzw. Kohn-Sham-Gleichungen für große Moleküle zu lösen. Indem in jeder SCF-Iteration die Zweielektronen-Integrale neu berechnet werden, vermeidet die direkte Methode die Berechnung und Speicherung einer quadratischen Anzahl von Zweielektronen-Integralen vor der ersten SCF-Iteration, im Gegensatz zum indirekten SCF-Verfahren, das auf ein konventionelles Schwarz-Screening (siehe Gleichung 2.34) zurückgreift. Auf diese Weise umgeht die direkte Methode einerseits die Beschränkungen aufgrund von Speicherkapazitäten und ermöglicht zudem, dass in den Iterationen die aktuelle Eielektronen-Dichtematrix in das Integralscreening miteinbezogen werden kann (siehe Gleichung 2.35). Dadurch kann man typischerweise sehr viel mehr Integrale vernachlässigen und linear-skalierende Methoden auf der Grundlage der Lokalität der Einteilchen-Dichtematrix ableiten (siehe z.B. Referenzen [22,23]). Andererseits bedingt das mehrmalige Berechnen von Zweielektronen-Integralen natürlich einen prinzipiell überschüssigen Mehraufwand beim direkten SCF-Verfahren.

Nachfolgend wird ein verändertes quasi-indirektes bzw. semidirektes SCF Verfahren präsentiert, mit dem sowohl ein lineares Skalenverhalten als auch eine effiziente Wiederverwendung

der Zweielektronen-Integrale erzielt werden kann. Die Genauigkeit und das Skalenverhalten des konventionellen direkten SCFs bleibt dabei unverändert, da die gleichen Integrale für die Integral-Dichte-Kontraktionen (siehe  $G_{\mu\nu}(\mathbf{P})$  in Gleichung 2.28) verwendet werden und nur ihr Verwendungsweg verändert wird. Die Effizienz des Verfahrens wird für verschiedene Systeme für den zeitbestimmenden Austauschteil gezeigt. Die Darstellung des linear-skalierendes indirekten bzw. semidirekten SCF-Verfahrens im folgenden Abschnitt ist in Teilen in der Publikation,

D. Flaig and C. Ochsenfeld,  
 „Combining the Advantages of Semi-Direct Schemes and Linear-Scaling  
 Self-Consistent Field Methods“,  
*Mol. Phys.* **108**, 2725 (2010).

enthalten. Die wesentlichen Inhalte sind in Abschnitt 3.2 als Teil dieser Arbeit einbezogen, teilweise durch weitere Aspekte ergänzt, so z.B. neue Details bei der Implementierung, Untersuchungen des Skalenverhaltens und der Kombination mit einer parallelen Implementierung für die Zweielektronen-Integrale.

Im Anschluss an die neue Integralspeicher-Methode wird eine ergänzende Screeningmethode beschrieben, die eine weitere Effizienzsteigerung gegenüber konventionellen SCF-Methoden ermöglicht. Die neue Screeningmethode berücksichtigt Abstandsinformationen zwischen den Bra- und Ket-Ladungsverteilungen (siehe Gleichung 2.30 und 2.31) des Zweielektronen-Integrals, wie ausführlich in einer kürzlich veröffentlichten Arbeit von S. A. Maurer et al. [60] dargestellt. Der Abschnitt 3.3 beschränkt sich auf die wesentlichen Beiträge zu der Veröffentlichung im Rahmen dieser Dissertation.

### 3.2 Linear-skalierendes quasi-indirektes bzw. semidirektes SCF-Verfahren

Das grundlegende Konzept des neuen linear-skalierenden in- bzw. semidirekten SCF-Verfahrens besteht darin zu vermeiden, eine quadratische Anzahl an Zweielektronen-Integralen vor dem SCF zu berechnen, indem die dünne Besetzung der Einteilchen-Dichtematrix schon beim erstmaligen Berechnen und Speichern ausgenutzt wird. Zu diesem Zweck, werden die Integrale nicht vor dem SCF, sondern erst in einer frühen SCF Iteration gespeichert, typischerweise in der ersten teuren SCF-Iteration, in der eine vollständige und aussagekräftige Einteilchen-

Dichtematrix für das Integralscreening zur Verfügung steht. Die Auswahl basiert auf dem gängigen Dichte-basierten Schwarz-Screening (siehe Gleichung 2.35). Die zweite Iteration ist als Speicheriteration typischerweise aus zwei Gründen eine gute Wahl: Zum einen ist die Anfangsschätzung in der ersten Iteration (z.B. Superposition atomarer Dichten etc.) noch schlecht für eine aussagekräftige Integrauswahl. Allerdings ist hier die Berechnung des Zweielektronenteils sehr schnell, da die SAD-Anfangsdichte sehr dünn besetzt ist. Andererseits, schrumpft der potentielle Zeitgewinn, wenn die Integrale erst in einer späteren als der zweiten Iteration abgespeichert werden. Dabei zeigt sich, dass die Einteilchen-Dichtematrix in der zweiten Iteration schon sehr repräsentativ für die konvergierte Dichtematrix ist.

Um den zusätzlichen Aufwand durch den Speicherzugriff zu minimieren, ist die neue Methode darauf ausgelegt, die Integrale im Hauptspeicher abzulegen. Gleichwohl ist die Speicherung auf der Festplatte natürlich prinzipiell möglich, allerdings erweist sie sich für konventionelle Speichermedien als nicht lohnenswert. Die Beschränkung auf den Hauptspeicher bedeutet allerdings zugleich, dass für große Moleküle oder Basissätze zusätzliche Kriterien für die Auswahl abzuspeichernder Integrale erforderlich werden können, trotz des stetigen Anwachsens der Hauptspeicher-Kapazitäten. Deshalb übernimmt die vorgestellte Methode die etablierten Zeit- und Größenabschätzungen der semidirekten SCF-Theorie [19]; das Zeitkriterium wird durch Zählung der Gleitkomma-Operationen bei der Integralberechnung verbessert.

Um die Stärken der CFMM- [24] und LinK-Methode [22, 23] von linear-skalierenden SCF-Verfahren ausnutzen zu können, bilden die derzeitigen direkten Methoden den Austausch- und Coulombteil getrennt. Dabei ist der Austauschteil in der Regel etwa drei- bis fünfmal so zeitaufwändig wie der Coulombteil, abhängig von den Integral-Grenzwerten, Basissätzen und dem molekularen System. Aus diesem Grund konzentrieren sich die Untersuchungen hier auf den Austauschteil, wobei die Methode prinzipiell auch auf die Nahfeld-Integrale des Coulombteils übertragen werden kann.

### 3.2.1 Implementierung

Die neue Integral-Speichermethode wurde in einer Entwicklungsversion von Q-Chem implementiert. Die Abbildung 3.1 fasst die Implementierung zusammen, die aus drei Hauptschritten besteht: Die erste Iteration ist vollkommen identisch zu einem konventionellen direkten SCF-Verfahren, das von einer Standardanfangsschätzung ausgeht, z.B. der Superposition ato-

- **1. Iteration**
  - konventioneller direkter Durchlauf basierend auf der Anfangsschätzung (z.B. SAD)
- **2. Iteration**
  - Auswahl-Durchlauf
  - Berechnung der signifikanten Zweielektronen-Integrale
  - Speicherung der ausgewählten Zweielektronen-Integrale
- **folgende Iterationen**
  - verändertes Integral-Screening
  - Berechnung der fehlenden Zweielektronen-Integrale
  - Einlesen bereits berechneter Zweielektronen-Integrale

Abb. 3.1: Übersicht über die Kernbestandteile der quasi-indirekten bzw. semidirekten Methode

marer Dichten (SAD), die typischerweise zu einer schnellen Berechnung der ersten Iteration führt. Die zweite Iteration ist diejenige, in der die Integrale ausgewählt, gescreent, berechnet und im Hauptspeicher abgelegt werden. In den nachfolgenden Iterationen wird dann ein verändertes Integralscreening durchlaufen, die Integrale werden wiedereingelesen und — falls nötig— zusätzliche Integrale berechnet. Der Auswahl-Durchlauf in der zweiten Iteration beinhaltet ein Dichte-basiertes Schwarz-Screening und setzt die zusätzlichen Auswahlkriterien um, sowohl das Zeit- als auch das Größenkriterium. Da weder die tatsächliche Größe eines bestimmten Zweielektronen-Integrals noch die erforderliche Zeit zu dessen Berechnung im Voraus bekannt ist, muss man wiederum auf Abschätzungen zurückgreifen. So ermöglicht ein zusätzlicher Benutzer-definierter Grenzwert  $\vartheta_{\text{size}}$  auf der Grundlage der konventionellen Schwarz-Schätzer große Zweielektronen-Integrale für die Speicherung auszuwählen. Denn die großen Zweielektronen-Integrale werden am wahrscheinlichsten in späteren SCF-Iterationen wiederverwendet, vor allem wenn das Integralscreening Differenzdichten verwendet und die Fock-Matrix inkrementell gebildet wird [19, 21].

Das zweite Kriterium, die Zeit zur Berechnung eines Integrals, kann man durch die Anzahl an Gleitkomma-Operationen  $n_{\text{flop}}$  abschätzen, die ein bestimmter Berechnungspfad für eine Integralklasse benötigt, geteilt durch die Anzahl an Integralen pro Klasse. Als Integralklasse ist dabei der Satz an Integralen definiert, der den gleichen Drehimpulstyp und Kontrakti-

onsgrad aufweist. Die Abbildung 3.3 zeigt am Beispiel des zyklischen  $S_6$ -Allotropes die gute Korrelation zwischen der Rechenzeit und der Anzahl an Gleitkomma-Operationen pro Integralklasse (gemittelt für 50 Bildungen der Austauschmatrix). In der derzeitigen Umsetzung wird ein Integral daher nur gespeichert, wenn es eine Minimalzahl  $\vartheta_{\text{flop}}$  an Gleitoperationen übersteigt. Die Minimalzahl  $\vartheta_{\text{flop}}$  gibt der Benutzer dabei nicht explizit an, sondern sie ergibt sich aus dem Größenkriterium  $\vartheta_{\text{size}}$  und den verfügbaren Speicherkapazitäten (wenn insgesamt genug Speicher verfügbar ist, findet also keine zusätzliche Auswahl aufgrund von  $\vartheta_{\text{flop}}$  statt).

Sobald die genaue Anzahl der zu speichernden Zweielektronen-Integrale feststeht, kann der gesamte Speicher zur Integralspeicherung freigegeben werden. In der anschließenden (unveränderten) Berechnung werden diejenigen Zweielektronen-Integrale hintereinander in den Integralspeicher abgelegt, die zum gleichen Schalenquartett gehören und der *Offset* jedes gespeicherten Schalenquartetts dann zusätzlich gespeichert (eine Schale bezeichnet den Satz an Integralen mit den gleichen Atomzentren und Exponenten, aber unterschiedlichen Drehimpulsen). Der Speicher für die *Offset*-Werte ist dabei so organisiert, dass er potentiell nur linear mit der Systemgröße anwächst.

Das veränderte Integralscreening in den folgenden Iterationen entscheidet darüber, welche Integrale gemäß des Differenzdichte-basierten Schwarz-Screenings signifikant sind, und welche davon wiedereingelesen werden können oder zuerst berechnet werden müssen. Innerhalb der vorgestellten Methode können also zwei Gründe dafür verantwortlich sein, dass Integrale in späteren Iterationen neu berechnet werden müssen: Erstens die Änderungen der Differenzdichte-Matrix und damit der Einfluss auf das aktuelle Dichte-basierte Schwarz-Screening und zweitens die zusätzlichen Auswahlkriterien  $\vartheta_{\text{size}}$  und  $\vartheta_{\text{flop}}$ . Die zusätzlich benötigten Integrale werden schließlich wie in einem gewöhnlichen direkten Verfahren berechnet und kontrahiert.

### 3.2.2 Abhängigkeiten der Geschwindigkeitszuwächse

Die Abbildung 3.2 zeigt die maximalen theoretischen Geschwindigkeitszuwächse eines Integralspeicher-Verfahrens verglichen mit dem direkten Verfahren, unter der Annahme, dass das Integralspeicher-Verfahren die gesamte Zeit der Integralberechnung einsparen könnte und keinerlei zusätzliche Zeit, z.B. für den Speicherzugriff, investiert werden müsste. Die maximalen theoretischen Geschwindigkeitszuwächse sind über verschiedenen Prozentsätzen aufgetragen,

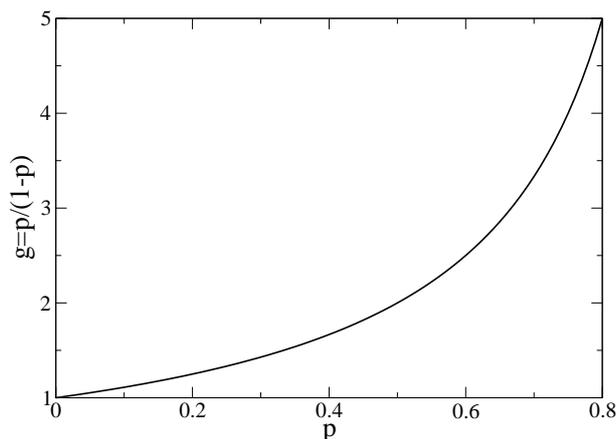


Abb. 3.2: Maximaler theoretischer Geschwindigkeitszuwachs  $g = t_{dir}/t_{semdir} = \frac{1}{1-p}$  für die Bildung der HF-Austauschmatrix abhängig vom Prozentsatz der Integralberechnung an der Austauschzeit.

die das Programm für die Integralberechnung gemessen an der gesamten Austauschzeit ausmacht. Wenn der Prozentsatz also beispielsweise 66.7 % (zwei Drittel) betragen würde, wäre die Integralspeicher-Methode bestenfalls dreimal so schnell bei der Bildung der Austauschmatrix wie das direkte Verfahren. Natürlich macht sich in der tatsächlichen Rechenzeit der Speicherzugriff bemerkbar und, noch wesentlicher, verbleibt die Rechenzeit der Integrale in der zweiten Iteration. Zum Vergleich führt die Tabelle 3.1 die CPU-Zeiten für die einzelnen Bestandteile der Integralspeicher-Methode und der direkten Methode auf, für das Beispiel eines Basenpaares Desoxyadenosin und Desoxythymidin DNA(A-T)<sub>1</sub> [61]. Die Zeitmessungen wurden seriell (auf einem *Core*) eines Intel-Xeon-Rechners (2x CPU E5420, quadcore, 64 GB RAM) durchgeführt. Für die Integralspeicher-Methode ist die Zeit für die Auswahl, die Speicherung und das Einlesen der Integrale angegeben, zusätzlich zur Screening-, Rechen-, und Kontraktionszeit. Die Integraalauswahl basiert auf der zweiten Dichtematrix. Für die Iterationen drei bis acht ist die Gesamtzeit für die jeweiligen Bestandteile aufgeführt (jeweils werden die Fockmatrizen inkrementell gebildet). Neben der spezifischen Iterationszeit geht aus der Tabelle auch die Gesamtzeit für das SCF hervor. Der angegebene Prozentsatz bezieht sich jeweils auf die gesamte Austauschzeit innerhalb der Integral-Speichermethode bzw. der direkten Methode. In der direkten Methode nimmt die Integralrechenzeit ungefähr 65.5 % der gesamten Austauschzeit ein. Natürlich kann die Integralspeicher-Methode nicht die Integralrechenzeit in der zweiten

Tab. 3.1: CPU-Zeit  $t$  (in Sekunden) und Prozentsatz  $p$  verschiedener Teilschritte bei der Bildung der Austauschmatrix im Integral-Speicherverfahren und der direkten SCF-Methode für DNA(A-T)<sub>1</sub> und einer SVP-Basis (Integralgrenzwert  $\vartheta_{\text{Int}} = 10^{-8}$ ; Konvergenzkriterium  $\vartheta_{\text{SCF}} = 10^{-5}$ ).

Ite.	Schritt	Integralspeicher-Methode		direkte Methode	
		t [s]	(p [%])	t [s]	(p [%])
1	gesamt	5	(2.6)	5	(1.7)
2	Screening <sup>a</sup>	3	(1.5)	3	(1.0)
	Berechnung <sup>b</sup>	40	(20.6)	40	(13.6)
	Kontraktion <sup>c</sup>	17	(8.8)	17	(5.6)
	Auswahl <sup>d</sup>	3	(0.1)	-	-
	Speicherung <sup>e</sup>	9	(4.6)	-	-
	gesamt	73	(37.6)	62	(20.5)
3-8	Screening <sup>a</sup>	16	(8.2)	12	(4.0)
	Berechnung <sup>b</sup>	1	(0.5)	156	(51.9)
	Kontraktion <sup>c</sup>	0	(0.2)	60	(19.9)
	Einl.+Kontr. <sup>f</sup>	96	(49.5)	-	-
	gesamt	116	(59.8)	234	(77.7)
total		194	(100)	301	(100)

<sup>a</sup>Dichte-basiertes Schwarz-Screening (ggf. verändert)

<sup>b</sup>Berechnung der Zweielektronen-Integrale

<sup>c</sup>Kontraktion der Zweielektronen-Integrale mit der Dichtematrix

<sup>d</sup>Lauf zur Auswahl der teuren und großen Zweielektronen-Integrale

<sup>e</sup>Speicherung der ausgewählten Zweielektronen-Integrale

<sup>f</sup>Einlesen und Kontraktion der Zweielektronen-Integrale

Tab. 3.2: a) Einfluss des Integral-Grenzwerts  $\vartheta_{\text{Int}}$  und b) des SCF-Konvergenzkriteriums  $\vartheta_{\text{SCF}}$  auf die Geschwindigkeitszuwächse  $g = t_{\text{dir.}}/t_{\text{sem.}}$  zur Bildung des Austauschteils für das DNA(A-T)<sub>1</sub>-System. Für a) wurde durchgehend  $\vartheta_{\text{SCF}} = 10^{-4}$  verwendet und für b)  $\vartheta_{\text{Int}} = 10^{-10}$ .

a)	$-\log(\vartheta_{\text{Int}})$	6	7	8	9	10
	Zuwachs	1.36	1.43	1.47	1.49	1.52
b)	$-\log(\vartheta_{\text{SCF}})$	3	4	5	6	7
	Zuwachs	1.35	1.52	1.60	1.65	1.68

Iteration einsparen, sondern nur diejenige der Einleseiterationen drei bis acht (51.9 %). Daher beträgt der bestmögliche Geschwindigkeitszuwachs für dieses Beispiel  $g = 2.08$ . Die Annahme, dass die Zeit für die Berechnung der zusätzlichen Integrale in den Iterationen drei bis acht vernachlässigbar ist, erweist sich hier als richtig: Die Zeit sinkt auf ungefähr 0.5 % der Austauschzeit und daraus zieht die Integralspeicher-Methode natürlich ihre Geschwindigkeitszuwächse. Auf der anderen Seite muss zusätzliche Zeit für die Integraalauswahl, -speicherung, das veränderte Screening und das Einlesen der Integrale verwendet werden, was den effektiven Geschwindigkeitszuwachs auf 1.51 im Vergleich zu dem bestmöglichen Zuwachs von 2.08 reduziert.

Der effektive Geschwindigkeitszuwachs der Integralspeicher-Methode hängt von verschiedenen Aspekten ab, die im Folgenden betrachtet werden. Die Tabelle 3.2 zeigt die Abhängigkeit vom Integralgrenzwert und SCF-Konvergenzkriterium, wieder am Beispiel von DNA(A-T)<sub>1</sub>. Für veränderte Integralgrenzwerte wurde ein konstantes SCF-Konvergenzkriterium verwendet (a) und umgekehrt (b). Die Geschwindigkeitszuwächse steigen, je strenger die Integralgrenzwerte sind, weil sich damit das Rechenzeit-Verhältnis zwischen Screening, Kontraktion und Integralberechnung in Richtung Integralberechnung verschiebt. Da man mithilfe der Methode nur ab der dritten Iteration Rechenzeit einsparen kann, steigt der Geschwindigkeitszuwachs zudem, je strenger das SCF-Konvergenzkriterium ist und je mehr SCF-Iterationen deshalb notwendig sind.

Darüber hinaus hängen die Geschwindigkeitszuwächse vom Basissatz und dem molekularen System ab. Wiederum steigen die Zuwächse, wenn die Integralberechnung in den späteren

Tab. 3.3: Geschwindigkeitszuwächse bei der Bildung des Austauschteils bei verschiedenen molekularen Systemen und Basissätzen für zwei verschiedene Grenzwertsätze. Zusätzlich ist der Energieunterschied zwischen den strengen (7/10) und weniger strengen (5/8) Grenzwerten angegeben.

Molekül	Basissatz	Zuwachs		<sup>b</sup> $\Delta E[\mu\text{H}]$
		<sup>a</sup> 5/8	<sup>a</sup> 7/10	
Alkane C <sub>20</sub> H <sub>42</sub>	6-31G*	1.39	1.51	7
DNA (AT) <sub>1</sub>	SVP	1.55	1.68	2
Peptide Gly <sub>6</sub>	TZP	1.59	1.72	20
Kohlenhydrate Glu <sub>2</sub>	cc-pVDZ	1.83	2.07	15
Perfluorierte Silane Si <sub>6</sub> F <sub>18</sub>	cc-pVTZ	2.10	2.39	70
Schwefelring S <sub>6</sub>	cc-pVQZ	2.79	3.70	13

<sup>a</sup>  $\alpha/\beta$  steht für ein SCF-Konvergenzkriterium  $\vartheta_{\text{SCF}} = 10^{-\alpha}$  und einen Integralgrenzwert  $\vartheta_{\text{Int}} = 10^{-\beta}$

<sup>b</sup> Absolute Energiedifferenz zwischen der Verwendung von 7/10- und 5/8-Grenzwerten

SCF-Iterationen teurer sind, verglichen mit den Screening- und Kontraktionsschritten. Deshalb erreicht die Methode die größten Zuwächse für schwere Atome und aufwändige Basissätze. Die Tabelle 3.3 listet einige Beispielrechnungen mit zunehmendem Gewinn der neuen Integralspeicher-Methode gegenüber dem direkten SCF auf. Für das Beispiel des cyclischen S<sub>6</sub>-Moleküls, einem Konvergenzkriterium von  $\vartheta_{\text{SCF}} = 10^{-7}$  und Integralgrenzwert von  $\vartheta_{\text{Int}} = 10^{-10}$  erreicht die Integralberechnung in den Einleseiterationen 78 % der gesamten Austauschzeit und der größte beobachtete Geschwindigkeitszuwachs von 3.70 wird so erreicht (der maximale Geschwindigkeitszuwachs gemäß den Annahmen in Abbildung 3.2 beträgt hier 4.55).

Die bislang beschriebenen Fälle verwenden alle eine Integralauswahl basierend auf der ersten berechneten Dichtematrix (die zweite Dichtematrix), ohne dass zusätzlich die Kriterien  $\vartheta_{\text{size}}$  und  $\vartheta_{\text{flop}}$  eingesetzt wurden, d.h. 100 % der durch die erste berechnete Dichtematrix ausgewählten Integrale wurden gespeichert. Allerdings können für große Systeme wegen Speicher-

Tab. 3.4: Prozentsatz gespeicherter Zweielektronen-Integrale in der zweiten Iteration und zusätzlich berechnete Integrale in späteren Iterationen für DNA(A-T)<sub>1</sub>, Basissatz TZP, Konvergenzkriterium  $\vartheta_{\text{SCF}} = 10^{-5}$ , Integralgrenzwert  $\vartheta_{\text{int}} = 10^{-8}$  und Integralspeicher-Grenzwert  $\vartheta_{\text{size}}$ . Der Prozentsatz bezieht sich auf die Anzahl an berechneten Integralen in der zweiten SCF-Iteration (4347.95 Millionen). In allen Fällen wird die Fockmatrix inkrementell gebildet.

	abspeichern	berechnen			
	iter. 2	iter. 3	ite. 4	iter. 5	iter. 6
direktes SCF	0.00	92.00	81.93	68.37	54.98
kein $\vartheta_{\text{size}}$	100.00	0.76	0.13	0.00	0.00
$\vartheta_{\text{size}} = 10^{-4}$	68.43	24.98	17.65	9.25	3.46

beschränkungen die zusätzlichen Kriterien  $\vartheta_{\text{size}}$  und  $\vartheta_{\text{flop}}$  wesentlich sein und ihr Einfluss auf die beschriebene Integralspeicher-Methode wird im Folgenden betrachtet, begonnen mit dem Effekt des Kriteriums  $\vartheta_{\text{size}}$ . Am Beispiel des Systems DNA(A-T)<sub>1</sub> und einer TZP-Basis analysiert die Tabelle 3.4 drei verschiedene Fälle:

- die direkte Methode (erster Fall) speichert keine Integrale
- beim zweiten Fall werden alle Integrale gespeichert, wie sie durch die erste berechnete Dichtematrix ausgewählt werden
- der dritte Fall schränkt die Integraalauswahl des zweiten Falls weiter ein, indem nur Integrale mit Schwarz-Schranken größer als  $\vartheta_{\text{size}} = 10^{-4}$  gespeichert werden.

Die Ergebnisse in Tabelle 3.4 zeigen, dass im dritten Fall die Anzahl benötigter Zweielektronen-Integralen in späteren Iterationen stark erhöht ist: Zum Beispiel müssen in der dritten SCF-Iteration 24.98 % anstatt von 0.76 % zusätzlich berechnet werden. Die Speicheranforderungen sinken dabei zwar von 32.4 GB auf 22.2 GB (68.43 %). Allerdings sinkt als Konsequenz der zusätzlich benötigten Integrale der Geschwindigkeitszuwachs auch deutlich erkennbar von 1.55 auf 1.40.

Andere geeignete Beispielrechnungen (siehe Abbildung 3.3) an einem zyklisches  $S_6$ -System betrachten den Einfluss des zweiten zusätzlichen Speicherkriteriums  $\vartheta_{\text{flop}}$  (cc-pVQZ-Basis, Integralgrenzwert  $\vartheta_{\text{int}} = 10^{-8}$ , Konvergenzkriterium  $\vartheta_{\text{int}} = 10^{-5}$ ). Die Auswahl nach der An-

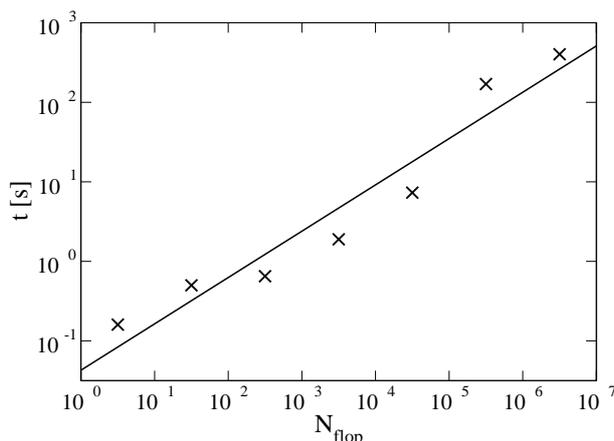


Abb. 3.3: Korrelation zwischen der CPU-Zeit  $t$  [s] der Integralberechnung und der Anzahl der benötigten Gleitkomma-Operationen  $N_{\text{flop}}$  (bei der Bildung des Austauschteils innerhalb der zweiten SCF-Iteration, gemittelt über 50 Durchläufe) für das zyklische  $S_6$ -Molekül und aug-cc-pVTZ-Basis. Die Zeiten für Integralklassen mit 1–10, 10–100, . . . , Gleitkomma-Operationen wurden gemittelt und nur ein Datenpunkt eingezeichnet.

zahl der Gleitkomma-Operationen pro Integral  $\vartheta_{\text{flop}}$  erweist sich als sehr effizient: Sie führt zu einem fast konstanten Geschwindigkeitszuwachs von 2.70 gegenüber 2.79, wenn alle Integrale abgespeichert werden. Zugleich reduzieren sich allerdings die Speicheranforderungen von 100 % (31.02 GB) auf 61.76 % (19.16 GB). Demgegenüber zeigt ein anderes Kriteriums, die Auswahl gemäß des Produkts von Gesamtdrehimpuls und Kontraktionsgrad, einen Zuwachs von 2.53 (bei ungefähr gleicher Reduktion der Speicheranforderungen). Wer während der Integralberechnung einfach so lange Integrale in ihrer Standardreihenfolge speichert, bis die Speichergrenze überschritten ist, erreicht in diesem Fall nur einen Zuwachs von 1.44.

### 3.2.3 Untersuchungen des Skalenverhaltens

Um das Skalenverhalten der neuen Integralspeicher-Methode exemplarisch zu untersuchen, werden zunächst drei Modellsysteme betrachtet: Lineare Alkane, helikale Oligopeptide und Amylosefragmente, die jeweils systematisch vergrößert werden können. Die Tabelle 3.5 stellt die Ergebnisse der neuen Integralspeicher-Methode zusammen und vergleicht sie mit einer konventionellen semidirekten Methode und der direkten Methode. Neben der CPU-Zeit zur Bildung des Austauschteils und dem Zeit-Skalenverhalten führt die Tabelle für die beiden Integralspeichermethoden auch die Speicheranforderungen und das Skalenverhalten des Speicherbe-

Tab. 3.5: Gesamte CPU-Zeit  $t$  (in Sekunden) zur Bildung des Austauschteils, Speicheranforderungen  $m$  (in GB) und Skalungsverhalten ( $n^x$ ) für lineare Alkane  $\text{Alk}_n$  ( $n$  C Atome), helikale Peptide  $\text{Gly}_n$  ( $n$  Glycin-Einheiten) und Amylose-Fragmente  $\text{Glu}_n$  ( $n$  Glucose-Einheiten) (Basissatz 6-31G\*, Konvergenzkriterium  $\vartheta_{\text{SCF}} = 10^{-5}$ , Integralgrenzwert  $\vartheta_{\text{Int}} = 10^{-7}$ ). Die vorgestellte Integral-Speichermethode (P-Screening) wird mit der direkten SCF-Methode (mit LinK-Screening) und einer konventionellen semidirekten Methode (kein P-Screening) verglichen, die keine Dichtematrix beim Integralscreening miteinbezieht, sondern Größen- und Zeitabschätzungen (siehe Text). Für die Abhängigkeit der Geschwindigkeitszuwächse vom Basissatz, den Integralgrenzwerten und Konvergenzkriterien siehe die Tabelle 3.3 (z.B werden für größere Basissätze größere Geschwindigkeitszuwächse erzielt).

molecule	$^a(N_{bas})$	Integralspeicher-Methode								direkt (LinK)	
		kein P-Screening				P-Screening (LinK)				t	$^b(x_t)$
		t	$^b(x_t)$	m	$^c(x_m)$	t	$^b(x_t)$	m	$^c(x_m)$		
$\text{Alk}_{10}$	(194)	9		0.14		6		0.36		10	
$\text{Alk}_{20}$	(384)	34	(1.95)	0.62	(2.18)	23	(1.97)	1.44	(2.03)	30	(1.60)
$\text{Alk}_{40}$	(764)	108	(1.68)	2.84	(2.21)	59	(1.37)	4.20	(1.56)	80	(1.43)
$\text{Alk}_{80}$	(1524)	344	(1.68)	11.66	(2.05)	145	(1.30)	9.81	(1.23)	191	(1.26)
$\text{Alk}_{160}$	(3044)	1136	(1.73)	47.26	(2.02)	343	(1.24)	21.03	(1.10)	421	(1.14)
$\text{Gly}_3$	(221)	19		0.25		13		0.71		20	
$\text{Gly}_6$	(419)	72	(2.08)	1.14	(2.37)	49	(2.07)	3.04	(2.27)	68	(1.91)
$\text{Gly}_{12}$	(815)	263	(1.95)	5.31	(2.31)	166	(1.83)	11.21	(1.96)	214	(1.72)
$\text{Gly}_{24}$	(1607)	835	(1.70)	23.00	(2.16)	416	(1.35)	29.36	(1.42)	520	(1.31)
$\text{Glu}_1$	(204)	18		0.27		13		0.73		20	
$\text{Glu}_2$	(389)	72	(2.15)	1.23	(2.35)	49	(2.06)	3.37	(2.37)	67	(1.87)
$\text{Glu}_4$	(759)	235	(1.77)	5.33	(2.19)	147	(1.64)	11.23	(1.80)	187	(1.54)
$\text{Glu}_8$	(1499)	747	(1.70)	21.67	(2.06)	375	(1.38)	27.99	(1.34)	476	(1.37)

<sup>a</sup> Anzahl an Basisfunktionen

<sup>b</sup> Skalierungs-Exponent der Gesamtzeit zur Bildung des Austauschteils

<sup>c</sup> Skalierungs-Exponent der Speicheranforderung

darfs auf. Die Speicherung der neuen Integralspeicher-Methode basiert wiederum auf den Schwarz-Schätzern und der ersten berechneten Dichtematrix. Demgegenüber wählt die konventionelle semidirekte Methode die Integrale durch Schwarz-Schranken vor dem SCF aus und bezieht also keine Dichtematrix in das Screening mit ein. Da eine solche Auswahl schnell zu einem Überschreiten der Speicherkapazität führt, z.B. würden 152.49 GB für  $\text{Alk}_{160}$  benötigt, werden für die konventionelle semidirekte Methode zusätzlich Größen- und Zeitkriterien verwendet. Die Kriterien werden über alle Molekülgrößen konstant gehalten und sind so gewählt, dass die maximalen Speicheranforderungen für die größten Moleküle erfüllt werden können ( $\vartheta_{\text{size}} = 10^{-4}$  und gemäß  $\vartheta_{\text{flop}}$  wird die aufwändigere Hälfte der Zweielektronen-Integrale gespeichert).

Der Vergleich der Gesamtzeiten zeigt, dass sich hier das lineare Skalenverhalten der direkten Methode schon für das zweitkleinste Molekül innerhalb der Serien auszahlt. Für größere Moleküle wird die konventionelle semidirekte Methode immer ineffizienter bezüglich der direkten Methode, da die Berechnung der Integrale vor dem SCF quadratisch skaliert. Dieses Problem ist in der neuen Integral-Speichermethode gelöst: das günstige Skalenverhalten der direkten Methode wird beibehalten. Das lineare Skalenverhalten ist dabei nicht nur für die Rechenzeiten erfüllt, sondern auch für die Speicheranforderungen der neuen Integral-Speichermethode.

Für die Beispiele in der Tabelle 3.5 sind die Geschwindigkeitszuwächse durch die vorgestellte Integralspeicher-Methode nicht so ausgeprägt, wie beispielsweise für größere Basissätze (höhere maximale Kontraktionsgrade und Drehimpulstypen) oder Atome tieferer Perioden (siehe Tabelle 3.3). Gleichzeitig verschiebt sich allerdings mit größeren Basissätzen und Atomen tieferer Perioden der Einsatz des linearen Skalenverhaltens, für die direkte Methode ebenso wie für die vorgestellte Integralspeicher-Methode. Um den prinzipiellen Vorteil der vorgestellten Integralspeicher-Methode für ein weiteres Beispiel zu verdeutlichen, sind ergänzende Rechnungen auf einem Workstation-Computer (Intel Xeon E7-4820, 2.00 GHz) mit insgesamt 500 GB RAM-Speicher an Kettensilikaten durchgeführt worden. Die Ergebnisse fasst die Abbildung 3.4 zusammen. Für die konventionelle semidirekte Methoden wurde bei den Rechnungen des Kettensilikats für alle Systemgrößen konstant die Kriterien  $\vartheta_{\text{size}} = 10^{-6}$  gewählt und gemäß  $\vartheta_{\text{flop}}$  die aufwändigsten zwei Drittel der Zweielektronenintegrale gespeichert. Für den gezeigten Fall bestätigt sich das günstige Skalenverhalten der vorgestellten Integralspeicher-Methode

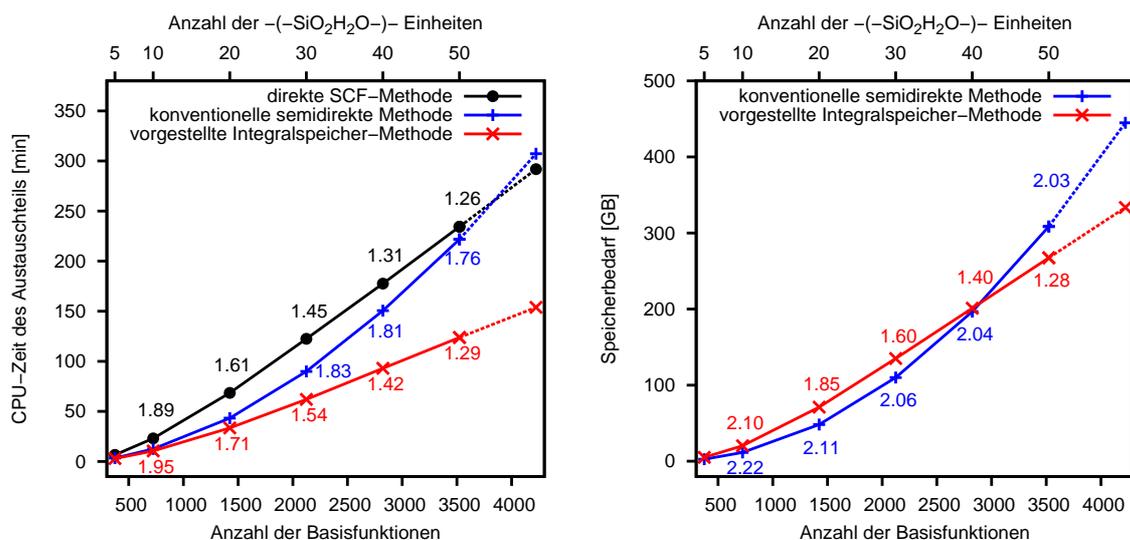


Abb. 3.4: Gesamte CPU-Zeit  $t$  (in Minuten) zur Bildung des Austauschteils (links) und Speicherbedarf (rechts) zur Bildung des Austauschteils (in GB) und Skalenerhalten für Kettensilikate mit  $n$   $-(\text{SiO}_2\text{H}_2\text{O})-$ -Einheiten, Basissatz cc-pVDZ, Konvergenzkriterium  $\vartheta_{\text{SCF}} = 10^{-7}$ , Integralgrenzwert  $\vartheta_{\text{Int}} = 10^{-10}$ . Die vorgestellte Integral-Speichermethode (P-Screening) wird mit einer konventionellen semidirekten Methode (kein P-Screening) verglichen, die keine Dichtematrix beim Integralscreening miteinbezieht, sondern Größen- und Zeitabschätzungen (siehe Text). Die letzten Punkte wurden basierend auf dem entsprechenden Skalenerhalten extrapoliert.

in Bezug auf die Rechenzeit und die Speicheranforderungen (Skalierungs-Exponent sinkt auf etwa 1.3). Zugleich zeigt sich hier auch der ausgeprägte Vorteil gegenüber der konventionellen direkten SCF-Methode (Geschwindigkeitszuwachs etwa 1.8).

### 3.2.4 Kombination mit einer parallelen Integralberechnung

Die bislang vorgestellten Zeiten wurden alle für einen seriellen Programmablauf gemessen. Die Abbildung 3.5 zeigt basierend auf zusätzlichen Untersuchungen, inwieweit sich die Geschwindigkeitszuwächse der neuen Integralspeicher-Methode mit einer parallelen Berechnung des Zweielektronen-Teils der Fockmatrix kombinieren lässt. Für das Beispiel eines Schwefel<sub>6</sub>-Rings und einem cc-pVQZ-Basissatz [62] vergleicht die Abbildung die Zeitmessung eines seriellen (1 *thread*) und parallelen Programmablaufs (10 *threads*). Die Parallelisierung basiert auf einer kürzlichen Implementierung [63] im Programmpaket Q-Chem [64] mittels Open-Multi-Processing (OpenMP, [65]). Dargestellt sind die tatsächlichen Zeiten zur Bildung des

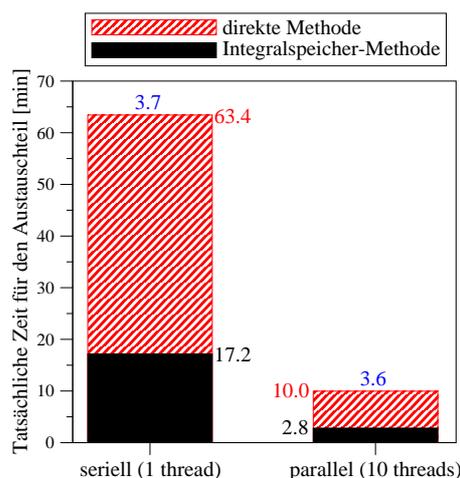


Abb. 3.5: Vergleich der tatsächlichen Zeit für den Austauschteil der direkten SCF-Methode und vorgestellten Integralspeicher-Methode im Falle einer seriellen Rechnung (1 *Thread*) und parallelen *Open-Multi-Processing*-Rechnung (10 *Threads*) für ein  $S_6$ -Ring, HF/cc-pVQZ; über den Balken sind die jeweiligen Geschwindigkeitszuwächse durch die Integralspeicher-Methode angegeben.

Austauschteils für ein direktes SCF-Verfahren und die neue Integralspeicher-Methode. Die Geschwindigkeitszuwächse für den seriellen Fall (3.7) lassen sich also gemäß den Ergebnissen sehr gut auf den parallelen Fall (3.6) übertragen: Die Geschwindigkeitszuwächse durch die Parallelisierung und Integralspeicher-Methode sind (nahezu) multiplikativ, für das gezeigte Beispiel ebenso wie auch für andere betrachtete Beispiele. So kann im vorliegenden Beispiel durch die Parallelisierung bei einem direkten SCF erreicht werden, dass der Austauschteil in 10 min statt in 63.4 min gebildet wird. Durch eine Kombination der Parallelisierung und der neuen Integralspeicher-Methode sinkt die tatsächliche Austausch-Rechenzeit dann auf 2.8 min. Die Kombination aus Parallelisierung und der neuen Integralspeicher-Methode kann gut auf derzeitigen Workstation-Computern eingesetzt werden. Die Zeitmessungen für den parallelen Programmablauf wurde mit 10 *Cores* auf einem Intel-Xeon(E5645)-Rechner (2.4 GHz, 96 GB, 12 *Cores*). durchgeführt.

### 3.3 Abstands-abhängiges Schwarz-Screening

Eine kürzlich veröffentlichte Arbeit von S. A. Maurer et al. [60] stellt eine neue Screening-methode für die bei Hartree-Fock- oder Kohn-Sham-DFT-Rechnungen massgebenden Zwei-

elektronen-Integrale vor. Die neue Methode trägt zu einer weiteren Effizienzsteigerung linear-skalierender SCF-Methoden (siehe Abschnitt 2.3.1) bei. Die nachstehende Darstellung beschränkt sich auf einen Abriss des Kerninhalts der neuen Methode und nennt die wesentlichen eigenen Beiträge zu der Veröffentlichung.

Das Schwarz-Screening (siehe Gleichung 2.34) bildet die Grundlage, um beim SCF-Verfahren den Anstieg mit der Systemgröße  $M$  des Rechenaufwands zur Bildung der Fock-Matrix (Kohn-Sham-Matrix) von formal  $M^4$  auf  $M^2$  zu reduzieren. Ein konventionelles Schwarz-Screening berücksichtigt zwar den exponentiellen Abfall innerhalb der Ladungsverteilung des Bra- bzw. Ket-Teil des Zweielektronen-Integrals (siehe Gleichungen 2.30 und 2.31). Hingegen berücksichtigt ein konventionelles Schwarz-Screening nicht, wie das Zweielektronen-Integral mit zunehmendem Abstand von Ladungsverteilung im Bra- und Ket-Teil aufgrund des  $\frac{1}{r_{12}}$ -Operators abfällt. Die neue Screeningmethode bezieht nun die Bra-Ket-Separation in die Screening-Gleichung mit ein:

$$|(\mu\nu|\kappa\lambda)| \approx \begin{cases} Q_{\mu\nu}Q_{\kappa\lambda}, & R - \text{ext}_{\mu\nu} - \text{ext}_{\kappa\lambda} \leq 1 \\ \frac{Q_{\mu\nu}Q_{\kappa\lambda}}{R - \text{ext}'_{\mu\nu} - \text{ext}'_{\kappa\lambda}}, & R - \text{ext}_{\mu\nu} - \text{ext}_{\kappa\lambda} > 1 \end{cases} \quad (3.1)$$

Die Ausdehnungen (*extents*)  $\text{ext}_{\mu\nu}$  und  $\text{ext}_{\kappa\lambda}$  für kontrahierte Ladungsverteilungen basieren dabei auf den Wohlsepariertheitskriterien für primitive Ladungsverteilungen der CFMM-Theorie [24, 60] und  $R$  steht für den Abstand von kontrahierten Zentren der Ladungsverteilungen für den Bra- und Ket-Teil. Das von Maurer et al. [60] vorgestellte neue Screeningverfahren zeigt sich für eine ausgiebige Reihe an Testrechnungen als überlegen gegenüber einer früheren, Multipol-basierten Screeningmethode [66, 67]. Mit der neuen Methode sinkt, ohne signifikanten Verlust an Genauigkeit, die benötigte Anzahl an Zweielektronen-Integralen für den Austauschteil und den betrachteten Systemen um Faktoren bis hin zu 2.3 (System aus 569 Wassermolekülen, SV(P)-Basissatz, Screening-Grenzwert  $10^{-8}$ ).

Für die Testrechnungen wurde im Rahmen dieser Dissertation ein neuer umfangreicher molekularer Testsatz erstellt. Der molekulare Testsatz umfasst ein großes Spektrum von metallischen bis zu nicht-metallischen, neutralen, ionischen und radikalischen Systemen, mit einer eindimensionalen bis hin zu dreidimensionalen Struktur mit Hauptgruppenelementen bis zur dritten Periode und einem Kupferkomplex (siehe Anhang B).

### 3.4 Zusammenfassung und Ausblick

Mit den vorgestellten Arbeiten wurde eine neue Integral-Speichermethode eingeführt, die Vorzüge semidirekter (bzw. indirekter) SCF-Methoden mit denen linear-skalierender direkter Methoden kombiniert. Der Schlüsselschritt besteht darin, mithilfe der ersten berechneten Einteilchen-Dichtematrix diejenigen Integrale auszuwählen, die für die späteren Iterationen gespeichert werden sollen. Es zeigt sich, dass die erste berechnete Einteilchen-Dichtematrix (die der zeitgünstigen Iteration einer SAD-Anfangsschätzung folgt) ausreichend genau ist, um die Integrale so auszuwählen, dass nur wenige Integrale in späteren Iterationen neu berechnet werden müssen. Die Geschwindigkeitszuwächse hängen vom molekularen System, dem Basissatz und den gewählten Grenzwerten ab, wobei für Austausch-artige Kontraktionen Beschleunigungen von typischerweise 1.5–3.7 beobachtet werden, recht nah an den abgeschätzten optimalen Geschwindigkeitszuwächsen. Das Verfahren ist nicht nur nützlich zur Berechnung von Energien, sondern insbesondere auch bei der Berechnung von Moleküleigenschaften, wo in der Regel strenge Grenzwerte benötigt werden. Mit der Verfügbarkeit von Computern mit mehr Hauptspeicher werden die Möglichkeiten zunehmen, Integrale abzuspeichern und ihre mehrmalige Berechnung zu umgehen, was die Nützlichkeit der vorgestellten Methode unterstreicht.

Mittels des von S. A. Maurer et. al. [60] vorgestellten neuen QQR-Screenings für Zweielektronen-Integrale kann der Abfall abhängig vom Abstand der Ladungsverteilungen für den Bra-Ket-Abstand gewinnbringend in das Integralscreening miteinbezogen werden. Zum Testen des entwickelten neuen Screening-Verfahrens und ebenso für zukünftige methodische Entwicklungen, wurde im Rahmen dieser Dissertation ein umfangreicher molekularer Testsatz entwickelt.



## Kapitel 4

# QM-Ansätze zur effizienten Beschreibung der Molekulardynamik

### 4.1 Einführung

Molekulardynamische Verfahren ermöglichen den Raum relevanter Strukturen eines molekularen Systems effizient abzutasten (siehe Abschnitt 2.2). Einerseits können MD-Verfahren verwendet werden, um ausgewiesene lokale Minimumstrukturen oder auch die globale Minimumstruktur zu finden. Andererseits ermöglichen MD-Verfahren chemische Umwandlungen unmittelbar zu simulieren und zudem verschiedene Zustandsgrößen gemäß den Gesetzmäßigkeiten der statistischen Thermodynamik zu berechnen.

MD-Verfahren, bei denen sich die Atomkerne auf der exakten Born-Oppenheimer-Hyperfläche bewegen (Born-Oppenheimer-MD, siehe Abschnitt 2.2), erfordern in jedem Zeitschritt die Berechnung der elektronischen Energie und der molekularen Kräfte (als negative Energiegradienten). Um die elektronische Schrödingergleichung in jedem Schritt zu lösen, muss prinzipiell in jedem Zeitschritt ein vollständiges SCF-Verfahren durchlaufen werden (bzw. für genauere Lösungen noch zusätzlich darauf aufbauende Elektronen-Korrelationsverfahren).

Aufgrund des großen Rechenaufwands, den diese Beschreibung mit sich bringt, basieren viele derzeitige MD-Untersuchungen von großen molekularen Systemen hauptsächlich auf Kraftfeld-Näherungen (siehe Abschnitt 2.5). Um für große Systeme genauere Beschreibungen (als durch die Kraftfeld-Verfahren) zu ermöglichen, sind in den letzten Jahren Ab-initio-MD-Verfahren als Näherungen der Born-Oppenheimer-MD entwickelt worden, allen voran das Car-Parinello-MD-Verfahren (CPMD, [68]). Das CPMD-Verfahren propagiert näherungsweise die elektronische Wellenfunktion, anstatt sie jeweils zeitaufwändig iterativ mittels eines SCF-

Durchläufe neu zu bestimmen: Im Rahmen der Hartree-Fock-Theorie (oder Kohn-Sham-DFT) propagiert die CPMD die der elektronischen Wellenfunktion zugrunde liegenden Spinorbitale  $\varphi_i$  (siehe Gleichung 2.16) gemäß der Gleichung

$$\mu \frac{\partial^2 \varphi_i}{\partial t^2} = -\hat{F}\varphi_i + \sum_j \epsilon_{ji} \varphi_j, \quad (4.1)$$

mit der fiktiven Elektronenmasse  $\mu$ , dem Fockoperator  $\hat{F}$  und den Lagrange-Multiplikatoren  $\epsilon_{ji}$  (vgl. die allgemeinen Hartree-Fock-Gleichungen 2.21). Aufgrund der numerischen Integration der elektronischen Propagationsgleichung 4.1 muss man den Zeitschritt bei einer Car-Parinello-MD typischerweise um eine Größenordnung kleiner als bei der Born-Oppenheimer-MD wählen, was die Gesamteffizienz der Car-Parinello-Methode senkt. Die Näherung der CPMD hängt zudem maßgeblich von der Wahl der fiktiven Elektronenmasse  $\mu$  ab [69–71]. In den letzten Jahren wurde daher vor allem intensiv an Weiterentwicklungen der Born-Oppenheimer-MD gearbeitet: Einen Fortschritt zur Effizienzsteigerung erbrachte die Methode der Fockmatrix-Extrapolation [72, 73], mit der die SCF-Durchläufe für die einzelnen MD-Schritte beschleunigt werden können. Die dabei auftretenden Probleme aufgrund einer signifikanten Fehlerfortpflanzung für lange Simulationsdauern (z.B. in Hinblick auf die Energieerhaltung) können weitestgehend mit Fortentwicklungen von Niklasson et al. überwunden werden [74–78].

Die Ansätze zur Fockmatrix-Extrapolation streben gute Anfangsschätzungen für das SCF-Verfahren an, was sich in einer Einsparung von SCF-Iterationen niederschlägt. Nichtsdestotrotz benötigen die Extrapolationsverfahren mindestens zwei, typischerweise drei bis sechs SCF-Iterationen für eine ausreichende Genauigkeit, abhängig vom molekularen System, Basissatz etc. Jedoch sind gerade die frühen, nicht einzusparenden SCF-Iterationen die zeitaufwändigsten bei einem SCF-Durchlauf, da in den nachfolgenden Iterationen ein Differenzdichte-Screening genutzt werden kann (siehe Gleichungen 2.37 und 2.36). Aus diesem Grund wird in Abschnitt 4.3 ein ergänzender Ansatz untersucht, um die Kosten zur Fock-Matrix-Bildung insbesondere in den frühen SCF-Iterationen zu senken. Der neue Ansatz besteht darin, die Zweielektronen-Integrale aus vergangenen Zeitschritten beim aktuellen Zeitschritt teilweise wiederzuverwenden, basierend auf Abschätzungen ihrer Änderung durch die Geometrieänderung. Der ergänzende Ansatz der Integralwiederverwendung kann dabei auf die im vorangehenden Kapitel vorgestellte SCF-Integralspeicher-Methode zurückgreifen.

Prinzipiell bieten außerdem QM/MM-Ansätze für die Berechnung molekularer Kräfte eine

weitere Möglichkeit, um den Rechenaufwand der MD maßgeblich zu senken und gleichzeitig eine zentrale Molekülregion möglichst genau (quantenchemisch) zu beschreiben (siehe Abschnitt 2.6). Der Abschnitt 4.4 stellt in diesem Zusammenhang die Grundzüge einer QM/MM-Implementierung vor, die linear-skalierende QM-Methoden im Rahmen des Programmpakets Q-Chem [64] mit den Funktionalitäten des ChemShell-Programmpakets [79] kombiniert.

Um während der Born-Oppenheimer-MD molekulare Eigenschaften zweiter Ordnung (z.B. NMR-Abschirmungen) zu berechnen, muss man neben den Energien und molekularen Kräften außerdem die gestörte Einteilchen-Dichtematrix bestimmen (siehe Abschnitt 2.4.2). Die Berechnung der Moleküleigenschaften in allen einzelnen (oder zumindest ausgewählten) Schritten der MD-Trajektorien ermöglicht es beispielsweise, explizite Schwingung- und Rotationskorrekturen sowie Temperatureinflüsse zu berechnen, oder auch Veränderungen der Moleküleigenschaften bei chemischen Umwandlungen zu simulieren (siehe z.B. Referenzen [80–83]). Um das CPSCF-Verfahren in aufeinanderfolgenden MD-Schritten zu beschleunigen, wird im Rahmen dieser Arbeit ein Extrapolations-Verfahren vorgestellt, mit dem sich CPSCF-Iterationen einsparen lassen, in analoger Weise wie beim SCF-Verfahren Iterationen durch die Methode der Fockmatrix-Extrapolation eingespart werden können.

Zusammenfassend stellen die folgenden Abschnitte also drei sich ergänzende Ansätze zur effizienten Beschreibung konformationeller Änderungen im Rahmen der Born-Oppenheimer-MD vor:

- Extrapolationsmethode für CPSCF-Anfangsschätzungen (Abschnitt 4.2)
- Integralwiederverwendung für veränderte Konformationen (Abschnitt 4.3)
- QM/MM-Implementierung für Molekulardynamik (Abschnitt 4.4)

## 4.2 Extrapolationsmethode für CPSCF-Anfangsschätzungen

Die von Pulay und Fogarasi [72] vorgeschlagene Methode der Fock-Matrix-Extrapolation gründet auf einer Polynomentwicklung  $M$ -ter Ordnung der Fockmatrizen  $\mathbf{F}(n)$  für  $N$  vorangegan-

gene Born-Oppenheimer-MD-Schritte  $n$ :

$$\begin{aligned} \mathbf{F}(1) &= \mathbf{K}_0 + 1 \cdot \mathbf{K}_1 + 1 \cdot \mathbf{K}_2 + \cdots + 1 \cdot \mathbf{K}_M \\ \mathbf{F}(2) &= \mathbf{K}_0 + 2 \cdot \mathbf{K}_1 + 4 \cdot \mathbf{K}_2 + \cdots + 2^M \cdot \mathbf{K}_M \\ &\vdots \\ \mathbf{F}(N) &= \mathbf{K}_0 + N \cdot \mathbf{K}_1 + N^2 \cdot \mathbf{K}_2 + \cdots + N^M \cdot \mathbf{K}_M \end{aligned} \quad (4.2)$$

Durch Lösung des linearen Gleichungssystem bzw. der zugehörigen Matrixgleichung

$$\mathbf{f} = \mathbf{X} \cdot \mathbf{k}, \quad (4.3)$$

mit  $\mathbf{f} = \begin{pmatrix} \mathbf{F}(1) \\ \vdots \\ \mathbf{F}(N) \end{pmatrix}$ ,  $\mathbf{k} = \begin{pmatrix} \mathbf{K}_0 \\ \vdots \\ \mathbf{K}_M \end{pmatrix}$  und  $X_{nm} = n^{m-1}$  ( $n = 1, 2, \dots, N; m = 1, 2, \dots, M + 1$ ) können die Koeffizienten-Matrizen  $\mathbf{K}_m$  bestimmt und dann die Fockmatrix  $\mathbf{F}(N + 1)$  der ersten SCF-Iteration für den Zeitschritt  $N + 1$  extrapoliert werden:

$$\mathbf{F}(N + 1) = \mathbf{K}_0 + (N + 1) \cdot \mathbf{K}_1 + (N + 1)^2 \cdot \mathbf{K}_2 + \cdots + (N + 1)^M \cdot \mathbf{K}_M \quad (4.4)$$

Zwar enthalten die Koeffizienten-Matrizen  $\mathbf{K}_m$  individuelle Koeffizienten für alle Elemente  $F_{\mu\nu}$  der Fockmatrix  $\mathbf{F}$ ; die Matrix  $\mathbf{X}$  ist allerdings unabhängig von den Indizes  $\mu$  und  $\nu$ . Deshalb muss das Gleichungssystem unabhängig von den Indizes  $\mu$  und  $\nu$  nur einmalig gelöst werden, wodurch das Extrapolationsverfahren sehr effizient wird.

Im Rahmen der vorliegenden Arbeit wird ein Verfahren zur Extrapolation für CPSCF-Anfangsschätzungen vorgestellt, das auf einer analogen polynomiellen Entwicklung wie die Fockmatrix-Extrapolation aufbaut. Für die erste Iteration des CPSCF-Verfahrens im Zeitschritt  $N+1$  kann demnach der Zweielektronen-Teil  $\mathbf{A}_2^\xi(N+1)$  (siehe Gleichung 2.64) aus den vorangehenden  $\mathbf{A}_2^\xi$ -Matrizen extrapoliert werden, indem die zur Gleichung 4.3 analoge Matrixgleichung

$$\mathbf{a}_2^\xi = \mathbf{X} \cdot \mathbf{k}, \quad (4.5)$$

mit  $\mathbf{a}_2^\xi = \begin{pmatrix} \mathbf{A}_2^\xi(1) \\ \vdots \\ \mathbf{A}_2^\xi(N) \end{pmatrix}$  gelöst wird (die Form von  $\mathbf{k}$  und  $\mathbf{X}$  ist gegenüber der Fockmatrix-Extrapolation unverändert). Indem man direkt den  $\mathbf{A}_2^\xi$ -Teil extrapoliert, kann in der ersten CPSCF-Iteration die zeitaufwändige Bildung des Terms  $\mathbf{G}(\mathbf{P}^\xi)$  (siehe Gleichung 2.48) entfallen, was nicht möglich wäre wenn man stattdessen die gestörte Einteilchen-Dichtematrix extrapolieren würde (vgl. den Vorteil der Extrapolation der Fockmatrix gegenüber einer Extrapolation der Einteilchen-Dichtematrix im SCF-Verfahren [73]).

Das vorgestellte Extrapolationsverfahren wurde in eine Entwicklungsversion des Programmpakets Q-Chem [64] für die Berechnung von NMR-Verschiebungen ( $\xi \equiv B^i$ , siehe Gleichung 2.62) in das Dichte-basierte Laplace-transformierte DL-CPSCF-Verfahren [7] implementiert. Die Tabelle 4.1 zeigt die Effizienz des Verfahrens für einen Testfall, ein Tetrafluorethen-Molekül  $C_2F_4$  und Schrittweiten von 0.5 fs, der auch im Zusammenhang der Fockmatrix-Extrapolation untersucht wurde. Die Tabelle vergleicht die benötigte Anzahl an CPSCF-Iterationen mit verschiedenen Extrapolationsschemata ( $N/M$ :  $N$ -Schritte, Polynom  $M$ -ter Ordnung) mit dem Fall, bei dem die erste Iteration des CPSCF-Verfahrens vom  $\mathbf{b}^\xi$ -Teil ausgeht (in der Tabelle 4.1 als „Ref.“ bezeichnet). Alle Rechnungen basieren auf einer Beschleunigung des CPSCF-Verfahrens basierend auf einer Übertragung der DIIS-Methode [18] (siehe Abschnitt 2.4.2). Für drei verschiedene Konvergenzkriterien ( $10^{-6}$ ,  $10^{-8}$  und  $10^{-10}$ ) sind zudem die Standardabweichungen (STD) und maximalen Fehler (MAX) gegenüber der Referenz-Rechnung gezeigt (für die Kohlenstoff- und Fluoratome). Die Ergebnisse sind über 1000 Schritte und fünf unabhängige Trajektorien gemittelt.

Für das mittlere Konvergenzkriterium von  $10^{-8}$  zeigt sich also beispielsweise, dass bereits durch die Verwendung des  $\mathbf{A}_2^\xi$ -Teils aus dem vorangehenden Schritt (formal eine 1/0-Extrapolation) statt 7.14 im Mittel nur 4.63 CPSCF-Iterationen benötigt werden. Mit den getesteten Extrapolationsschemata über mehrere Schritte hinweg und Polynomen höherer Ordnung kann man die mittlere Iterationszahl weiter senken auf bis zu 2.02. Für den betrachteten Fall liegen dabei alle Abweichungen gegenüber der Referenz unterhalb von 0.01 ppm. Für das gezeigte Beispiel lassen sich somit in etwa die durch die Fockmatrix-Extrapolation erreichten Einsparungen hinsichtlich der SCF-Iterationszahl auch auf das CPSCF-Extrapolationsverfahren übertragen. Für die getesteten Konvergenzkriterien ( $10^{-6}$ ,  $10^{-8}$ ,  $10^{-10}$ ) sinken die Iterationszahlen ausgehend von kleinen Werten für  $N$  (Anzahl der Extrapolationspunkte) und  $M$  (Ordnung des Extrapolationspolynoms), die Iterationszahlen durchlaufen ein Minimum bevor sie wieder ansteigen (vgl. ebenfalls die Fockmatrix-Extrapolation). Für den getesteten Fall erbringen die Kombinationen 4/2, 6/3 und 8/4 von den getesteten  $N/M$ -Kombinationen die größten Einsparungen hinsichtlich der CPSCF-Iterationsanzahl.

Die Tabelle 4.2 zeigt weitere Ergebnisse der neuen Extrapolationsmethode für CPSCF-Anfangsschätzungen bei einer Reihe weiterer molekularer Systeme. Die Ergebnisse basieren auf Mittelwerten aus drei Trajektorien über jeweils 50 fs und der größeren 6-311G\*\*-Basis.

Tab. 4.1: Mittlere Anzahl der CPSCF-Iterationen  $N_{\text{Ite.}}$ , die ohne Extrapolation (Referenz, Ref.) und mit unterschiedlichen Extrapolationsschemata  $N/M$  benötigt werden sowie Standardabweichungen (STD) und maximale Fehler (MAX) der NMR-Verschiebungen für Kohlenstoff und Fluor für  $\text{C}_2\text{F}_4$  bei DIIS-Fehlernormen von a)  $10^{-6}$ , b)  $10^{-8}$  und c)  $10^{-10}$  als DL-CPSCF-Konvergenzkriterium; Mittelwerte aller Schritte in fünf Trajektorien (jeweils 0.5 ps,  $T = 500\text{K}$ , Zeitschritt 0.5 fs, GIAO-HF/3-21G).

	$N_{\text{Ite.}}$	$\text{MAX}_\text{C}$	$\text{STD}_\text{C}$	$\text{MAX}_\text{F}$	$\text{STD}_\text{F}$
Ref.	6.00	-	-	-	-
1/0	2.88	$5.65 \cdot 10^{-2}$	$8.27 \cdot 10^{-3}$	$2.91 \cdot 10^{-1}$	$4.70 \cdot 10^{-2}$
2/1	2.01	$1.43 \cdot 10^{-2}$	$3.21 \cdot 10^{-3}$	$7.76 \cdot 10^{-2}$	$2.37 \cdot 10^{-2}$
4/2	2.01	$7.08 \cdot 10^{-3}$	$1.79 \cdot 10^{-3}$	$6.39 \cdot 10^{-2}$	$1.73 \cdot 10^{-2}$
a) 6/3	2.03	$1.34 \cdot 10^{-2}$	$1.74 \cdot 10^{-3}$	$7.92 \cdot 10^{-2}$	$1.73 \cdot 10^{-2}$
8/4	2.01	$6.65 \cdot 10^{-3}$	$1.70 \cdot 10^{-3}$	$7.37 \cdot 10^{-2}$	$1.71 \cdot 10^{-2}$
12/6	2.03	$6.78 \cdot 10^{-3}$	$1.70 \cdot 10^{-3}$	$6.70 \cdot 10^{-2}$	$1.73 \cdot 10^{-2}$
16/8	2.05	$6.73 \cdot 10^{-3}$	$1.70 \cdot 10^{-3}$	$7.15 \cdot 10^{-2}$	$1.74 \cdot 10^{-2}$
Ref.	7.14	-	-	-	-
1/0	4.63	$3.20 \cdot 10^{-3}$	$7.24 \cdot 10^{-4}$	$2.20 \cdot 10^{-2}$	$4.38 \cdot 10^{-3}$
2/1	3.01	$4.56 \cdot 10^{-3}$	$9.77 \cdot 10^{-4}$	$3.29 \cdot 10^{-2}$	$5.10 \cdot 10^{-3}$
4/2	2.02	$3.05 \cdot 10^{-3}$	$7.29 \cdot 10^{-4}$	$3.11 \cdot 10^{-2}$	$5.76 \cdot 10^{-3}$
b) 6/3	2.06	$1.21 \cdot 10^{-2}$	$4.80 \cdot 10^{-4}$	$5.76 \cdot 10^{-2}$	$4.39 \cdot 10^{-3}$
8/4	2.04	$1.55 \cdot 10^{-3}$	$4.20 \cdot 10^{-4}$	$1.56 \cdot 10^{-2}$	$3.78 \cdot 10^{-3}$
12/6	2.08	$1.52 \cdot 10^{-3}$	$4.19 \cdot 10^{-4}$	$1.63 \cdot 10^{-2}$	$3.72 \cdot 10^{-3}$
16/8	2.11	$1.47 \cdot 10^{-3}$	$4.18 \cdot 10^{-4}$	$1.63 \cdot 10^{-2}$	$3.72 \cdot 10^{-3}$
Ref.	8.99	-	-	-	-
1/0	6.06	$2.40 \cdot 10^{-4}$	$6.10 \cdot 10^{-5}$	$1.65 \cdot 10^{-3}$	$3.37 \cdot 10^{-4}$
2/1	4.70	$2.49 \cdot 10^{-4}$	$5.63 \cdot 10^{-5}$	$1.25 \cdot 10^{-3}$	$2.78 \cdot 10^{-4}$
4/2	3.71	$4.97 \cdot 10^{-4}$	$7.75 \cdot 10^{-5}$	$2.72 \cdot 10^{-3}$	$3.45 \cdot 10^{-4}$
c) 6/3	2.87	$1.19 \cdot 10^{-2}$	$1.96 \cdot 10^{-4}$	$5.71 \cdot 10^{-2}$	$1.27 \cdot 10^{-3}$
8/4	2.11	$2.72 \cdot 10^{-4}$	$7.10 \cdot 10^{-5}$	$3.22 \cdot 10^{-3}$	$5.41 \cdot 10^{-4}$
12/6	2.13	$1.31 \cdot 10^{-4}$	$2.89 \cdot 10^{-5}$	$1.27 \cdot 10^{-3}$	$2.38 \cdot 10^{-4}$
16/8	2.31	$1.05 \cdot 10^{-4}$	$1.85 \cdot 10^{-5}$	$1.05 \cdot 10^{-3}$	$2.00 \cdot 10^{-4}$

Tab. 4.2: Mittlere Anzahl  $N_{\text{Ite.}}$  der CPSCF-Iterationen, die ohne Extrapolation (Referenz, Ref.) und mit unterschiedlichen Extrapolationsschemata  $N/M$  benötigt werden sowie Standardabweichungen (STD) der NMR-Verschiebungen aller Atome, angegeben als  $N_{\text{Ite.}}$  (STD); Mittelwerte aller Schritte in drei Trajektorien (jeweils 50 fs,  $T = 300\text{K}$ , Zeitschritt 0.5 fs, GIAO-HF/6-311G\*\*, DIIS-Fehlernorm  $10^{-8}$  als DL-CPSCF-Konvergenzkriterium).

Molekül	Ref.	1/0	2/1	4/2	6/3	8/4	12/6
CH <sub>4</sub>	5.88	4.28 ( $2.5 \cdot 10^{-4}$ )	3.02 ( $3.3 \cdot 10^{-4}$ )	3.60 ( $2.1 \cdot 10^{-4}$ )	2.41 ( $5.0 \cdot 10^{-4}$ )	2.29 ( $3.2 \cdot 10^{-4}$ )	2.32 ( $2.9 \cdot 10^{-4}$ )
CH <sub>3</sub> CH <sub>3</sub>	6.00	4.80 ( $1.6 \cdot 10^{-4}$ )	3.34 ( $1.7 \cdot 10^{-4}$ )	3.79 ( $1.3 \cdot 10^{-4}$ )	2.71 ( $3.2 \cdot 10^{-4}$ )	2.06 ( $5.2 \cdot 10^{-4}$ )	2.13 ( $4.9 \cdot 10^{-4}$ )
CH <sub>3</sub> F	7.00	4.88 ( $1.8 \cdot 10^{-3}$ )	2.56 ( $1.8 \cdot 10^{-3}$ )	3.30 ( $1.6 \cdot 10^{-3}$ )	2.29 ( $1.5 \cdot 10^{-3}$ )	2.06 ( $1.2 \cdot 10^{-3}$ )	2.09 ( $1.1 \cdot 10^{-3}$ )
CH <sub>3</sub> OH	7.00	5.04 ( $9.8 \cdot 10^{-4}$ )	3.71 ( $1.5 \cdot 10^{-3}$ )	3.96 ( $1.6 \cdot 10^{-3}$ )	3.45 ( $2.0 \cdot 10^{-3}$ )	2.95 ( $2.3 \cdot 10^{-3}$ )	2.98 ( $2.0 \cdot 10^{-3}$ )
CH <sub>3</sub> NH <sub>2</sub>	7.00	5.00 ( $5.1 \cdot 10^{-4}$ )	3.69 ( $7.7 \cdot 10^{-4}$ )	3.97 ( $8.4 \cdot 10^{-4}$ )	3.35 ( $1.1 \cdot 10^{-3}$ )	2.65 ( $1.5 \cdot 10^{-3}$ )	2.68 ( $1.3 \cdot 10^{-3}$ )
CH <sub>3</sub> CHO	9.00	6.61 ( $1.5 \cdot 10^{-3}$ )	2.95 ( $8.5 \cdot 10^{-3}$ )	3.88 ( $6.6 \cdot 10^{-3}$ )	2.47 ( $5.1 \cdot 10^{-3}$ )	2.32 ( $3.1 \cdot 10^{-3}$ )	2.37 ( $2.8 \cdot 10^{-3}$ )
CH <sub>3</sub> COCH <sub>3</sub>	8.58	5.99 ( $1.2 \cdot 10^{-3}$ )	4.12 ( $4.8 \cdot 10^{-3}$ )	4.17 ( $5.8 \cdot 10^{-3}$ )	4.23 ( $4.7 \cdot 10^{-3}$ )	4.65 ( $4.1 \cdot 10^{-3}$ )	5.26 ( $3.5 \cdot 10^{-3}$ )
CH <sub>3</sub> CN	8.00	5.64 ( $1.8 \cdot 10^{-3}$ )	2.86 ( $2.1 \cdot 10^{-3}$ )	3.60 ( $1.9 \cdot 10^{-3}$ )	2.39 ( $2.2 \cdot 10^{-3}$ )	2.32 ( $9.5 \cdot 10^{-4}$ )	2.48 ( $8.0 \cdot 10^{-4}$ )
CO <sub>2</sub>	7.00	4.08 ( $2.3 \cdot 10^{-3}$ )	2.15 ( $1.8 \cdot 10^{-3}$ )	2.11 ( $5.8 \cdot 10^{-3}$ )	2.19 ( $6.3 \cdot 10^{-4}$ )	2.33 ( $5.3 \cdot 10^{-4}$ )	3.41 ( $9.0 \cdot 10^{-4}$ )
CF <sub>4</sub>	6.34	4.22 ( $3.9 \cdot 10^{-3}$ )	2.00 ( $3.4 \cdot 10^{-3}$ )	2.00 ( $5.2 \cdot 10^{-3}$ )	2.00 ( $3.8 \cdot 10^{-3}$ )	2.00 ( $4.7 \cdot 10^{-3}$ )	2.22 ( $5.4 \cdot 10^{-3}$ )
HCN	7.97	5.99 ( $2.2 \cdot 10^{-3}$ )	2.72 ( $5.2 \cdot 10^{-3}$ )	3.36 ( $3.3 \cdot 10^{-3}$ )	2.29 ( $5.1 \cdot 10^{-3}$ )	2.11 ( $2.5 \cdot 10^{-3}$ )	2.16 ( $1.3 \cdot 10^{-3}$ )
CH <sub>2</sub> CH <sub>2</sub>	7.00	5.45 ( $4.2 \cdot 10^{-4}$ )	3.44 ( $4.8 \cdot 10^{-4}$ )	3.87 ( $4.5 \cdot 10^{-4}$ )	2.95 ( $5.5 \cdot 10^{-4}$ )	2.55 ( $7.6 \cdot 10^{-4}$ )	2.63 ( $7.6 \cdot 10^{-4}$ )
CHCH	7.97	5.97 ( $6.7 \cdot 10^{-4}$ )	3.67 ( $6.9 \cdot 10^{-4}$ )	3.95 ( $7.5 \cdot 10^{-4}$ )	3.37 ( $7.6 \cdot 10^{-4}$ )	3.33 ( $7.1 \cdot 10^{-4}$ )	3.60 ( $7.6 \cdot 10^{-4}$ )
C <sub>6</sub> H <sub>6</sub>	7.13	5.63 ( $3.6 \cdot 10^{-4}$ )	4.99 ( $3.2 \cdot 10^{-4}$ )	4.93 ( $3.2 \cdot 10^{-4}$ )	5.00 ( $3.5 \cdot 10^{-4}$ )	5.10 ( $4.3 \cdot 10^{-4}$ )	5.79 ( $3.7 \cdot 10^{-4}$ )
CH <sub>2</sub> O	8.84	6.31 ( $3.7 \cdot 10^{-3}$ )	2.94 ( $8.5 \cdot 10^{-3}$ )	3.54 ( $7.8 \cdot 10^{-3}$ )	2.66 ( $6.6 \cdot 10^{-3}$ )	2.60 ( $5.5 \cdot 10^{-3}$ )	2.74 ( $4.4 \cdot 10^{-3}$ )
CH <sub>2</sub> CCH <sub>2</sub>	8.00	5.53 ( $3.4 \cdot 10^{-4}$ )	3.72 ( $3.3 \cdot 10^{-4}$ )	3.98 ( $3.4 \cdot 10^{-4}$ )	3.19 ( $4.9 \cdot 10^{-4}$ )	2.89 ( $6.7 \cdot 10^{-4}$ )	3.14 ( $5.9 \cdot 10^{-4}$ )
Thymin	9.00	5.79 ( $1.5 \cdot 10^{-3}$ )	4.76 ( $1.6 \cdot 10^{-3}$ )	4.44 ( $1.9 \cdot 10^{-3}$ )	5.05 ( $1.5 \cdot 10^{-3}$ )	5.36 ( $1.7 \cdot 10^{-3}$ )	5.76 ( $1.7 \cdot 10^{-3}$ )
Ribose	7.95	5.32 ( $7.5 \cdot 10^{-4}$ )	5.05 ( $6.4 \cdot 10^{-4}$ )	5.00 ( $6.7 \cdot 10^{-4}$ )	5.12 ( $6.3 \cdot 10^{-4}$ )	5.51 ( $6.1 \cdot 10^{-4}$ )	6.00 ( $4.5 \cdot 10^{-4}$ )
Glycin	8.55	5.61 ( $2.2 \cdot 10^{-3}$ )	4.23 ( $2.1 \cdot 10^{-3}$ )	4.18 ( $2.3 \cdot 10^{-3}$ )	4.33 ( $2.3 \cdot 10^{-3}$ )	4.85 ( $2.3 \cdot 10^{-3}$ )	5.36 ( $2.2 \cdot 10^{-3}$ )

Für alle molekulare Systeme zeigt sich ein ähnlicher Trend, wie beim  $C_2F_4$ -System mit der kleineren 3-21G-Basis, wenn auch die Iterationszahlen und effektiven Einsparungen etwas voneinander abweichen, beispielsweise bedingt durch die abweichende Flexibilität der molekularen Systeme.

### 4.3 Integralwiederverwendung für veränderte Konformationen

Bei molekulardynamischen Methoden kann der Zeitschritt wegen der numerischen Integration der Kernbewegungsgleichung (siehe Gleichung 2.12) nicht zu groß gewählt werden; typischerweise liegt er im Bereich unterhalb von 1 fs. Als Folge der kleinen Schrittweite weichen die aufeinanderfolgenden Geometrien einer Born-Oppenheimer-MD typischerweise nur geringfügig voneinander ab. Wenn man also immer aufeinanderfolgende Zeitschritte miteinander vergleicht, verschieben sich die Zentren der Zweielektronen-Integrale ebenfalls nur geringfügig, abhängig vom molekularen System und der Simulationstemperatur (Details siehe unten). Als Konsequenz sind auch die Änderungen sehr vieler Zweielektronen-Integrale bei einem Schritt sehr klein (oder auch null), insbesondere bei Zweielektronen-Integralen über Basisfunktionen

- mit sehr entfernten Zentren,
- mit betragsmäßig großen Basisfunktions-Exponenten (schnell abfallend),
- mit Zentren, die alle an einem Atom lokalisiert sind.

Die kleinen schrittweisen Änderungen vieler Zweielektronen-Integrale motiviert ein Born-Oppenheimer-MD-Verfahren, bei dem die Änderungen der Zweielektronen-Integrale aufgrund der Geometrieänderung im Voraus abgeschätzt werden. Die numerisch unveränderten Zweielektronen-Integrale sollen aus dem vorangehenden MD-Schritt wiederverwendet werden. Nur wenn die Integraländerungen durch die Geometrieänderung einen bestimmten Grenzwert übersteigt, müsste somit das zugehörige Zweielektronen-Integral neu berechnet werden. Prinzipiell ließe sich die Gesamtänderung der Zweielektronen-Integrale  $\Delta_{1 \rightarrow N}$  für mehrere Schritte dann als Summe der schrittweisen Änderung erhalten:

$$\Delta_{1 \rightarrow N} = \sum_{n=1}^N \underbrace{(\mu\nu|\lambda\sigma)^{(n)} - (\mu\nu|\lambda\sigma)^{(n-1)}}_{\Delta^{(n)}} \quad (4.6)$$

Ein nützliches MD-Verfahren, das auf der Abschätzung der schrittweisen Integraländerung gründet, muss sich dabei gegenüber Methoden lohnen, bei denen die SCF-Durchläufe für

jede einzelne Geometrie bereits konventionelle Integralscreening-Methoden verwenden (siehe Abschnitt 2.3.1).

Damit stellt sich zu Beginn die Frage, ob es prinzipiell vorteilhaft ist, nach Integraldifferenzen zu screenen, wie im neuen Ansatz, anstatt nach absoluten Integralwerten, wie bei den konventionellen Screening-Methoden. Das Leistungsvermögen eines Screenings nach Integraldifferenzen (und ebenso nach absoluten Integralwerten) hängt natürlich in der Praxis jeweils von den zur Verfügung stehenden Abschätzungen ab. Für den ersten Schritt der Untersuchung soll die Abhängigkeit von Abschätzungen jedoch zunächst ausgeblendet werden und so werden für den Vergleich zunächst die exakten absoluten Integralwerte bzw. exakten Integraldifferenzen verwendet: Für das Beispiel eines DNA-Basenpaares  $\text{DNA}(\text{A-T})_1$  und einer 6-31G\*\*<sup>\*</sup>-Basis zeigt die Abbildung 4.1a einen Fall A, bei dem im SCF für eine bestimmte Molekülkonformation alle Zweielektronen-Integrale für den Austauschteil vernachlässigt werden, deren exakter Betrag kleiner als bestimmte Grenzwerte sind. Im Fall B werden die Differenzen der Zweielektronen-Integrale zur vorangehenden Molekülkonformation exakt berechnet und für den Austauschteil alle Zweielektronen-Integrale wiederverwendet, die sich unterhalb bestimmter Grenzwerte geändert haben; nur die übrigen Zweielektronen-Integrale werden neu berechnet. Die Ergebnisse sind über drei Schritte von 41 a.u.  $\approx$  1 fs innerhalb zwei unabhängiger Trajektorien gemittelt (Temperatur 295 K), um statistisch aussagekräftig zu sein. Das potentiell vorteilhafte Fehlerverhalten bei der Integralwiederverwendung (basierend auf der Integraldifferenz) gegenüber der Integralvernachlässigung (basierend auf dem absoluten Integralwert) zeigt sich auch für weitere untersuchte Beispiele. Die Abbildung 4.1b zeigt als weiteres Beispiel die Ergebnisse einer analogen Untersuchung für ein Kohlenstoffnanoröhren-Fragment ( $\text{C}_{10}\text{H}_{20}$ ).

Das linke Teilschaubild zeigt den Prozentsatz der insgesamt für den Austauschteil zu berechnenden Zweielektronen-Integrale (in allen SCF-Iterationen) gegenüber einem Referenzfall mit Integralgrenzwerten von  $10^{-10}$ ; das mittlere und rechte Teilschaubild zeigt den Energiefehler bzw. die Standardabweichungen der molekularen Kräfte als Funktion des Prozentsatzes zu berechnender Zweielektronen-Integrale. Für alle verschiedenen Grenzwerte ( $10^{-4}, \dots, 10^{-10}$ ) wurden im SCF-Durchlauf dafür konstant 13 SCF-Iterationen (unter Verwendung einer DIIS-Beschleunigung [18]) durchlaufen.

Wie sich an diesem Beispiel zeigt, kann man mit einer Grenzwert-kontrollierten Integralwiederverwendung (aus dem vorangehenden Born-Oppenheimer-MD-Schritt) hier potentiell

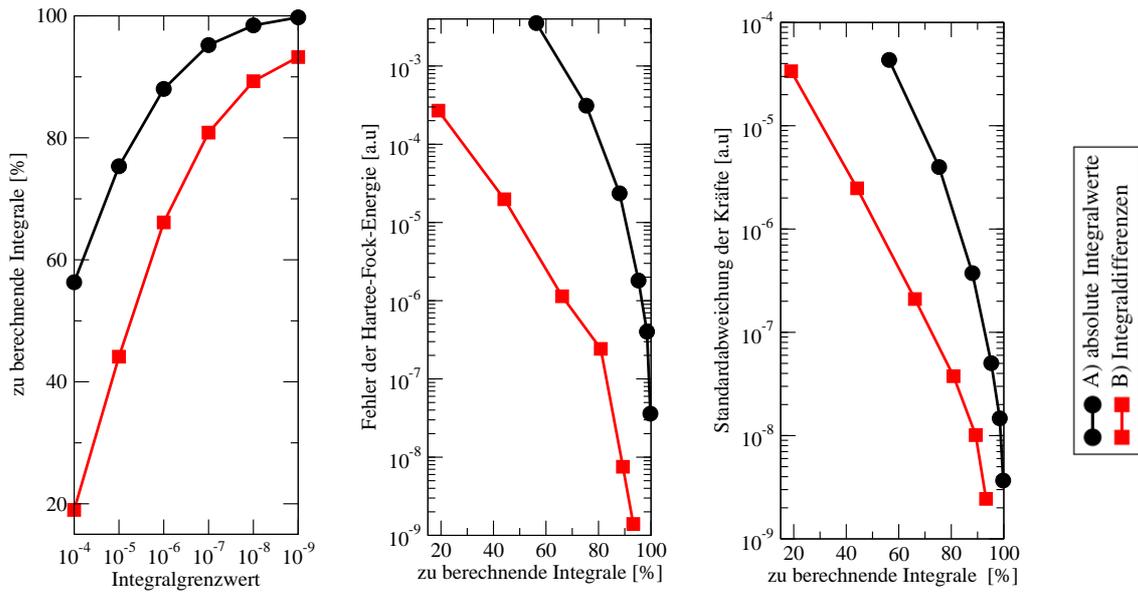


Abb. 4.1a: Vergleich der Vernachlässigung von Zweielektronen-Integralen (gemäß ihrer absoluten Werte) mit der Wiederverwendung von Zweielektronen-Integralen (gemäß ihrer Änderung bei einem Born-Oppenheimer-MD-Schritt) im Austauschteil für das Beispiel eines DNA-Basenpaares DNA(A-T)<sub>1</sub> (HF/6-31G\*\*, 13 SCF-Iterationen, MD-Schritt = 1 fs, T = 295 K).

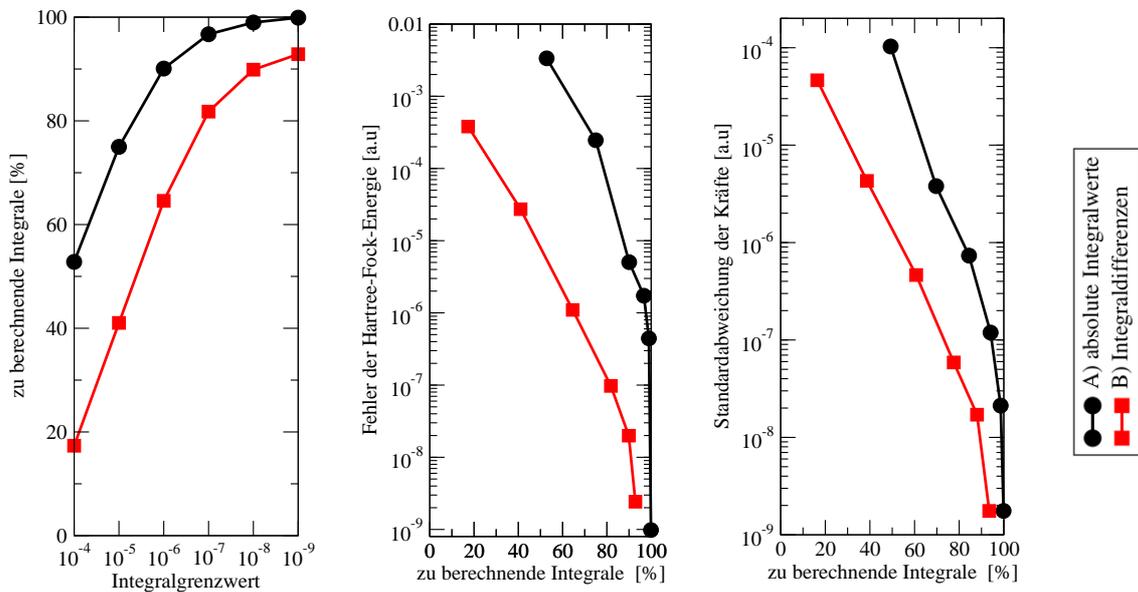


Abb. 4.1b: Vergleich der Vernachlässigung von Zweielektronen-Integralen (gemäß ihrer absoluten Werte) mit der Wiederverwendung von Zweielektronen-Integralen (gemäß ihrer Änderung bei einem Born-Oppenheimer-MD-Schritt) im Austauschteil für das Beispiel eines Kohlenstoffnanoröhren-Fragments C<sub>10</sub>H<sub>20</sub> (HF/6-31G\*\*, 12 SCF-Iterationen, MD-Schritt = 1 fs, T = 295 K).

um eine (bis zu zwei) Größenordnungen kleinere Energie- und Kräftefehler erreichen, als wenn man die Zweielektronen-Integrale gemäß ihrer absoluten Größe vollständig vernachlässigt— bei der gleichen Anzahl zu berechnender Zweielektronen-Integrale und somit einem vergleichbaren Rechenaufwand.

Für ein nutzbares Verfahren müssen natürlich sowohl die exakten absoluten Integralwerte als auch die exakten Integraldifferenzen im Voraus abgeschätzt werden. Zur Abschätzung der exakten absoluten Integralwerte kann man standardmäßig auf die Schwarz-Schranken (siehe Abschnitt 2.3.1) oder aber die neu entwickelten Abstands-abhängigen Schwarz-Schätzwerte (QQR, siehe Abschnitt 3.3) zurückgreifen (das Screening bezieht dann typischerweise Differenzdichte-Elemente mit ein, siehe Gleichung 2.35). Für die Integraldifferenzen durch die Änderung der Molekülgeometrie stehen jedoch bisher keine etablierten Abschätzungen zur Verfügung; gute Schätzwerte dafür herzuleiten wird somit zur Kernvoraussetzung für die neue Methode.

Der auf den ersten Blick naheliegende Ansatz, nämlich die Differenz der Zweielektronen-Integrale als Differenz der konventionellen Schwarz-Schätzwerte abzuschätzen, kann nicht bestehen. Zwar sind die Schwarz-Schätzwerte obere Schranken des Zweielektronen-Integralwerts, sie liegen aber nicht notwendigerweise nahe am tatsächlichen Wert. Daher ist die Differenz der Schwarz-Schranken für zwei unterschiedliche Molekülkonformationen häufig sehr unterschiedlich von der tatsächlichen Änderung des Zweielektronen-Integrals, wie die Abbildung 4.2 veranschaulicht: Die Differenz oberer Schranken  $\Delta_{oo}^{(n)}$  schätzt die tatsächliche Änderung  $\Delta^{(n)}$  nicht notwendigerweise gut ab. Die Abbildung 4.2 zeigt zudem, dass die Änderung eines Zweielektronen-Integrals  $\Delta^{(n)}$  prinzipiell mit der Hilfe von oberen und unteren Schranken streng abgeschätzt werden kann:

$$\Delta^{(n)} \leq \max\left(\Delta_{ou}^{(n)}, \Delta_{uo}^{(n)}\right), \quad (4.7)$$

Das Problem liegt hierbei darin, eng anliegende untere Schranken (bzw. verlässliche Abschätzungen) für den Betrag eines Zweielektronen-Integrals herzuleiten. Einen zukünftigen Ansatzpunkt können hier möglicherweise die in Abschnitt 3.3 vorgestellten Abstands-abhängigen Schwarz-Abschätzungen (QQR) liefern, da sie häufig dem exakten Betrag eines Zweielektronen-Integrals sehr viel näher kommen als die konventionellen Schwarz-Schranken.

Im Rahmen der vorgestellten Arbeit wurde bislang allerdings ein anderer Ansatz zur Abschätzung der Zweielektronen-Integraldifferenz  $\Delta^{(n)}$  verfolgt. Der Ansatz basiert formal auf

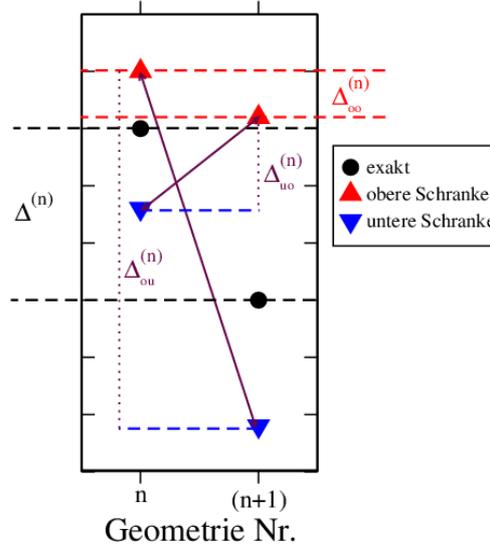


Abb. 4.2: Schema zur Ableitung von Schätzwerten für die Änderung eines Zweielektronen-Integrals zwischen zwei aufeinanderfolgenden Molekülgeometrien  $n$  und  $n+1$ ;  $\Delta^{(n)}$  bezeichnet für ein Beispiel die tatsächliche Änderung,  $\Delta_{oo}^{(n)}$  die Änderung oberer Schranken (z.B. Schwarz-Schranken) und  $\Delta_{ou}^{(n)}$  bzw.  $\Delta_{uo}^{(n)}$  die Änderung zwischen oberer und unterer Schranke bzw. unterer und oberer Schranke.

einer Taylorentwicklung um den Integralwert bei einem MD-Schritt  $n$ :

$$\begin{aligned}
 \Delta^{(n)} &= \sum_{i=x,y,z} \sum_{A=1}^4 \frac{\partial}{\partial R_A^i} (\mu\nu|\lambda\sigma) \Big|_n \cdot \overbrace{\left[ R_A^{i(n+1)} - R_A^{i(n)} \right]}^{\Delta R_A^{i(n)}} \\
 &+ \frac{1}{2} \sum_{i=x,y,z} \sum_{j=x,y,z} \sum_{A=1}^4 \sum_{B=1}^4 \frac{\partial^2}{\partial R_A^i \partial R_B^j} (\mu\nu|\lambda\sigma) \Big|_n \cdot \Delta R_A^i{}^{(n)} \cdot \Delta R_B^j{}^{(n)} \\
 &+ \dots
 \end{aligned} \tag{4.8}$$

Die Größen  $R_A^i$  und  $R_B^j$  bezeichnen dabei die x-, y-, und z-Koordinaten der vier Zentren des Zweielektronen-Integrals. Da die Geometrieänderungen  $\Delta R_A^i{}^{(n)}$  bei einem Born-Oppenheimer-MD-Schritt typischerweise klein sind, kann man schon für Potenzentwicklungen niedriger Ordnung kleine Abbruchfehler erwarten. Allerdings würden schon die zweiten partiellen Ableitungen  $\frac{\partial^2}{\partial R_A^i \partial R_B^j} (\mu\nu|\lambda\sigma) \Big|_n$  einen enormen Zusatzaufwand bedingen (der den eigentlichen Rechenaufwand des Zweielektronen-Integrals sogar noch übersteigen würde). Bedeutsamerweise werden hingegen erste partielle Ableitungen  $\frac{\partial}{\partial R_A^i} (\mu\nu|\lambda\sigma) \Big|_n$  ohnehin berechnet, da sie bei der Born-Oppenheimer-MD zur Berechnung der molekularen Kräfte  $F_A^i$  benötigt werden

(siehe Gleichung 2.60 in Verbindung mit Gleichung 2.48):

$$F_A^i = 2 \cdot \text{tr}\{\mathbf{P}\mathbf{S}^{R_A^i}\mathbf{P}\mathbf{F}\} - 2 \cdot \text{tr}\{\mathbf{P}\mathbf{h}^{R_A^i}\} - \text{tr}\{\mathbf{P}\mathbf{G}^{R_A^i}(\mathbf{P})\} - V_{\text{NN}}^{R_A^i} \quad (4.9)$$

mit

$$\mathbf{G}_{\mu\nu}^{R_A^i}(\mathbf{P}) = \sum_{\kappa\lambda} P_{\kappa\lambda} \left\{ 2 \frac{\partial}{\partial R_A^i} (\mu\nu|\kappa\lambda) - \frac{\partial}{\partial R_A^i} (\mu\lambda|\kappa\nu) \right\}. \quad (4.10)$$

Die ersten partiellen Ableitungen können also bei der Berechnung der molekularen Kräfte abgespeichert werden und innerhalb einer linearen Entwicklung herangezogen werden, um die Integraldifferenz abzuschätzen.

Aus Effizienzgründen basieren Screening-Methoden für die Zweielektronen-Integrale in der Regel auf Schalenquartett-Maxima und ebenso wird das vorgestellte Screening nach Integraldifferenzen deshalb auf Schalenquartett-Maxima ausgerichtet. Die partiellen Ableitungen  $\left. \frac{\partial}{\partial R_A^i} (\mu\nu|\lambda\sigma) \right|_n$  werden dazu über alle zu einem Schalenquartett *MNLS* gehörenden Basisfunktionen maximiert und außerdem über die Ableitungen nach den x-, y-, und z-Komponenten der vier Zentren. So erhält man für jedes Schalenquartett *MNLS* eine gemeinsame maximale Ableitung  $I_{MNLS}^{(n)}$ ,

$$I_{MNLS}^{(n)} \equiv \max_{\substack{\mu \in M, \nu \in N, \\ \lambda \in L, \sigma \in S, \\ i=x,y,z, \\ 1 \leq A \leq 4}} \left| \frac{\partial}{\partial R_A^i} (\mu\nu|\lambda\sigma) \right|_n, \quad (4.11)$$

und somit genau einen gemeinsamen Schätzwert  $\Delta_{MNLS}^{(n)}$  für die zum Schalenquartett gehörenden Integraländerungen:

$$\Delta_{MNLS}^{(n)} \equiv I_{MNLS}^{(n)} \sum_{i=x,y,z} \sum_{A=1}^4 \left| \Delta R_A^i \right|^{(n)} \quad (4.12)$$

Um die Leistungsfähigkeit und das Fehlerverhalten des vorgeschlagenen Ansatzes untersuchen zu können, wurde die vorgestellte Integralspeicher-Born-Oppenheimer-MD-Methode für einzelne MD-Schritte in eine Entwicklungsversion des Programmpakets Q-Chem [64] eingebaut. Die Implementierung gründet dabei auf der Umsetzung des in Abschnitt 3.2 vorgestellten linear-skalierenden indirekten bzw. semidirekten SCF-Verfahrens und beinhaltet im Wesentlichen zusätzlich Routinen, um

- die Schätzwerte  $\Delta_{MNLS}^{(n)}$  während der Berechnung der molekularen Kräfte zu berechnen und zu speichern,

- die Indizierung des Integralspeichers auf die neue Geometrie zu übertragen (inklusive einer Anpassung der Schalenpaarindizierung),
- die Schätzwerte in der veränderten Geometrie innerhalb eines veränderten Integralscreening einzubeziehen und
- statistisch mit den exakten Integraländerungen vergleichen zu können.

Für drei Testsysteme (ein Basenpaar DNA(A-T)<sub>1</sub>, ein Kohlenstoffnanoröhren-Fragment und ein Glycin-Molekül) ist auf der linken Seite der Abbildung 4.3 gezeigt, wie sich die maximalen Schalenquartettänderungen bei einem Born-Oppenheimer-MD-Schritt (1 fs, T = 298 K) statistisch verteilen. Um das Leistungsvermögen der vorgeschlagene Abschätzung zu beurteilen, gibt die Abbildung neben der exakten statistischen Verteilung (rot) auch die Verteilung wieder, wie sie die Abschätzung gemäß der Gleichung 4.12 ergibt (blau). Für alle drei Testsysteme bildet die Abschätzung den Verlauf der exakten Verteilung ab, allerdings um bis zu einer Größenordnung nach links verschoben, was einer Überschätzung der tatsächlichen Änderung um bis zu einer Größenordnung entspricht (im Mittel). Die Teilschaubilder auf der rechten Seite erlauben eine genauere Analyse der Überschätzung (oder aber auch Unterschätzung) der tatsächlichen Änderung. Die Teilschaubilder zeigen die Größenverteilung der Quotienten aus der abgeschätzten und tatsächlichen maximalen Schalenquartettänderung. Im Intervall unterhalb von 10<sup>0</sup> ist also der Anteil unterschätzter Änderungen gezeigt und in den Intervallen 10<sup>0</sup> – 10<sup>1</sup>, 10<sup>1</sup> – 10<sup>2</sup> bzw. 10<sup>2</sup> – 10<sup>3</sup> eine Überschätzung um eine, zwei bzw. drei Größenordnung(en).

Für die betrachteten Testfälle werden also jeweils über 90 Prozent der Schalenquartettänderungen innerhalb eines Faktors von 1–10 richtig abgeschätzt und eine Unterschätzung tritt nicht auf. Die erzielten Ergebnisse der Approximation nach Gleichung 4.12 erweisen sich somit als vielversprechend. Unterhalb von zehn Prozent der Schalenquartettänderungen werden allerdings um mehr als eine Größenordnung überschätzt, was das Einsparungspotential drosselt. Zudem unterliegt die beschriebene Abschätzung einer weiteren Einschränkung: Ein konventionelles Screening des Zweielektronen-Teils bei der Berechnung der Energiegradienten kann für jeden Integralwert (anders als im SCF-Verfahren) nicht nur ein Element der Einteilchen-Dichtematrix miteinbeziehen, sondern das Produkt zweier Dichtematrix-Elemente (vgl. Gleichungen 4.9 und 4.10). Insbesondere deshalb müssen in der Praxis nicht für alle

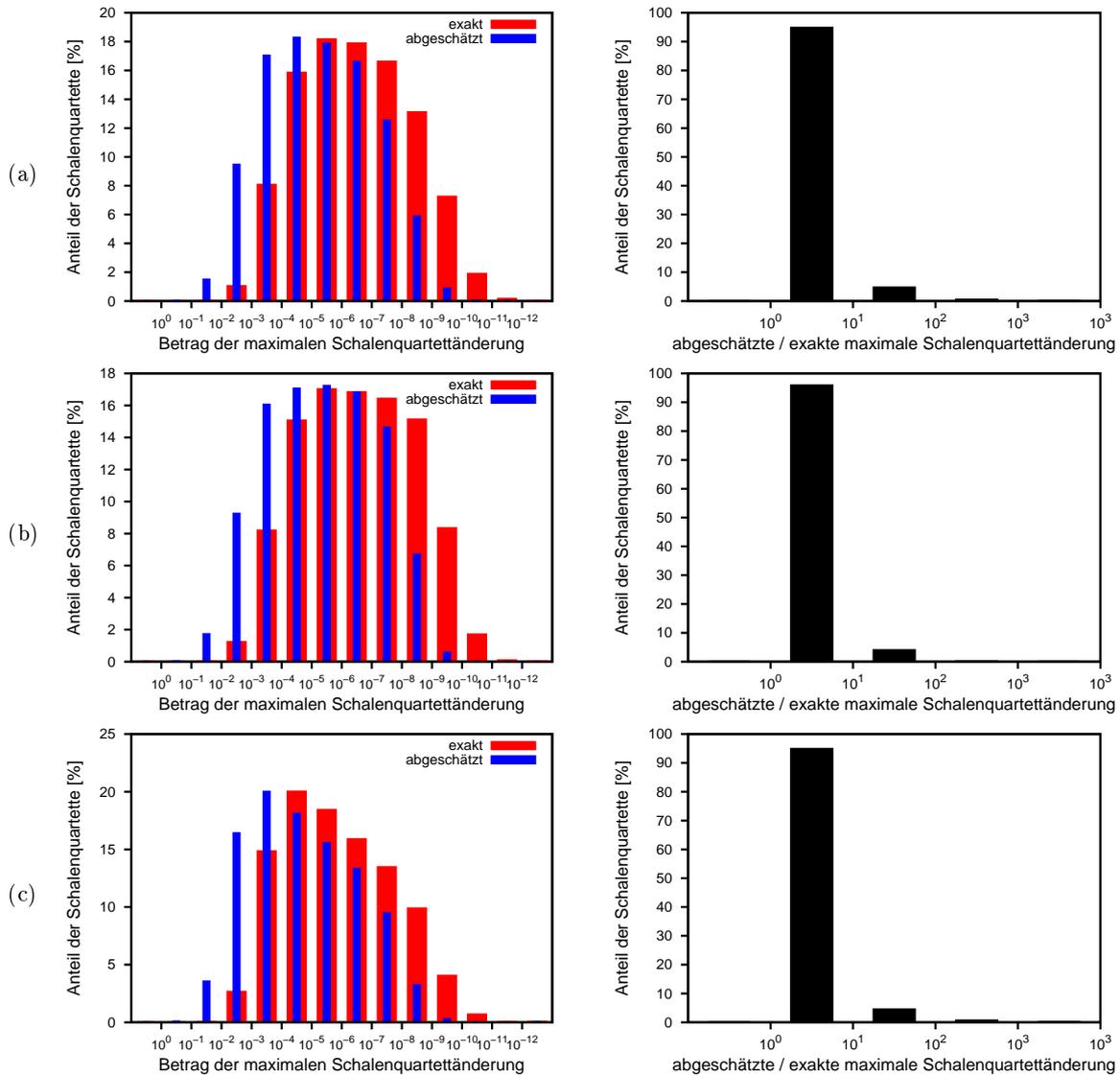


Abb. 4.3: Exakte und abgeschätzte Größenverteilung der maximalen Schalenquartettänderung (links) und Verteilung des Quotienten aus der abgeschätzten und exakten maximalen Schalenquartettänderung (rechts) für ein (a) Basenpaar DNA(A-T)<sub>1</sub> (Basissatz 6-31G\*\*), (b) Kohlenstoffnanoröhren-Fragment C<sub>10</sub>H<sub>20</sub> (Basissatz 6-31G\*\*) und (c) Glycin-Molekül (Basissatz cc-pVTZ). Ausgewertet sind die signifikanten Schalenquartette zur Bildung der 1. Austauschmatrix bei Verwendung der Einteilchen-Dichtematrix aus der vorangehenden Geometrie (SCF-Konvergenz  $10^{-7}$ ; Integralgrenzwert bei der Kraftberechnung für den Austauschteil  $10^{-14}$ , sonst  $10^{-10}$ ; BOMD-Schritt 1 fs; T = 295 K; gemittelt über jeweils 3 Schritte aus 2 Trajektorien).

signifikanten Zweielektronen-Integrale auch alle ersten partiellen Ableitungen  $\left. \frac{\partial}{\partial R_A^i}(\mu\nu|\lambda\sigma) \right|_n$  berechnet werden. Um für die statistischen Aussagen in Abbildung 4.3 alle ersten partiellen Ableitungen  $\left. \frac{\partial}{\partial R_A^i}(\mu\nu|\lambda\sigma) \right|_n$  zur Verfügung zu haben, wurde der Integralgrenzwert für die Gradientenberechnung des Austauschanteils auf  $10^{-14}$  gesetzt. Für die in der Praxis üblichere Wahl von  $10^{-10}$  werden bei den Systemen (a)/(b)/(c) demgegenüber nur für 73/73/70% Prozent der Schalenquartette die ersten partiellen Ableitungen berechnet. Nur für sie kann sich für zukünftige praktische Anwendungen die Abschätzung der Integraländerung auf die dargestellte Näherung stützen. Um die restlichen Integral-Änderungen effizient abzuschätzen, muss zukünftig ein weiterer Ansatz ausgearbeitet werden, evt. auf Basis des im Abschnitt 3.3 dargestellten Abstands-abhängigen Schwarz-Screenings. Die bisherige Implementierung erlaubt hauptsächlich die dargestellte statistische Auswertung bei einzelnen MD-Schritten. Um zu einer optimalen Effizienzsteigerung bei mehreren MD-Schritten zu gelangen, muss die bisherige Implementierung optimiert und erweitert werden.

#### 4.4 QM/MM-Implementierung für Molekulardynamik

Neben den voranstehenden methodischen QM-Ansätzen zur Effizienzsteigerung wurden im Rahmen der vorgestellten Arbeit bestehende linear-skalierende quantenchemische Methoden zur Berechnung von Energien und molekularen Kräften mit bestehenden QM/MM-Beschreibungen kombiniert. Dazu wurde eine Schnittstelle zwischen dem QM-Programmpaket Q-Chem [64] und MM-Programmpaketen (CHARMM [84], Amber [53], DL-Poly) in der ChemShell-Programmsuite [79] erweitert, ausgehend von Arbeiten von C. V. Sumowski et al. [85]. Die geschaffene Implementierung verwirklicht die Berechnung molekularer Kräfte auf Basis der elektrostatischen Einbettung (gemäß Gleichung 2.81) als Grundlage von beispielsweise Molekulardynamik-Rechnungen. Auf diese Weise ermöglicht die Implementierung effiziente und zugleich verlässliche Energie- und Gradientenrechnungen, beispielsweise eine Beschreibung der Dynamik einer zentralen Molekülregion mittels der in Q-Chem enthaltenen linear-skalierenden Methoden und zugleich eine effiziente Beschreibung der Umgebungseinflüsse. Die Details der Implementierung beschreibt der Anhang A.

## 4.5 Zusammenfassung und Ausblick

Im Rahmen dieser Arbeit wurde ein Ansatz vorgestellt, um CPSCF-Rechnungen in aufeinanderfolgenden Schritten einer Born-Oppenheimer-Molekulardynamik zu beschleunigen. In Anlehnung an die Methode der Fockmatrix-Extrapolation beim SCF-Verfahren gelingt es, auch beim CPSCF-Verfahren Iterationen einzusparen, indem man den gestörten Zweielektronenteil aus vorangehenden MD-Schritten extrapoliert, basierend auf einer einfachen Polynomentwicklung. So lassen sich für betrachtete NMR-Verschiebungsrechnungen an aufeinanderfolgenden Molekülgeometrien ohne signifikante Genauigkeitseinbußen bis maximal fünf CPSCF-Iterationen gegenüber CPSCF-Rechnungen einsparen, die nur ausgehen vom gestörten Ein-elektronen-Teil und dem von der gestörten Einteilchen-Dichtematrix unabhängigen Teil. In zukünftigen Rechnungen muss überprüft werden, inwiefern sich die beschriebenen Einsparungen auch auf sehr große molekulare Systeme übertragen lassen. Wünschenswert wäre zudem eine Übertragung auf eine Extrapolation über mehrere Zeitschritte hinweg.

Darüber hinaus wurde für SCF-Rechnungen ein Ansatz untersucht, um die Kosten zur Fock-Matrix-Bildung (bzw. Kohn-Sham-Matrix-Bildung) zu senken. Der Ansatz besteht darin, die Zweielektronen-Integrale aus vergangenen MD-Schritten zur Lösung des aktuellen SCF-Problems wiederzuverwenden, wenn die Integrale gemäß einer Abschätzung numerisch invariant sind. Der ergänzende Ansatz der Integralwiederverwendung kann dabei auf die im vorangehenden Kapitel vorgestellte SCF-Integralspeicher-Methode zurückgreifen. Die präsentierte Abschätzung der Integraländerungen basiert auf einer linearen Entwicklung mit den bei der Kraftberechnung gespeicherten ersten partiellen Integralableitungen. Die erzielten Testergebnisse der Approximation erweisen sich als vielversprechend, wenn auch die Approximation dazu tendiert, die Integraländerungen zu überschätzen. Darüber hinaus wird das Einsparungspotential dadurch gesenkt, dass die vorgestellte Approximation typischerweise nicht für alle Zweielektronen-Integrale zur Verfügung steht. Zukünftig bleibt daher zu untersuchen, ob sich die vorgestellte Näherung mit anderen Abschätzungen kombinieren lässt, beispielsweise mit den im vorangehenden Kapitel dargestellten QQR-Schätzwerten. Die bisherigen Implementierungen erlauben die Integralwiederverwendung nur für den Austauschteil sowie einzelnen MD-Schritte und eine Implementierung für den Coulomb-Teil sowie für mehrere MD-Schritte steht also noch aus.

Der abschließende Abschnitt des vorangegangenen Kapitels skizziert die Erweiterung ei-

ner Schnittstelle zwischen dem QM-Programmpaket Q-Chem [64] und MM-Programmpaketen (CHARMM [84], Amber [53], DL-Poly) innerhalb der ChemShell-Programmsuite [79]. Mittels der Programmschnittstelle können linear-skalierende quantenchemische Methoden zur Berechnung von Energien und molekularen Kräften mit bestehenden QM/MM-Methoden in zukünftigen Anwendungsrechnungen routinemäßig kombiniert werden.

## Kapitel 5

### QM- und QM/MM-Berechnung von NMR-Verschiebungen

#### 5.1 Einführung

Die kernmagnetische Resonanzspektroskopie (NMR-Spektroskopie) hat sich in den letzten Jahrzehnten zu einer der zentralen spektroskopischen Methoden in der Chemie entwickelt. Auf der Basis von NMR-Daten (chemischen Verschiebungen, Intensitäten, Spin-Spin-Kopplungskonstanten, etc.) können heute organische und anorganische Verbindungen häufig weitreichend strukturell charakterisiert werden. Routinemäßige Auswertungen im Labor können sich dabei einerseits auf vielfältige Erfahrungswerte stützen oder auch auf empirische Inkrementverfahren (siehe z.B. Referenz [86] und darin enthaltene Referenzen).

Allerdings kann die Interpretation des gemessenen Spektrums auf Grundlage empirischer Daten auch sehr schwierig oder unmöglich werden, insbesondere z.B. für Systeme mit komplexen Bindungssituationen oder aber für makro- und supramolekulare Systeme. Hier kann die ab-initio Berechnung von NMR-Daten, angefangen von NMR-Verschiebungen, einen wesentlichen Beitrag zur Strukturaufklärung leisten. Neben der prinzipiellen Molekülkonstitution und -konfiguration kann man mit Ab-initio-Rechnungen auch sehr genaue strukturelle Daten ermitteln, wie z.B. Bindungslängen und -winkel etc. (siehe z.B. Referenz [45,80]). Gerade für genaue konformationelle Charakterisierungen erweisen sich hingegen empirische Modelle für makro- und supramolekulare Systeme häufig als unzureichend (siehe z.B. Referenz [85]).

Zentral für die Berechnung von NMR-Verschiebungen an großen Systemen ist einerseits der methodische Fortschritt der quantenchemischen Methoden. So erbrachte die Entwicklung linear-skalierender Methoden zur Berechnung von NMR-Verschiebungen auf Hartree-Fock- oder DFT-Niveau (siehe Abschnitt 2.4 und Referenzen [5–7]) in den letzten Jahren einen entscheidenden methodischen Fortschritt. Ein wichtiges Ziel liegt darin, auch für Post-Hartree-

Fock-Methoden den Rechenaufwand maßgeblich und ohne signifikante Genauigkeitseinbußen zu senken. Andererseits spielt es für Anwendungsrechnungen auch eine große Rolle, bereits bestehende Methoden optimal ausnutzen zu können, beispielsweise linear-skalierenden quantenchemischen NMR-Methoden auf HF- und KS-DFT-Niveau in Kombination einerseits mit aufwändigeren Methoden (z.B. MP2) oder andererseits mit günstigeren QM/MM-Methoden. Grundvoraussetzung ist dafür insbesondere, dass die Fehlerbalken der derzeitigen Methoden möglichst genau eingeschätzt werden können, sowohl für die QM-Beschreibung an sich (Hartree-Fock, DFT-Funktional, Basissatz etc.) als auch in Hinblick auf die benötigten Größen der QM-Regionen zur expliziten Beschreibung von Umgebungseinflüssen auf eine zentrale Molekülregion. In diesem Zusammenhang präsentieren die nachstehenden Abschnitte

- eine Genauigkeitsuntersuchung von NMR-Verschiebungen auf HF/DFT-Niveau (siehe Abschnitt 5.2)
- eine Analyse, wie sich mit intermediären Referenzen die Genauigkeiten von NMR-Verschiebungen steigern lassen (siehe Abschnitt 5.3)
- die Kombination linear-skalierender QM-Methoden mit QM/MM-Ansätzen zur Untersuchung der QM-Größenkonvergenz am Beispiel von NMR-Abschirmungen (siehe Abschnitt 5.4)

Die Genauigkeitsuntersuchungen von NMR-Verschiebungen auf GIAO-HF/DFT- und -MP2-Niveau im Rahmen dieser Arbeit zielt auf eine Ergänzung bereits bestehender Untersuchungen (siehe z.B. Referenzen [87–100]) in mehrerer Hinsicht: Zum einen soll der molekulare Testsatz ausgebaut werden und dabei insbesondere weitere relevante Stoffgruppen, vor allem der organischen Chemie, ergänzt werden. Die Studie verfolgt außerdem das Ziel auch neue Entwicklungen von DFT-Funktionalen und verschiedene Basissätze miteinzubeziehen. Als Referenzwerte strebt sie sehr genaue theoretische Ergebnisse an, CCSD(T) mit großen Basissätzen. Die Fehler bei der Berechnung auf GIAO-HF/DFT- und -MP2-Niveau sollen schließlich nach aussagekräftigen Kriterien für relative Abschirmungen (den chemischen Verschiebungen) bewertet werden.

Bereits frühere Arbeiten (siehe z.B. Referenzen [85, 101, 102]) greifen auf ein weiterführendes Konzept bei der ab-initio Berechnung von NMR-Verschiebungen zurück, um die Genauigkeit gegenüber dem HF- oder DFT-Niveau für eine zentrale Molekülregion möglichst effizient

steigern zu können. Das Konzept gründet darauf, eine zweckmäßige intermediäre Referenz einzuführen und auf hohem theoretischen Niveau zu berechnen. Im Rahmen der vorliegenden Arbeit wird nun untersucht, wie gewinnbringend das Konzept intermediärer Referenzen allgemein für Polypeptide eingesetzt werden kann. Die Darstellung in den Abschnitten 5.2 und 5.3 baut teilweise auf Ergebnissen auf, die im Rahmen der Zusammenarbeit bei drei Bachelor-Arbeiten [103] erzielt wurden.

Der Einfluss der chemischen Umgebung auf die NMR-Verschiebungen einer zentralen Molekülregion wird für verschiedene supramolekulare Systeme im abschließenden Abschnitt des Kapitels untersucht. Dazu wurde die QM-Region vergrößert bis die Änderungen der NMR-Abschirmungen für die betrachteten Kerne vernachlässigbar sind. Um die Größenkonvergenz zu erreichen wurden QM-Regionen mit bis zu rund 1700 Atomen und 16000 Basisfunktionen mit Dichte-basierten linear-skalierenden CPSCF-Methoden behandelt [5–7]. Die Ergebnisse liefern Einsichten in die Lokalität und Konvergenz der elektronischen Struktur an sich und werden damit ebenso in anderem Kontext relevant. Darüber hinaus widmet sich der abschließende Abschnitt des Kapitels auch der Kombination linear-skalierender QM-Methoden mit QM/MM-Ansätzen. Auf diese Weise kann ermittelt werden, in welchem Ausmaß die Einbeziehung der chemischen Umgebung als partielle Punktladungen in einem QM/MM-Ansatz das Konvergenzverhalten mit ansteigender QM-Größe verbessert. Die wesentlichen Inhalte des Abschnitts 5.4 sind in der Veröffentlichung,

D. Flaig, M. Beer, and C. Ochsenfeld,

„Convergence of Electronic Structure with the Size of the QM region:

Example of QM/MM NMR Shieldings“,

*J. Chem. Theory Comput.* **8**, 2260 (2012).

enthalten und sind als Teil dieser Arbeit einbezogen.

## 5.2 Genauigkeiten von NMR-Verschiebungen auf GIAO-SCF- und MP2-Niveau

Die vorgestellte Genauigkeitsuntersuchung für  $^1\text{H}$ - und  $^{13}\text{C}$ -Verschiebungen greift zunächst auf die Strukturauswahl einer theoretischen Studie von A. A. Auer et al. [94] zurück, für die auch experimentelle Abschirmungen in der Gasphase gemessen wurden [104]. Darüber hinaus

wurde die Auswahl um weitere wichtige molekulare Systeme erweitert, so dass die Auswahl nun ein breites Spektrum an organischen Molekülen abdeckt (die Moleküle CO, CO<sub>2</sub>, HCN, und CF<sub>4</sub> sind ebenfalls enthalten) und ebenso einen weiten Bereich der <sup>1</sup>H- bzw. <sup>13</sup>C-NMR-Verschiebungsskala (siehe Tabelle 5.1 und 5.2). Im Einzelnen hinzugekommen sind gegenüber der ursprünglichen Auswahl die Verbindungen Benzol, Furan, Imidazol, Pyridin und Pyrimidin als Vertreter von aromatischen Systemen, deren Abkömmlinge beispielsweise auch in biochemischen Systemen (z.B. Peptide, Nucleinsäuren) eine wichtige Rolle spielen. Dimethylether bzw. Ameisensäure dienen als Repräsentanten der Ether bzw. der Carbonsäuren. Als wichtiges weiteres System wurde das Standardreferenzmolekül Tetramethylsilan (TMS) mit in die Auswahl hinzugenommen, so dass zugleich auch eine Berechnung der Standardverschiebungen ermöglicht wird.

Alle NMR-Berechnungen basieren in Analogie zur Untersuchung von A. A. Auer et al. [94] auf den globalen Minimumstrukturen der isolierten Moleküle, wie sie bei sehr genauen Geometrieoptimierungen auf CCSD(T)/cc-pVTZ-Niveau [62] mithilfe des Programmpakets Cfour [105] gefunden wurden. Die Untersuchung wendet sich zunächst dem methodischen Fehler von MP2- und SCF-Methoden zu und dann dem Einfluss der Basissatzgröße auf die SCF-Ergebnisse. Bei der Bestimmung des methodischen Fehlers wird durchgängig ein großer Basissatz (qz2p, [94, 106]) verwendet, um so den Methoden- und Basissatzfehler (weitgehend) getrennt betrachten zu können. Für die Wasserstoffatome wurde ebenfalls die große qz2p-Basis mit dem Kontraktionsschema H:(7s2p)→[4s2p] verwendet und nicht wie in Referenz [94] eine verkleinerte Basis.

Die Auswahl der DFT-Methoden umfasst die reinen GGA-Funktionale BP86 [107], PBE [39], KT2 [108], B3LYP<sup>0.05</sup> [109] und Hybrid-GGA-Funktionale B3LYP [41, 42], PBE0 [43], B97-2 [44]. Die SCF-Rechnungen sind mit dem Programmpaket Q-Chem [64], sowie einem QM-Paket von J. Kussmann [110] durchgeführt, das neu entwickelte DFT-Funktionale aus der Libxc-Datenbank [111] einbindet. Als Referenzwerte dienen für alle folgenden Beurteilungen sehr genaue theoretische Ergebnisse auf CCSD(T)-Niveau auf der Grundlage der getroffenen Strukturauswahl. Im Blickpunkt der vorliegenden Untersuchung stehen somit die Fehler bei der Berechnung von NMR-Abschirmungen an sich, d.h. basierend auf einzelnen Konformationen der isolierten Moleküle, entkoppelt von der Frage nach dem Einfluss von Molekülschwingungen (Zeitmittelung über verschiedene Konformationen) oder der chemischen Umgebung. Neben

den Referenzwerten auf CCSD(T)-Niveau wurden mithilfe des Programmpakets Cfour [105] auch Abschirmungen auf MP2-Niveau berechnet.

Die Tabellen 5.1 und 5.2 fassen die Ergebnisse der Rechnungen für alle getesteten Methoden zusammen. Darin sind für alle  $^1\text{H}$ - bzw.  $^{13}\text{C}$ -Kerne die Verschiebungen  $\delta_A = \sigma_{\text{TMS}} - \sigma_A$  gegenüber TMS angegeben (siehe Gleichung 2.46). Abschließend führen die Tabellen zusätzlich drei Kriterien zur Bewertung der Methodengenauigkeit auf: die mittlere vorzeichenbehaftete Abweichung (*mean signed deviation*, MSD), die mittlere absolute Abweichung (*mean absolute deviation*, MAD) und die Standardabweichung (*standard deviation*, STD),

$$\begin{aligned} \text{MSD}^{\text{Methode}} &= \frac{1}{N_K} \sum_A^{N_K} \overbrace{\delta_A^{\text{Methode}} - \delta_A^{\text{CCSD(T)}}}^{\Delta\delta_A^{\text{Methode}}} \\ \text{MAD}^{\text{Methode}} &= \frac{1}{N_K} \sum_A^{N_K} \left| \Delta\delta_A^{\text{Methode}} \right| \\ \text{STD}^{\text{Methode}} &= \sqrt{\frac{1}{N_K - 1} \sum_A^{N_K} (\text{MSD}^{\text{Methode}} - \Delta\delta_A^{\text{Methode}})^2}, \end{aligned} \quad (5.1)$$

über  $N_K$  Atomkerne. Dabei ist die Standardabweichung invariant gegenüber der analogen Formulierung mit absoluten isotropen Abschirmungen statt mit Verschiebungen, solange die Referenz (hier TMS) jeweils miteinbezogen wird. Dadurch ist die STD unabhängig von der Wahl der Referenz, im Gegensatz zu der MSD und MAD, was beispielhaft in Tabelle 5.3 gezeigt ist. Bei der MSD und MAD kann es durch den Fehler der Referenzabschirmung selbst, abhängig von der Wahl der Referenz, prinzipiell zu einer Kompensation oder aber Vermehrung des Verschiebungsfehlers kommen. Die allgemeine Beurteilung der Genauigkeit von Verschiebungen stützt sich daher im Folgenden auf die von der Wahl der Referenz unabhängige Standardabweichung.

Tab. 5.1:  $^1\text{H}$ -NMR-Verschiebungen (in ppm) auf GIAO-CCSD(T)-, -MP2-, -DFT- und -HF-Niveau gegenüber TMS<sup>a</sup> und mittlere vorzeichenbehaftete Abweichung (MSD), mittlere absolute Abweichung (MAD) und Standardabweichung (STD) gegenüber den CCSD(T)-Ergebnissen (Basissatz qz2p [94, 106] und CCSD(T)/cc-pVTZ-Geometrien).

Molekül/ Kern	CCSD(T)	MP2	B3LYP <sup>0.05</sup> [109]	B97-2 [44]	B3LYP [41, 42]	PBE0 [43]	HF	KT2 [108]	BP86 [107]	PBE [39]
CH <sub>3</sub> O <u>H</u>	-0.33	-0.20	-0.46	-0.44	-0.45	-0.47	0.18	-0.87	-0.70	-0.72
TMS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CH <sub>3</sub> N <u>H</u> <sub>2</sub>	0.13	0.18	0.04	0.05	0.10	0.03	0.26	-0.03	-0.02	-0.02
CH <sub>4</sub>	0.28	0.26	0.11	0.19	0.21	0.16	0.48	0.01	0.07	0.05
C <sub>2</sub> H <sub>6</sub>	0.91	0.90	0.91	0.89	0.93	0.87	0.95	0.90	0.88	0.88
C <sub>2</sub> H <sub>2</sub>	1.48	1.50	1.29	1.31	1.29	1.30	1.80	1.42	1.15	1.14
CH <sub>3</sub> CN	1.65	1.67	1.74	1.76	1.78	1.74	1.84	1.79	1.74	1.73
CH <sub>3</sub> COCH <sub>3</sub> , in COC-Ebene	1.79	1.77	1.74	1.89	1.88	1.90	2.09	1.71	1.81	1.80
<u>C</u> H <sub>3</sub> CHO, in CCO-Ebene	1.88	1.85	1.83	2.00	2.00	2.01	2.19	1.83	1.93	1.92
<u>C</u> H <sub>3</sub> CHO, außer CCO-Ebene	2.01	2.02	2.10	2.14	2.16	2.15	2.06	2.09	2.18	2.18
CH <sub>3</sub> COCH <sub>3</sub> , außer COC-Eb.	2.05	2.08	2.17	2.17	2.20	2.18	2.02	2.22	2.26	2.27
<u>C</u> H <sub>3</sub> NH <sub>2</sub> , gauche zu fr. Elektr.	2.36	2.40	2.48	2.38	2.47	2.38	2.30	2.51	2.45	2.47
<u>C</u> H <sub>3</sub> NH <sub>2</sub> , anti zu fr. Elektr.	2.59	2.61	2.83	2.68	2.79	2.66	2.37	3.03	2.87	2.91
HCN	2.63	2.66	2.50	2.54	2.51	2.51	2.92	2.79	2.44	2.42
CH <sub>3</sub> OCH <sub>3</sub> , außer COC-Ebene	3.10	3.14	3.16	3.12	3.19	3.12	2.94	3.25	3.24	3.24
<u>C</u> H <sub>3</sub> OH, anti zu OH	3.37	3.44	3.53	3.40	3.50	3.42	3.30	3.51	3.50	3.52
<u>C</u> H <sub>3</sub> OH, gauche zu OH	3.49	3.54	3.74	3.58	3.70	3.58	3.29	3.86	3.77	3.81
CH <sub>3</sub> OCH <sub>3</sub> , in COC-Ebene	3.63	3.72	3.80	3.64	3.74	3.66	3.45	3.87	3.78	3.80
CH <sub>3</sub> F	4.19	4.24	4.42	4.30	4.40	4.31	4.13	4.45	4.44	4.47
CH <sub>2</sub> CCH <sub>2</sub>	4.70	4.78	4.67	4.91	4.89	4.93	4.99	4.80	4.86	4.87
C <sub>2</sub> H <sub>4</sub>	5.45	5.53	5.57	5.84	5.82	5.91	5.84	5.69	5.85	5.88
HCOO <u>H</u>	5.73	5.84	5.69	6.03	6.03	6.06	6.24	5.64	5.94	5.94
Furan (3, 4)	6.18	6.39	6.18	6.39	6.40	6.44	6.38	6.35	6.37	6.40
Imidazole (5)	6.91	7.10	6.82	7.08	7.09	7.12	7.21	7.13	7.05	7.07
Pyrimidin (5)	7.14	7.55	7.01	7.32	7.31	7.41	7.24	7.07	7.29	7.31
Furan (2, 5)	7.18	7.37	7.10	7.36	7.37	7.39	7.41	7.44	7.34	7.35
Imidazole (4)	7.22	7.37	7.15	7.43	7.44	7.47	7.56	7.40	7.39	7.40
Pyridin (3, 5)	7.25	7.59	7.12	7.44	7.42	7.51	7.44	7.21	7.39	7.42
Imidazole (2)	7.40	7.43	7.26	7.59	7.58	7.62	7.85	7.55	7.52	7.52
<u>H</u> COOH	7.69	7.68	7.90	7.97	8.04	7.98	7.76	8.07	8.08	8.08
Pyridin (4)	7.70	7.84	7.54	7.88	7.86	7.96	8.18	7.67	7.79	7.83
C <sub>6</sub> H <sub>6</sub>	7.83	8.07	7.70	8.01	7.99	8.09	8.15	7.84	7.96	7.99
Imidazole (1)	8.19	8.69	8.15	8.42	8.43	8.47	8.31	8.52	8.43	8.47
Pyrimidin (4, 6)	8.76	8.85	8.63	9.01	8.99	9.07	9.17	8.79	8.99	9.01
Pyridin (2, 6)	8.77	8.89	8.66	9.04	9.03	9.11	9.13	8.83	9.05	9.08

Tab. 5.1: (Fortsetzung)

Molekül/ Kern	CCSD(T)	MP2	B3LYP <sup>0.05</sup> [109]	B97-2 [44]	B3LYP [41,42]	PBE0 [43]	HF	KT2 [108]	BP86 [107]	PBE [39]
Pyrimidin (2)	9.31	9.41	9.24	9.60	9.60	9.66	9.52	9.39	9.67	9.69
H <sub>2</sub> CO	9.48	9.42	9.90	10.19	10.28	10.29	9.52	10.29	10.56	10.63
CH <sub>3</sub> C $\overline{\text{H}}$ O	9.66	9.66	10.02	10.27	10.36	10.33	9.60	10.44	10.62	10.67
MSD		0.09	0.01	0.14	0.17	0.17	0.16	0.12	0.16	0.17
MAD		0.09	0.12	0.17	0.19	0.20	0.21	0.19	0.22	0.24
STD		0.12	0.16	0.18	0.19	0.21	0.21	0.25	0.27	0.29

<sup>a</sup> Die isotrope Abschirmung von Tetramethylsilan ist (in ppm): 31.89 [CCSD(T)], 31.82 [MP2], 31.78 [B3LYP<sup>0.05</sup>], 31.89 [B97-2], 31.94 [B3LYP], 31.78 [PBE0], 32.21 [HF], 31.75 [KT2], 31.78 [BP86], 31.72 [PBE]

Tab. 5.2: <sup>13</sup>C-NMR-Verschiebungen (in ppm) auf GIAO-CCSD(T)-, -MP2-, -DFT- und -HF-Niveau gegenüber TMS<sup>a</sup> und mittlere vorzeichenbehaftete Abweichung (MSD), mittlere absolute Abweichung (MAD) und Standardabweichung (STD) gegenüber den CCSD(T)-Ergebnissen (Basissatz qz2p [94,106] und CCSD(T)/cc-pVTZ-Geometrien).

Molekül/ Kern	CCSD(T)	MP2	B3LYP <sup>0.05</sup> [109]	KT2 [108]	B97-2 [44]	BP86 [107]	PBE [39]	B3LYP [41,42]	PBE0 [43]	HF
CH <sub>4</sub>	-3.7	-3.8	-7.5	-7.9	-4.9	-7.2	-8.0	-6.6	-5.8	-1.3
TMS	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
CH <sub>3</sub> CN	3.3	3.5	2.3	1.9	3.2	2.4	1.9	2.0	2.6	2.9
C <sub>2</sub> H <sub>6</sub>	9.4	9.8	9.9	9.4	9.9	9.3	9.0	9.7	9.1	10.5
CH <sub>3</sub> NH <sub>2</sub>	31.1	32.7	33.6	32.9	32.4	33.2	33.0	33.0	32.0	30.4
CH <sub>3</sub> COCH <sub>3</sub>	31.2	32.7	32.1	29.9	31.6	32.4	32.1	32.2	32.2	30.4
CH <sub>3</sub> CHO	33.0	34.6	35.1	33.3	33.9	35.3	35.4	34.9	34.6	31.9
CH <sub>3</sub> OH	52.2	54.8	55.8	54.3	53.8	55.8	55.6	55.1	53.9	49.9
CH <sub>3</sub> OCH <sub>3</sub>	61.5	64.7	65.8	63.9	62.8	65.6	65.4	64.6	63.4	57.8
CH <sub>3</sub> F	71.2	74.4	75.1	72.7	73.1	75.9	75.8	74.8	73.7	68.1
C <sub>2</sub> H <sub>2</sub>	71.8	73.6	73.7	67.8	75.0	74.8	75.1	75.9	78.3	79.6
CH <sub>2</sub> CCH <sub>2</sub>	75.2	76.5	74.3	70.4	77.7	76.9	76.7	78.3	79.2	80.1
HCN	108.4	109.3	106.8	101.4	113.6	111.1	111.3	114.2	116.9	124.2
Furan (3, 4)	108.5	111.0	110.3	103.4	112.3	112.9	112.7	115.2	114.9	114.6
Imidazole (5)	114.5	117.7	115.1	108.7	117.6	117.4	117.2	120.5	120.2	122.5
CH <sub>3</sub> CN	117.2	120.2	116.7	111.4	122.4	120.4	120.8	123.8	125.3	131.6
Pyrimidin (5)	122.6	129.7	124.4	117.8	124.7	126.7	127.1	128.4	127.9	122.7
C <sub>2</sub> H <sub>4</sub>	122.7	126.3	127.5	122.5	131.2	132.6	132.7	134.2	134.8	134.4
Pyridin (3, 5)	124.8	130.2	126.1	119.4	127.4	128.6	128.9	131.1	130.5	128.7

Tab. 5.2: (Fortsetzung)

Molekül/ Kern	CCSD(T)	MP2	B3LYP <sup>0.05</sup> [109]	KT2 [108]	B97-2 [44]	BP86 [107]	PBE [39]	B3LYP [41,42]	PBE0 [43]	HF
CF <sub>4</sub>	127.0	131.1	138.9	130.2	130.0	137.0	137.0	134.1	131.7	113.2
Imidazole (4)	132.5	135.0	134.4	127.1	136.6	137.3	137.1	140.5	139.6	140.9
C <sub>6</sub> H <sub>6</sub>	133.0	135.7	134.1	127.5	136.6	137.3	137.5	140.6	139.9	142.1
CO <sub>2</sub>	134.2	133.1	129.2	124.9	132.9	130.0	130.2	134.8	135.6	144.7
Imidazole (2)	135.7	135.1	135.0	128.4	138.5	137.2	137.0	141.9	141.3	149.9
Pyridin (4)	136.3	136.5	136.2	129.4	139.8	139.4	139.6	143.4	143.1	149.6
Furan (2, 5)	139.8	142.9	140.1	134.5	142.8	142.9	142.8	146.5	145.8	148.8
Pyridin (2, 6)	152.9	153.7	153.5	147.2	157.3	157.9	158.2	161.5	160.8	165.0
Pyrimidin (4, 6)	158.3	157.9	158.3	151.7	163.1	163.0	163.2	167.0	166.6	173.5
HCOOH	160.8	161.9	161.5	154.7	165.2	164.4	164.3	168.7	168.6	171.6
Pyrimidin (2)	163.1	163.4	164.1	157.6	168.3	168.8	169.1	172.4	171.7	176.1
H <sub>2</sub> CO	187.2	189.3	194.6	189.5	201.7	205.8	206.8	206.2	207.4	201.2
CO	189.0	185.5	184.3	178.6	200.3	195.9	196.7	203.3	205.7	221.8
CH <sub>3</sub> C $\underline{H}$ O	193.1	194.8	198.5	192.2	204.9	207.4	207.9	210.0	209.7	207.6
CH <sub>3</sub> C $\underline{O}$ CH <sub>3</sub>	200.5	202.2	205.7	198.5	211.1	213.2	213.4	217.1	215.4	215.6
CH <sub>2</sub> C $\underline{C}$ H <sub>2</sub>	216.5	223.4	220.1	213.7	227.4	227.8	227.9	234.6	232.1	238.4
MSD		1.9	1.5	-3.4	4.1	4.5	4.6	6.7	6.6	7.8
MAD		2.2	2.6	4.2	4.3	5.1	5.2	7.0	6.8	9.3
STD		2.1	3.2	3.6	3.8	4.7	4.9	5.4	5.4	8.8

<sup>a</sup> Die isotrope Abschirmung von Tetramethylsilan ist (in ppm): 195.3 [CCSD(T)], 197.2 [MP2], 186.6 [B3LYP<sup>0.05</sup>], 190.1 [KT2], 187.1 [B97-2], 183.0 [BP86], 183.3 [PBE], 182.7 [B3LYP], 187.6 [PBE0], 193.9 [HF]

Unter den untersuchten Methoden erreicht die MP2-Methode gemäß der Tabellen 5.1 und 5.2 mit Standardabweichungen von 0.12/2.1 ppm für <sup>1</sup>H/<sup>13</sup>C den kleinsten Wert gegenüber den CCSD(T)-Referenzen. Die SCF-Ergebnisse liegen zwischen 0.16/3.2 und 0.29/8.8 ppm für <sup>1</sup>H/<sup>13</sup>C. Die Abfolge der Methoden hinsichtlich ihrer Genauigkeit weichen für <sup>1</sup>H- bzw. <sup>13</sup>C-Verschiebungen etwas voneinander ab: Bei <sup>1</sup>H-Verschiebungen erweist sich Hartree-Fock als gleich auf mit getesteten DFT-Funktionalen, während für <sup>13</sup>C-Verschiebungen alle getesteten DFT-Funktionale um mehr als 3.4 ppm bessere Resultate als HF liefern. Für sowohl <sup>1</sup>H- als auch <sup>13</sup>C-Verschiebungen schneiden die Funktionale B3LYP<sup>0.05</sup> [109] und B97-2 [44] besonders gut ab, das Funktional KT2 [108] nur im Falle der <sup>13</sup>C-Verschiebungen. Abschließend zeigt die Tabelle 5.4 für <sup>13</sup>C-Verschiebungen die Standardabweichungen der SCF-Methoden gegenüber den CCSD(T)/qz2p-Referenzwerten, wenn die SCF-Rechnungen für die gleiche Strukturaus-

Tab. 5.3: Abhängigkeit der mittleren vorzeichenbehafteten Abweichung (MSD) und mittleren absoluten Abweichung (MAD) von der Wahl der Referenz bei der Berechnung von  $^{13}\text{C}$ -Verschiebungen. Als vier Beispielreferenzen sind hier TMS,  $\text{C}_2\text{H}_2$ ,  $\text{C}_6\text{H}_6$  und  $\text{CH}_3\text{COCH}_3$  ausgewählt. Zum Vergleich ist die von der Wahl der Referenz unabhängige Standardabweichung (STD) gezeigt. Bei der Berechnung der Abweichungen sind jeweils die vier Referenzwerte selbst nicht berücksichtigt (Basissatz und Geometrien analog zu Tabelle 5.2).

Referenz		MP2	B3LYP <sup>0.05</sup> [109]	KT2 [108]	B97-2 [44]	BP86 [107]	PBE [39]	B3LYP [41, 42]	PBE0 [43]	HF
TMS	MSD	1.9	1.4	-3.4	3.9	4.3	4.4	6.5	6.3	7.5
	MAD	2.3	2.6	4.2	4.1	4.9	5.0	6.7	6.5	9.2
$\text{C}_2\text{H}_2$	MSD	0.1	-0.4	0.6	0.9	1.6	1.4	2.7	0.1	-0.1
	MAD	1.6	2.5	3.1	2.8	3.3	3.4	4.6	4.2	7.2
$\text{C}_6\text{H}_6$	MSD	-0.8	0.3	2.1	0.3	0.0	-0.1	-1.1	-0.6	-1.4
	MAD	1.8	2.4	3.2	2.7	3.1	3.3	4.1	4.1	7.2
$\text{CH}_3\text{COCH}_3$	MSD	0.2	-3.8	-1.4	-6.6	-8.4	-8.5	-10.1	-8.6	-7.7
	MAD	1.7	4.4	3.5	7.0	8.9	9.1	10.4	9.2	9.1
STD		2.2	3.4	3.7	3.9	4.8	5.0	5.4	5.5	8.9

wahl kleinere Basissätze verwenden, beispielsweise die *Double-zeta*-Basis SVP [106] oder *Triple-zeta*-Basis tz2p [94, 106]. Zum Vergleich sind auch noch einmal die qz2p-Ergebnisse aus Tabelle 5.2 aufgeführt und andererseits auch die Ergebnisse für die minimale Basis STO-3G [112]. Die Tabelle zeigt, dass vor allem bei kleineren Basissätzen als SVP ein deutlicher Fehleranstieg auftritt, während der Unterschied der Standardabweichungen ausgehend vom SVP-Niveau zu größeren Basissätzen hin deutlich weniger ausgeprägt ist. Zudem ändert sich die Abfolge der DFT-Funktionale bzw. Hartree-Fock hinsichtlich ihres Fehlers maßgeblich in Abhängigkeit des Basissatzes: Einerseits zeigen sich besonders die speziell für NMR-Berechnungen entwickelten GGA-Funktionale B3LYP<sup>0.05</sup> [109], KT2 [108] als empfindlich gegenüber kleinen Basissätzen, denn unterhalb des tz2p-Niveaus scheint es hier überwiegend zu einer Fehlervermehrung für NMR-Verschiebungen gegenüber dem methodischen Fehler zu kommen. Hingegen erreichen die Hybrid-GGA-Funktionale B97-2 [44] und PBE0 [43] und besonders auch die Hartree-Fock-Ergebnisse für kleinere Basissätze bis hin zum SVP-Niveau sogar kleinere Standardabweichungen als auf qz2p-Niveau, was auf eine überwiegende Fehlerkompensation von Methoden- und Basissatzfehler hindeutet.

Tab. 5.4: Abhängigkeit der Standardabweichung (STD) von der Wahl des Basissatzes bei der Berechnung von  $^{13}\text{C}$ -Verschiebungen (Geometrien analog zu Tabelle 5.2).

Basissatz	B3LYP <sup>0.05</sup> [109]	KT2 [108]	B97-2 [44]	BP86 [107]	PBE [39]	B3LYP [41,42]	PBE0 [43]	HF
STO-3G [112]	22.1	21.8	19.4	20.5	20.3	19.3	18.7	16.0
SVP [106]	6.2	6.4	2.9	4.3	4.2	3.0	2.8	6.0
tz2p [94, 106]	3.2	4.0	3.0	4.9	5.0	4.2	4.3	7.9
qz2p [94, 106]	3.2	3.6	3.8	4.7	4.9	5.4	5.4	8.8

### 5.3 Intermediäre Referenzen bei der Berechnung von NMR-Verschiebungen

Die Untersuchung im vorangehenden Abschnitt hat gezeigt, dass die MP2-Methode die Genauigkeit von NMR-Verschiebungen auf HF- oder DFT-Niveau durchgängig übertrifft. Für größere Systeme stoßen die derzeit zur Verfügung stehenden MP2-Methoden zur Berechnung von NMR-Verschiebungen mit zunehmenden Systemgrößen jedoch aufgrund des gravierenden Anstiegs des Rechenaufwands noch schnell an die Grenze des Machbaren bzw. Zweckmäßigen. Eine potentiell gewinnbringende Näherung für große Systeme liegt darin, eine geeignete intermediäre Referenz einzuführen, um für einen zentralen Bereich von Interesse mit möglichst geringem Aufwand über die SCF-Genauigkeit hinweg zu gelangen. Die intermediäre Referenz sollte dabei typischerweise einen kleinen Systemausschnitt um den zentralen Bereich in einer repräsentativen Konformation umfassen. Die intermediäre NMR-Verschiebung  $\delta_{A, i. Ref.}^{\text{MP2}}$  (für einen Atomkern  $A$ ) gegenüber der Standardreferenz (z.B. Tetramethylsilan) kann dann auf höherem MP2-Niveau berechnet und die Effekte durch die molekulare Umgebung auf niedrigerem SCF-Niveau als Inkrement  $\Delta\sigma_A^{\text{SCF}}$  einbezogen werden. Die getroffene Näherung für die NMR-Verschiebung gegenüber einer Standardreferenz (Ref.) lässt sich für den Ansatz wie folgt formulieren:

$$\begin{aligned}
 \delta_A^{\text{MP2}} &= \sigma_{\text{Ref.}}^{\text{MP2}} - \sigma_A^{\text{MP2}} = \overbrace{\sigma_{\text{Ref.}}^{\text{MP2}} - \sigma_{A, i. \text{Ref.}}^{\text{MP2}}}^{\delta_{A, i. \text{Ref.}}^{\text{MP2}}} + \overbrace{(\sigma_{A, i. \text{Ref.}}^{\text{MP2}} - \sigma_A^{\text{MP2}})}^{\Delta\sigma_A^{\text{MP2}}} \\
 &\approx \sigma_{\text{Ref.}}^{\text{MP2}} - \sigma_{A, i. \text{Ref.}}^{\text{MP2}} + \underbrace{(\sigma_{A, i. \text{Ref.}}^{\text{SCF}} - \sigma_A^{\text{SCF}})}_{\Delta\sigma_A^{\text{SCF}}} = \delta_A^{\text{SCF/MP2}}
 \end{aligned} \tag{5.2}$$

Die Verbesserung der Resultate  $\Delta_A^{\text{SCF}/\text{MP2} \rightarrow \text{SCF}}$  gegenüber der nur auf einem niedrigen Niveau berechneten Verschiebung  $\delta_A^{\text{SCF}}$  ergibt sich nach diesem Ansatz als Differenz der MP2- und HF-Verschiebungswerte für die intermediäre Referenz selbst:

$$\Delta_A^{\text{SCF}/\text{MP2} \rightarrow \text{SCF}} = \delta_A^{\text{SCF}/\text{MP2}} - \overbrace{(\sigma_{\text{Ref.}}^{\text{SCF}} - \sigma_A^{\text{SCF}})}^{\delta_A^{\text{SCF}}} = \delta_{A, \text{i.Ref.}}^{\text{MP2}} - \delta_{A, \text{i.Ref.}}^{\text{SCF}} \quad (5.3)$$

Solange die intermediäre Referenz und die betrachtete Molekülregion im tatsächlichen System strukturell eng verwandt sind (die Inkremente  $\Delta\sigma_A^{\text{MP2}}$  also klein sind), ist demnach davon auszugehen, dass man mithilfe intermediärer Referenzen einerseits einen wesentlichen Teil der kurzreichweitigen Elektronenkorrelation beschreibt und andererseits effizient langreichweitige Umgebungseinflüsse miteinbeziehen kann. Prinzipiell lässt sich das Konzept analog auch auf andere Methoden übertragen, beispielsweise könnten zukünftig noch genauere intermediäre Verschiebungswerte auf einem hohen *Coupled-Cluster*-Niveau mit MP2-Inkrementen kombiniert werden.

Einen besonderen Vorteil verspricht das Konzept intermediärer Referenzen für makromolekulare Systeme aus definierten Momomereinheiten. Einmalige Rechnungen für die isolierten Momomereinheiten können hier genaue intermediäre Verschiebungswerte liefern und in tabellierter Form als Grundlage für NMR-Rechnungen an den verschiedenen polymeren Derivaten dienen. In diesem Zusammenhang entwickelt die vorliegende Arbeit das Konzept intermediärer Referenzen für Polypeptide (am Beispiel von  $^{13}\text{C}$ -Verschiebungen) und untersucht im Anschluss die Fehler des Konzepts durch den Vergleich gegenüber exakten MP2-Ergebnissen für das vollständige System. Die Abbildung 5.1 zeigt schematisch das Vorgehen für peptidische Systeme. Für stellvertretende Strukturen der isolierten 20 Aminosäure-Moleküle mit den Termini Acetyl (ACE) und N-Methylamid (NME) wurden auf MP2-Niveau die Verschiebungen  $\delta_{A, \text{i.Ref.}}^{\text{MP2}}$  gegenüber TMS berechnet. Die stellvertretenden Strukturen der Aminosäure-Moleküle ergaben sich dabei gemäß einer Amber-Geometrieoptimierung (FF10, [53]) und TMS wurde auf CCSD(T)/cc-pVTZ-Niveau optimiert. Zu den intermediären Verschiebungen  $\delta_{A, \text{i.Ref.}}^{\text{MP2}}$  können nun näherungsweise die Inkremente  $\Delta\sigma_A^{\text{SCF}}$  addiert werden, um den Effekt der tatsächlichen

- Konformation des Aminosäurerests und
- der Polypeptid- und Wasser-Umgebung etc.

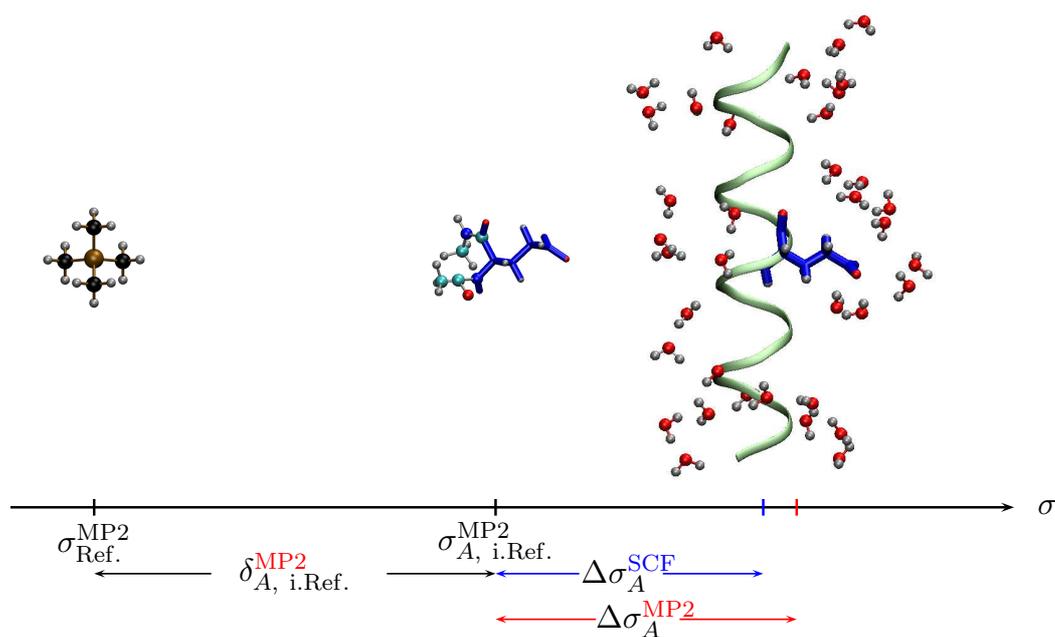


Abb. 5.1: Schema für das Konzept intermediärer Referenzen am Beispiel von Polypeptiden, links: Standardreferenz (TMS), Mitte: intermediäre Referenz (isolierte Aminosäure mit Endgruppen ACE und NME), rechts: der gleiche Aminosäurerest in abweichender Molekülkonformation in der tatsächlichen Polypeptid- und Wasserumgebung.

zu berücksichtigen. Die berechneten Werte sind in den Tabellen 5.5 und 5.6 zusammengestellt, so dass zukünftige Rechnungen darauf zurückgreifen können, um gemäß der Gleichung 5.2 NMR-Verschiebungen von Polypeptiden zu ermitteln. Die Tabelle 5.5 listet die  $^{13}\text{C}$ -Verschiebungen für die verschiedenen Kohlenstoffatome in den isolierten Aminosäuren mit ACE- und NME-Termini gegenüber TMS auf MP2-Niveau auf (entspricht dem Term  $\delta_{A, \text{i.Ref.}}^{\text{MP2}}$ ). Die Tabelle 5.6 stellt dann ergänzend die berechneten Abschirmungen auf HF/qz2p-Niveau zur Verfügung (entspricht dem Term  $\sigma_{A, \text{i.Ref.}}^{\text{SCF}}$ ). Die Werte der durch die Molekülkonstitution nicht eindeutig festgelegten Atomkerne sind in den Tabellen ausgelassen und mit „n.e.“ gekennzeichnet. Die Nomenklatur der Kohlenstoffatome in den Aminosäuren richtet sich nach dem üblichen Standard [113].

Die Fehler der auf niedrigem Niveau berechneten Inkremente  $\Delta\sigma_A^{\text{SCF}}$  gegenüber den MP2-Inkrementen  $\Delta\sigma_A^{\text{MP2}}$  werden im Folgenden exemplarisch für Inkremente aufgrund einer Konformationsänderung oder einer Wasserumgebung untersucht. Für die Untersuchung der Fehler wurden jeweils neun stellvertretende Strukturen der Aminosäure-Systeme erzeugt, indem im

Tab. 5.5:  $^{13}\text{C}$ -Verschiebungen der isolierten Aminosäuren mit ACE- und NME-Termini gegenüber TMS (MP2/qz2p [94, 106], isotrope Abschirmung von TMS: 197.2 ppm, Amber-optimierte Aminosäure-Strukturen (FF10 [53]), TMS optimiert auf CCSD(T)/cc-pVTZ-Niveau); die durch die Molekülkonstitution nicht eindeutig festgelegten Kerne sind mit „n.e.“ bezeichnet.

	COO <sup>-</sup>	C <sub>α</sub>	C <sub>β</sub>	C <sub>γ</sub>	C <sub>δ</sub>	C <sub>ε</sub>	C <sub>ζ</sub>
Ala	175.1	55.8	25.8	-	-	-	-
Arg	177.5	65.3	37.7	33.3	51.4	-	152.1
Asn	175.0	54.5	45.7	170.6	-	-	-
Asp	179.6	62.4	55.9	174.7	-	-	-
Cys	167.5	58.4	35.9	-	-	-	-
Gln	174.9	59.4	41.8	38.2	171.4	-	-
Glu	178.5	65.2	46.1	49.0	175.3	-	-
Gly	168.2	48.9	-	-	-	-	-
Leu	175.5	56.5	54.3	30.8	n.e.	-	-
Lys	177.9	62.7	44.8	29.6	33.4	50.6	-
Met	174.7	59.5	41.7	37.7	-	22.1	-
Ser	168.2	58.9	73.7	-	-	-	-
Thr	169.1	65.6	73.8	28.1	-	-	-
Val	167.5	62.8	41.3	n.e.	-	-	-
Ile	COO <sup>-</sup> 172.9	C <sub>α</sub> 63.7	C <sub>β</sub> 50.5	C <sub>γ</sub> (3) 16.2	C <sub>γ</sub> (3 <sup>1</sup> ) 34.6	C <sub>δ</sub> 16.6	
His	COO <sup>-</sup> 175.0	C <sub>α</sub> 60.7	C <sub>β</sub> 42.7	2 136.6	4 150.5	5 118.6	
Pro	COO <sup>-</sup> 175.8	2 65.4	3 36.5	4 31.7	5 51.8		
Phe	COO <sup>-</sup> 167.3	C <sub>α</sub> 61.8	C <sub>β</sub> 50.7	1 145.7	2, 6 n.e.	3, 5 n.e.	4 135.2
Trp	COO <sup>-</sup> 175.6	C <sub>α</sub> 59.3	C <sub>β</sub> 39.0	2 127.5 5 128.1	3 118.9 6 128.2	3a 140.1 7 115.8	4 124.8 7a 141.6
Tyr	COO <sup>-</sup> 174.9	C <sub>α</sub> 61.9	C <sub>β</sub> 50.0	1 139.7	2, 6 n.e.	3, 5 n.e.	4 165.8

Tab. 5.6:  $^{13}\text{C}$ -Abschirmungen der isolierten Aminosäuren mit ACE- und NME-Termini (HF/qz2p [94, 106], Amber-optimierte Strukturen (FF10 [53])); die durch die Molekülkonstitution nicht eindeutig festgelegten Kerne sind mit „n.e.“ bezeichnet.

	$\text{COO}^-$	$\text{C}_\alpha$	$\text{C}_\beta$	$\text{C}_\gamma$	$\text{C}_\delta$	$\text{C}_\epsilon$	$\text{C}_\zeta$
Ala	6.8	145.1	170.2				
Arg	5.0	136.9	160.5	164.0	148.2	-	36.3
Asn	6.8	147.1	152.9	13.4	-	-	-
Asp	3.0	140.6	144.4	6.3	-	-	-
Cys	13.7	143.6	161.6	-	-	-	-
Gln	6.8	142.7	158.2	160.0	12.6	-	-
Glu	3.7	137.3	155.1	150.3	7.0	-	-
Gly	13.3	150.2	-	-	-	-	-
Leu	6.2	145.0	146.7	168.0	n.e.		
Lys	4.5	139.8	154.4	168.2	164.5	147.4	
Met	7.0	142.4	157.5	161.3	-	173.6	-
Ser	13.2	142.6	128.7	-	-	-	-
Thr	12.5	136.9	130.2	168.1	-	-	-
Val	13.7	139.6	158.5	n.e.	-	-	-
Ile	$\text{COO}^-$	$\text{C}_\alpha$	$\text{C}_\beta$	$\text{C}_\gamma(3)$	$\text{C}_\gamma(3^1)$	$\text{C}_\delta$	
	7.9	138.8	150.9	163.3	178.5	177.9	
His	$\text{COO}^-$	$\text{C}_\alpha$	$\text{C}_\beta$	2	4	5	
	6.7	140.8	157.1	39.8	35.6	72.5	
Pro	$\text{COO}^-$	2	3	4	5		
	5.9	137.0	163.4	166.1	148.6		
Phe	$\text{COO}^-$	$\text{C}_\alpha$	$\text{C}_\beta$	1	2, 6	3, 5	4
	13.8	140.1	150.7	40.9	n.e.	n.e.	53.4
Trp	$\text{COO}^-$	$\text{C}_\alpha$	$\text{C}_\beta$	2	3	3a	4
	6.0	142.5	160.8	55.6	73.4	50.5	59.4
				5	6	7	7a
				62.7	55.4	75.5	44.1
Tyr	$\text{COO}^-$	$\text{C}_\alpha$	$\text{C}_\beta$	1	2, 6	3, 5	4
	6.4	139.8	151.3	51.8	n.e.	n.e.	23.4

Abstand von 40 ps jeweils drei Strukturen aus drei unabhängigen MD-Trajektorien entnommen wurden (Amber-Kraftfeld FF-10 [53], Zeitschritt 1 fs, T=298 K). Das MD-Verfahren wurde dabei für die

- isolierten Aminosäuren,
- Aminosäuren in Wasser mit eingefrorener Aminosäure-Konformation,
- Aminosäuren in Wasser mit veränderlicher Aminosäure-Konformation,

durchgeführt, um so die Untersuchung der Einzeleffekte (Konformation vs. Wasserumgebung) sowie des gemeinsamen Effekts zu ermöglichen. Die Wasserumgebung wurde basierend auf einer TIP3P-Wasserbox [114] zur Aminosäure hinzugefügt und die Anfangsstruktur durch NVT- und NPT-Simulationen verbessert. Die Wasserumgebung wurde nach der MD auf jeweils 2 Å um die Aminosäure, der wichtigsten Solvatationssphäre, eingeschränkt. Für alle Rechnungen dient als intermediäre Referenz die optimierte isolierte Aminosäure-Struktur (siehe oben).

Die Tabelle 5.7 stellt die Ergebnisse der Fehleruntersuchung zusammen. Die Fehler beziehen sich darin auf MP2-Vergleichsrechnungen und die Fehlerkriterien sind analog zu den Gleichungen 5.1 mit MP2-Referenzen gewählt. Neben den Fehlern der intermediär bestimmten  $^{13}\text{C}$ -Verschiebungen führt die Tabelle zum Vergleich auch die Fehler einer reinen HF-Beschreibung auf. Die Fehler sind zudem aufgeschlüsselt nach einfach gebundenen C-Atomen (gesättigt, gesät.) und sonstigen C-Atomen. Alle  $^{13}\text{C}$ -Verschiebungen wurden gegenüber TMS berechnet (für die isotrope Abschirmungen der CCSD(T)/cc-pVTZ-optimierten Struktur auf HF- bzw. MP2-Niveau siehe Tabelle 5.2).

Die statistische Auswertung bei der Wasserumgebung stützt sich (aus Gründen des Rechenaufwands) auf weniger individuelle Aminosäuren als bei den Vakuumstrukturen (siehe Fußnoten der Tabelle), allerdings ebenfalls auf mindestens fünf verschiedene Aminosäuren (Gly, Ser, Cys, Thr, Phe) mit jeweils neun Strukturen (45 Einzelrechnungen). Die damit sicherlich statistisch aussagekräftigen Ergebnisse zeigen, dass das Konzept der intermediären Referenzen sowohl für reine Konformationsänderungen bzw. Wassereinflüsse als auch den kombinierten Effekten eine signifikante Verbesserung gegenüber reinen Hartree-Fock-Rechnungen erlaubt. So sinkt beispielsweise die mittlere Abweichung über alle C-Atome von 7.3 ppm auf unter 1.1 ppm ab und somit deutlich unter die mittlere absolute Größe der Einflüsse auf die Verschiebungen selbst (4.2 ppm). Damit ist mithilfe intermediärer Referenzen eine wichtige

Voraussetzung erfüllt, um mittels der berechneten NMR-Verschiebung die Konformation oder Wasserumgebung von Polypeptiden charakterisieren zu können.

Tab. 5.7: Mittlere vorzeichenbehaftete Abweichung (MSD), mittlere absolute Abweichung (MAD) und Standardabweichung (STD) gegenüber MP2-Verschiebungen für Hartree-Fock-Verschiebungen im Vergleich zu intermediär bestimmten Verschiebungen (tot.: Gesamtabweichung, gesät.: Abweichung für gesättigte C-Atome, sonst.: Abweichung für die sonstigen C-Atome). Die Fehler der intermediär bestimmten Verschiebungen sind gezeigt für die Beschreibung von Konformationsänderungen (Konf.), einer Wasserumgebung und beiden Effekten gleichzeitig (Details siehe Text). Neben den Fehlern ist zudem aufgeführt, wie groß der mittlere absolute Einfluss der Effekte ist (auf MP2-Niveau).

	Hartree-Fock			Intermediär bestimmte Verschiebungen (HF/MP2, siehe Text)								
	tot.	gesät.	sonst.	Konf. <sup>a</sup>			Wasser <sup>b</sup>			Konf.+ Wasser <sup>c</sup>		
				tot.	gesät.	sonst.	tot.	gesät.	sonst.	tot.	gesät.	sonst.
MSD	-1.1	5.9	-8.9	0.0	0.1	-0.1	0.2	0.1	0.2	0.1	0.2	0.0
MAD	7.3	5.9	8.9	0.7	0.3	1.2	0.4	0.2	0.7	1.1	0.5	1.6
STD	8.2	2.6	3.8	1.1	0.4	1.5	0.8	0.2	1.0	1.5	0.6	2.0
mittlerer absoluter Einfluss				2.4	2.7	2.2	1.3	0.7	1.8	4.2	3.9	4.4

<sup>a</sup> gemittelt für die Systeme Ile, Asp, Asn, Arg, Pro, Tyr und <sup>b</sup>

<sup>b</sup> gemittelt für die Systeme Val, Trp, Ala, Leu und <sup>c</sup>

<sup>c</sup> gemittelt für die Systeme Gly, Ser, Cys, Thr, Phe

## 5.4 Konvergenz von QM/MM-Abschirmungen mit der Größe der QM-Region

Der folgende Abschnitt gibt einen Überblick über die Kombination eines QM/MM-Ansatzes von Cui und Karplus [4] mit linear-skalierenden Methoden zur Berechnung der kernmagnetischen Abschirmtensoren auf Hartree-Fock- und DFT-Niveau [5–7]. Im Anschluss wird die Konvergenz reiner QM- und QM/MM-NMR-Verschiebungen mit der Größe der QM-Region untersucht. Die Ergebnisse sollen dabei nicht nur wichtige Einblicke für die Berechnung von NMR-Verschiebungen liefern, sondern können allgemeiner interpretiert werden, um die Lokalität und Konvergenz der elektronischen Struktur selbst zu bemessen. Die Untersuchung zeigt die Ergebnisse für eine Reihe repräsentativer molekularer Systeme.

### 5.4.1 Kombination linear-skalierender NMR-Methoden mit einer QM/MM-Beschreibung

Nach Gleichung 2.62 ist ein Element des kernmagnetisches Abschirmtensors auf Hartree-Fock- oder KS-DFT-Niveau gegeben als

$$\sigma_A^{ij} = \frac{\partial^2 E_0}{\partial B^i \partial \mu_A^j} = 2 \cdot \text{tr}\{\mathbf{P}^{B^i} \mathbf{h}^{\mu_A^j}\} + 2 \cdot \text{tr}\{\mathbf{P} \mathbf{h}^{B^i, \mu_A^j}\} \quad (5.4)$$

mit  $i, j = x, y, z$ . Vor Kurzem führten M. Beer und C. Ochsenfeld das Dichtematrix-basierte Laplace-transformierte-CPSCF-(DL-CPSCF)-Verfahren ein, mit dem die gestörte Einteilchen-Dichtematrix  $\mathbf{P}^B$  innerhalb einer vollständig Dichte-basierten Formulierung bestimmt werden kann, basierend auf einer Laplace-Transformation mit  $\tau$  Laplace-Punkten  $\alpha$  und Matrixexponentialen [7]:

$$\mathbf{P}_{n+1}^{B^i} = \overbrace{\sum_{\alpha}^{\tau} \omega_{\alpha} \{e^{-t_{\alpha} \mathbf{Q} \mathbf{F} \mathbf{Q}}\} \cdot \mathbf{b}_n^{B^i} \cdot \{e^{t_{\alpha} \mathbf{P} \mathbf{F} \mathbf{P}}\}}^{\mathbf{P}_{vo,n}^{B^i}} - \left(\mathbf{P}_{vo,n}^{B^i}\right)^{\dagger} - \mathbf{P} \mathbf{S}^{B^i} \mathbf{P}. \quad (5.5)$$

Dabei steht  $n$  für den CPSCF-Iterationsindex und  $\mathbf{Q}$  bzw.  $\mathbf{P}$  für die unbesetzte und besetzte Einteilchen-Dichtematrix. Der Term  $\mathbf{b}_n^{B^i}$  wird bestimmt als

$$\mathbf{b}_n^{B^i} = \mathbf{S}^{B^i} \mathbf{P} \mathbf{F} - \mathbf{F}_n^{B^i} \mathbf{P} \mathbf{S}. \quad (5.6)$$

$\mathbf{F}$  und  $\mathbf{S}$  sind die Fock-Matrix (bzw. Kohn-Sham-Matrix) bzw. die Überlappmatrix (siehe Abschnitt 2.3). Die Notation  $\mathbf{M}^{B^i}$  bedeutet die Matrix der partiellen Ableitungen nach  $B^i$  der zugehörigen Matrix  $\mathbf{M}$ . Die Matrix  $\mathbf{F}_n^{B^i}$  ist die Summe des Eielektronen-Teils  $\mathbf{h}^{B^i}$  und Zweielektronen-Teils  $\{\mathbf{G}[\mathbf{P}]\}_n^{B^i}$ ,

$$\mathbf{F}_n^{B^i} = \mathbf{h}^{B^i} + \{\mathbf{G}[\mathbf{P}]\}_n^{B^i} = \mathbf{h}^{B^i} + \mathbf{G}^{B^i}[\mathbf{P}] + \mathbf{G}[\mathbf{P}_n^{B^i}]. \quad (5.7)$$

Wie in Abschnitt 2.6 dargestellt, wird bei den QM/MM-Schemata mit einer elektrostatischen Einbettung die elektrostatischen QM/MM-Wechselwirkungen im QM-Teil berücksichtigt. Die elektrostatischen QM/MM-Wechselwirkungen lassen sich zum Eielektronen-Hamilton-Operator  $\hat{h}$  hinzufügen (siehe Gleichungen 2.78 und 2.79). Sie sind für die nachfolgende Studie als Coulomb-Wechselwirkung zwischen partiellen Punktladungen an den Orten der MM-Atome beschrieben, wie es mittlerweile zum Standardvorgehen geworden ist, obwohl auch Verfahren

basierend auf Multipoltermen höherer Ordnung hergeleitet worden sind (siehe z.B. Referenzen [115, 116] und darin enthaltene Referenzen):

$$\tilde{h}_{\text{QM/MM}} = \hat{h}_{\text{QM}} - \sum_n^{N_q} \frac{q_n}{r_n} \quad (5.8)$$

Die Größen  $q_n$  stehen hierbei für die MM-Teilladungen (als Vielfache der absoluten Elektronenladung in atomaren Einheiten) und  $r_n$  ist der Abstand eines QM-Elektrons zu der externen Punktladung. Wegen des veränderten Eielektronen-Hamilton-Operators  $\tilde{h}$  ergibt das SCF-Verfahren eine veränderte Einteilchen-Dichtematrix  $\tilde{\mathbf{P}}$ . Indem man  $\mathbf{P}$  durch  $\tilde{\mathbf{P}}$  ersetzt verändert sich der diamagnetische Beitrag (zweiter Term) in der Gleichung 5.4. Darüber hinaus verändert sich der paramagnetische Beitrag (erster Term) in der Gleichung 5.4, da die Berücksichtigung der externen Punktladungen die Lösungen  $\frac{\partial}{\partial B_i} \mathbf{P}$  des DL-CPSCF-Verfahrens beeinflusst: Da die Punktladungen zu den Eielektronen-Integral-Ableitungen in der Gleichung (5.7) beitragen, muss man den Term  $\mathbf{h}^{B^i}$  durch  $\tilde{\mathbf{h}}^{B^i}$  ersetzen, wobei für  $\tilde{\mathbf{h}}^{B^i}$  gilt:

$$\tilde{h}_{\mu\nu}^{B^i} = h_{\mu\nu}^{B^i} - \sum_n^{N_q} q_n \left\{ \underbrace{\left\langle \frac{\partial}{\partial B^i} \mu \mid r_n^{-1} \mid \nu \right\rangle}_{V_{\mu^{B^i} \nu}^n} + V_{\mu\nu}^{B^i} \right\}, \quad (5.9)$$

Dadurch erhält man zugleich veränderte Matrizen  $\tilde{\mathbf{F}}_n^{B^i}$  (gemäß der Gleichung 5.7),  $\tilde{\mathbf{b}}_n^{B^i}$  (gemäß der Gleichung 5.6) und schließlich  $\tilde{\mathbf{P}}_{n+1}^{B^i}$  (gemäß der Gleichung 5.5).

#### 5.4.2 Grundzüge der Konvergenzuntersuchung

Vor einer ausführlichen Diskussion der Beispielsysteme und Ergebnisse in den nachfolgenden Abschnitten werden hier die methodischen Grundzüge vorgestellt, um die Konvergenz mit der QM-Größe zu untersuchen. Die Abbildung 5.2 veranschaulicht das prinzipielle Vorgehen: Ausgehend von einer zentralen Molekülregion des Interesses (z.B. der gelöste Stoff) werden die QM-Ausschnitte nach und nach vergrößert, bis die Veränderungen in den NMR-Abschirmungen unter ein bestimmtes Kriterium fallen (siehe Abschnitt 5.4.5). Die reinen QM-Rechnungen vernachlässigen den restlichen Teil des molekularen Systems vollständig, während die QM/MM-Rechnungen den restlichen Teil des molekularen Systems als MM-Teilsystem mit einbeziehen. Ein Abstandskriterium definiert die Größe des QM-Teils, nachfolgend als QM-Umgebungsabstand  $r$  bezeichnet. Der QM-Umgebungsabstand hat dabei die folgende Bedeutung: Ein Molekül (oder Residuum) der Umgebung wird im QM-Teil berücksichtigt, wenn



Abb. 5.2: Schema der Konvergenzuntersuchung. Links: Reine QM-Rechnungen mit zunehmender QM-Größe bis zur Konvergenz; Rechts: QM/MM-Rechnungen mit zunehmender QM-Größe bis zur Konvergenz, wobei das MM-Teilsystem den restlichen Teil des molekularen Systems enthält.

irgendein Atom des Moleküls (oder Residuums) näher als der QM-Umgebungsabstand zu irgendeinem Atom der zentralen Molekülregion ist. Dadurch bilden die QM-Ausschnitte die Form der zentralen Molekülregion nach und ein optimales Konvergenzverhalten mit  $r$  wird angestrebt, d.h. kleinste Abweichungen für eine gegebene Anzahl an QM-Atomen. Kovalente Bindungsbrüche bei der Aufteilung in ein QM- und MM-Teilsystem werden falls möglich vermieden (z.B. werden bei den Systemen in Lösung gesamte Moleküle an der QM/MM-Grenzfläche entweder im QM- oder aber MM-Teil berücksichtigt); ansonsten (z.B. bei dem DNA-Enzym-Komplex) wird die QM-Region durch Wasserstoffatome abgesättigt (siehe Abschnitt 5.4.3). Die vorgestellte Untersuchung beschränkt sich wegen des Rechenaufwands auf die Größenkonvergenz für einzelne Konformationen der verschiedenen molekularen Systeme. Jedoch werden die Ergebnisse über alle zentralen Atome eines Atomtyps gemittelt um ein allgemeineres Bild der Größenkonvergenz zu ergeben.

Um QM-Regionen mit bis zu ungefähr 1700 Atomen zu erschließen, beschränkt sich die Untersuchung zudem auf die Niveaus von GIAO-HF bzw. -DFT und verwendet hauptsächlich Basissätze wie 6-31G\*\* [117,118] und SVP [106]. Für die GIAO-HF- und -DFT-Methoden und die Basissätze 6-31G\*\* und SVP können die Fehler der absoluten isotropen Abschirmungen noch recht groß sein. Allerdings werden für die vorgestellte Größenkonvergenz-Untersuchung nur relative Änderungen gegenüber der Berechnung mit der größten QM-Region betrachtet; für analoge relative Änderungen haben sich die SCF-Methoden mit Basissätzen auf SVP-Niveau im vorangehenden Kapitel als weitaus weniger anfällig gezeigt. Inwieweit das Konvergenzverhalten durch die Wahl verschiedener Basissätze und Methoden beeinflusst wird, bemisst zudem eine stellvertretende Untersuchung anhand eines Beispielsystems (größere Basissätze und verschiedene SCF-Methoden).

Alle Rechnungen wurden mit einer Entwicklungsversion des Programmpakets Q-Chem [64] durchgeführt, basierend auf der kürzlich eingeführten DL-CPSCF-Methode [7] mit ausreichend strengen Integral-, SCF- und CPSCF-Konvergenzgrenzwerten und einer überprüften Genauigkeit der Laplace-Entwicklung.

### 5.4.3 Molekulare Systeme

Die Konvergenzuntersuchung betrachtet verschiedene ausgewählte molekulare Systeme:

- Aminopyrazol in wässriger Lösung [55]
- Molekularer Clip in wässriger Lösung [119]
- Glutathion in Methanol [120]
- DNA-Enzym-Komplex [56]
- Hexa-peri-hexabenzocoronene [121]

Die ersten beiden Beispiele sind in Wasser gelöste Systeme: Ausgehend vom isolierten Molekül wird immer mehr der Wasserumgebung explizit quantenmechanisch behandelt. Der QM/MM-Ansatz bezieht den restlichen Teil des molekularen Systems dann jeweils basierend auf dem TIP3P-Wassermodell [114] ein. Um kovalente Bindungsbrüche an der QM-MM-Grenzfläche zu vermeiden, werden nicht vollständig enthaltene Wassermoleküle aus der aktuellen QM-Region ausgeschlossen und erst ab der nächstgrößeren Rechnung im QM-Teil behandelt. Für das Aminopyrazol-Beispiel wurde das Monomer zunächst aus der Kristallstruktur des Hexamers [55] ausgeschnitten, dann im Vakuum mittels MP2/6-31G\*\* (frozen-core) optimiert und in eine Wasserumgebung (50 Å) eingebettet. Anschließend wurde für das System eine MD-Simulation durchgeführt (300 K, Gesamtzeit 20 ps, Zeitschritt 1.5 fs, Merck-Kraftfeld MMFF94 [122]) bei eingefrorenen Kernpositionen des gelösten Stoffs mithilfe des Programmpakets Macromodel [123]. Von der letzten Geometrie der MD-Simulation wurde ein Gesamtsystem von 3359 Atomen ausgeschnitten. Das System umfasst das Aminopyrazol und 15 Å der Wasserumgebung. Für die Konvergenzuntersuchung wurden bis zu 1367 Atome in der QM-Region berücksichtigt.

Ein zweites Beispiel wurde ausgehend vom optimierten Wirt-Gast-Komplex eines Naphthalin-Clips [119] aufbereitet (Reste R:  $\text{P}(\text{CH}_3)\text{O}_2\text{CH}_3$ , Gast: N-Methylnicotinamid). Die Was-

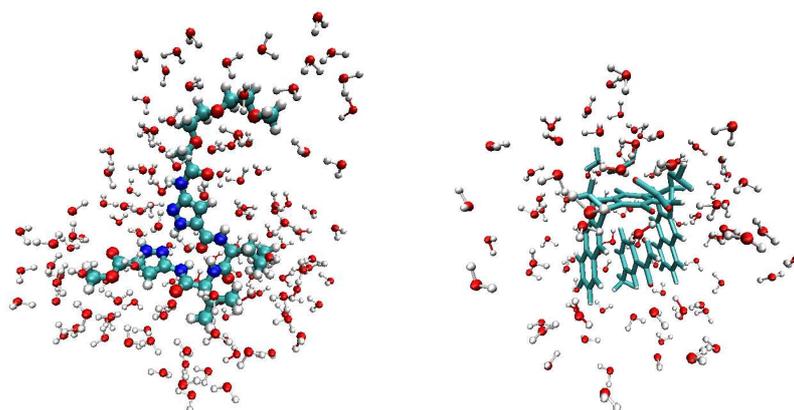


Abb. 5.3a: Aminopyrazol-Peptid [55] (links) und molekularer Clip (rechts) in Wasser [119], gezeigt innerhalb einer Wasserumgebung von 5 Å (die größte betrachtete QM-Sphäre umfasst 10 Å der Umgebung, 1367 bzw. 1150 QM-Atome).

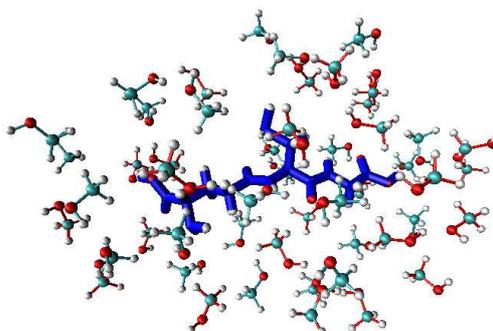


Abb. 5.3b: Glutathion [120], gezeigt innerhalb einer Methanol-Umgebung von 5 Å (die größte betrachtete QM-Sphäre umfasst 10 Å der Umgebung, 961 QM-Atome).

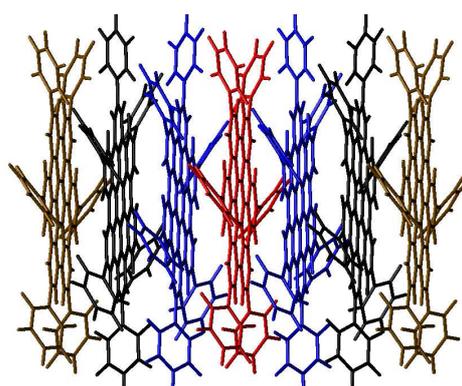


Abb. 5.3c: Heptamer-Ausschnitt aus der gestapelten Struktur des Phenyl-substituierten Hexa-peri-hexabenzocoronens [121], 840 QM-Atome.

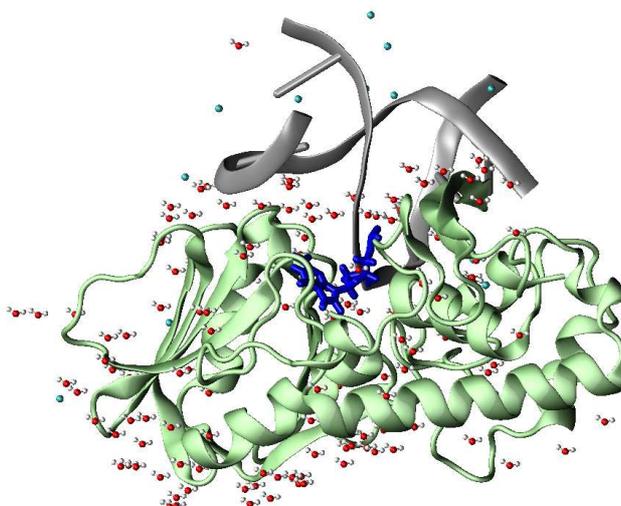


Abb. 5.3d: Komplex des Enzyms MutM (grün) und DNA (grau) mit 8-oxoG-Schaden (blau) [56], gezeigt mit den in der Kristallstruktur enthaltenen Wassermolekülen. Die größte betrachtete QM-Sphäre umfasst 10 Å der Umgebung um den Schaden; darin sind 1614 QM-Atome enthalten bzw. 1752 QM-Atome im System mit einer zusätzlichen Wasserumgebung (für Details siehe Text).

serumgebung wurde analog zum ersten Beispiel hinzugefügt. Für das zweite Beispiel umfasst das Gesamtsystem 3241 Atome und der größte QM-Ausschnitt besteht aus 1150 Atomen.

Das dritte Beispiel zielt darauf ab, das Konvergenzverhalten eines weiteren dipolaren Lösungsmittels zu untersuchen. Die ideale Struktur des Tripeptids Glutathion (PDB code 1DUG [120]) wurde in eine Methanol-Umgebung (50 Å) eingebettet, mithilfe des Desmond-Programms der Schrödinger-Suite [124]. Anschließend wurde eine MD-Simulation mit dem Merck-Kraftfeld (MMFF94 [122]) innerhalb des Programmpakets Macromodel [123] für die Methanol-Umgebung mit eingefrorenen Kernpositionen des gelösten Stoffs durchgeführt (300 K, Gesamtzeit 20 ps, Zeitschritt 1.5 fs). Die letzte Systemkonformation aus der Trajektorie wurde für die Konvergenzuntersuchung verwendet. Bei der Aufteilung in eine QM- und MM-Region wurden wiederum analog zu den vorangehenden Beispielen kovalente Bindungsbrüche an der QM/MM-Grenzfläche vermieden. Die partiellen Punktladungen des Methanol für die QM/MM-Rechnungen wurden aus dem MMFF94-Parametersatz entnommen. Für dieses Beispiel besteht das Gesamtsystem aus 2059 Atomen und der größte QM-Ausschnitt aus 961 Atomen.

Bei den bisherigen Systemen bildet der gelöste Stoff die zentrale Molekülregion für die es von besonderem Interesse ist, die zugehörigen NMR-Verschiebungen zu kennen. In analoger Weise liegt das Hauptaugenmerk in DNA-Enzym-Komplexen häufig auf einer bestimmten Molekülregion, bei Untersuchungen enzymatischer DNA-Reparatur (siehe Kapitel 6): auf einzelnen DNA-Schäden und dem Einfluss der umgebenden Gruppen. Als Beispiel wird nachfolgend ein DNA-Protein-Komplex untersucht, der einen einzelnen DNA-Schaden (8-oxoG) aufweist, basierend auf der Kristallstruktur mit dem PDB-Code 1R2Y [56]. Die Rechnungen verwenden die Standard-AMBER-Parameter für die DNA- und Proteinresiduen und Wassermoleküle und Parameter aus [125] für den 8-oxoG-Schaden. Die Kristallstruktur wurde mit Wasserstoffatomen abgesättigt und mit  $\text{Na}^+$ -Gegenionen neutralisiert. Anschließend wurden die Positionen der Wasserstoffatome mithilfe des NAMD-Kraftfeldprogramms [126] optimiert. Das Gesamtsystem ohne zusätzliche Wassermoleküle umfasst 5515 Atome. Für ein weiteres Beispielsystem wurde eine Wasserumgebung von 20 Å zu der Struktur hinzugefügt, minimiert ( $10^5$  Schritte) und mithilfe einer MD-Simulation angepasst (298 K, Zeitschritt 1.0 fs, 0.1 ns Simulationsdauer). Ein Ausschnitt der letzten MD-Konformation mit 15 Å der Wasserumgebung wurde als Gesamtsystem in wässriger Umgebung ausgewählt (30647 Atome).

Für die Konvergenzstudie wurden immer mehr der umgebenden Proteinresiduen und DNA-Residuen (5'-Nukleotide) im QM-Teil berücksichtigt mit bis zu 1752 QM-Atomen. Anders als in den vorangehenden Beispielen können hier kovalente Bindungsbrüche bei der Aufteilung in eine QM- und MM-Region nicht vermieden werden. Deshalb wurden die gebrochenen Bindungen an der QM/MM-Grenzfläche mit Wasserstoffatomen abgesättigt, mithilfe des ChemShell-Programms [79]. MM-Punktladungen in der Nähe der Grenzfläche wurden dabei gemäß [79] und [127] angepasst.

Das letzte Beispiel konzentriert sich auf ein molekulares System, dessen Umgebungseinflüsse durch Ringströme benachbarter Aromaten bestimmt sind und bei denen man eine geringere Rolle von elektrostatischen Effekten erwarten kann: In der gestapelten Struktur von Phenylsubstituiertem Hexa-peri-hexabenzocoronon (HBC, [121]) wird der Einfluss der Nachbarringe (Ringebenenabstand 3.5 Å) auf die zentrale Einheit betrachtet. Als Referenzsystem wurde das Heptamer gewählt für sowohl die reine QM- als auch QM/MM-Rechnung. Für die QM/MM-Rechnungen wurden die externen Punktladungen mithilfe des RESP-Verfahrens (*restrained electrostatic potentials*, [128]) für das Monomer bestimmt.

#### 5.4.4 Auswertung der QM-Größenkonvergenz

Die Grundaspekte des Auswertungsverfahrens hinsichtlich der

- statistischen Bewertung der Ergebnisse für die einzelnen Kerne
- Abhängigkeit vom Basissatz
- Abhängigkeit von der QM-Methode
- Größenkonvergenz-Kriterien

werden im Folgenden am Beispiel des Aminopyrazol-Wasser-Systems dargestellt. Anschließend werden dann die Ergebnisse für die anderen untersuchten Systeme analog für die  $^1\text{H}$ - und  $^{13}\text{C}$ -Kerne zusammengefasst. Der Anhang C beinhaltet für alle untersuchten Systeme weitere Abbildungen für maximale Abweichungen, das beste und schlechtestes Abschneiden von QM/MM gegenüber reinem QM und Informationen für Stickstoff- und Sauerstoffkerne.

Eine statistische Auswertung der Ergebnisse für die einzelnen Kerne ist hauptsächlich aus zwei Gründen unentbehrlich für allgemeinere Schlussfolgerungen hinsichtlich der Größenkonvergenz: Zum einen beeinflusst die Umgebung die Kerne der zentralen Molekülregion in unterschiedlichem Ausmaß aufgrund der genauen Bindungssituation. Zum anderen kann die Qualität der MM-Beschreibung stark für die einzelnen Kerne schwanken. Die Abbildung 5.4 veranschaulicht den ersten Gesichtspunkt, die unterschiedlich starken Umgebungseinflüsse. Die Abbildung zeigt die Konvergenz der isotropen Abschirmungen gegenüber dem Ergebnis der Rechnung mit der größten QM-Region für den am meisten bzw. am wenigsten beeinflussten  $^1\text{H}$ - und  $^{13}\text{C}$ -Kern im Aminopyrazol-Molekül: Die Änderungen reichen von 7.2 ppm bis hin zu 0.1 ppm, wenn man die Werte des isolierten Aminopyrazols ( $r = 0$ ) mit denen innerhalb der größten Umgebung ( $r = 10$ ) vergleicht.

Der zweite Gesichtspunkt, der unterschiedliche Vorteil durch die QM/MM-Beschreibung, zeigt sich in allen Systemen als ebenfalls sehr ausgeprägt. Die Abbildung 5.5 vergleicht den Fall des größten Vorteils mit dem Fall des geringsten Vorteils durch die QM/MM-Beschreibung: Während es gemäß der Abbildung 5.5a) sehr lohnenswert erscheint, die Umgebung mittels QM/MM einzubeziehen, würde man gemäß der Abbildung 5.5b) überhaupt keinen Vorteil darin sehen. Hier sind sogar die QM/MM-Resultate mit den kleinsten QM-Sphären etwas schlechter als die reinen QM-Ergebnisse, was darauf hinweist, dass die spezifische Bindungssi-

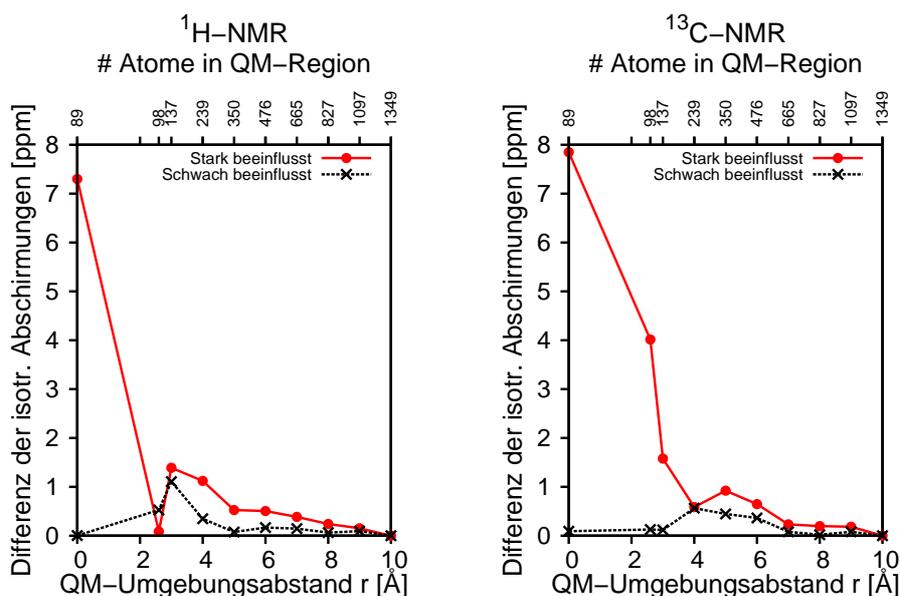


Abb. 5.4: Differenzbetrag der isotropen Abschirmung gegenüber den Ergebnissen der größten QM-Rechnung ( $r = 10 \text{ \AA}$ ) für den am meisten und am wenigsten beeinflussten  $^1\text{H}$ - und  $^{13}\text{C}$ -Kern im Aminopyrazol-Wasser-System, reine QM-Beschreibung: GIAO-B3LYP/SVP.

tuation schlecht durch die Standard-MM-Parameter beschrieben ist. Der Anhang C beschreibt die Details, wie im Rahmen der Untersuchung der beste und schlechteste Fall ausgewählt ist. Aus den beschriebenen Gründen gründen die folgenden Ergebnisse auf Standardabweichungen (*root mean square deviations*, RMS) über alle zentralen Kerne (einer jeden Sorte):

$$\text{RMS}(r) = \sqrt{\frac{\sum_A^N \{\sigma_A(r) - \sigma_A^{\text{QM/MM}}(r_{\text{max}})\}^2}{N}} \quad (5.10)$$

Dabei geht die Summe über alle N-Kerne einer Sorte (z.B.  $^1\text{H}$ ) und  $\sigma_A$  steht für die isotrope Abschirmung eines zentralen Atoms, die entweder auf reinem QM-Niveau oder auf QM/MM-Niveau berechnet wird. Gemäß den Erklärungen in den Abschnitten 5.4.2 und 5.4.3 gibt der Abstand  $r$  die Dicke der QM-Sphäre um die zentrale Molekülregion an und  $r_{\text{max}}$  ist die Dicke der größten behandelten QM-Sphäre. Sowohl die reinen QM-Rechnungen als auch QM/MM-Rechnungen beziehen sich dabei auf die größte QM/MM-Rechnung, da von ihr die genauesten Ergebnisse zu erwarten sind. Das bedeutet allerdings, dass die Standardabweichung der größten reinen QM-Rechnung noch ungleich null sein kann, wenn nämlich die größte QM- und QM/MM-Rechnung immer noch aufgrund eines restlichen MM-Teils voneinander abweichen.

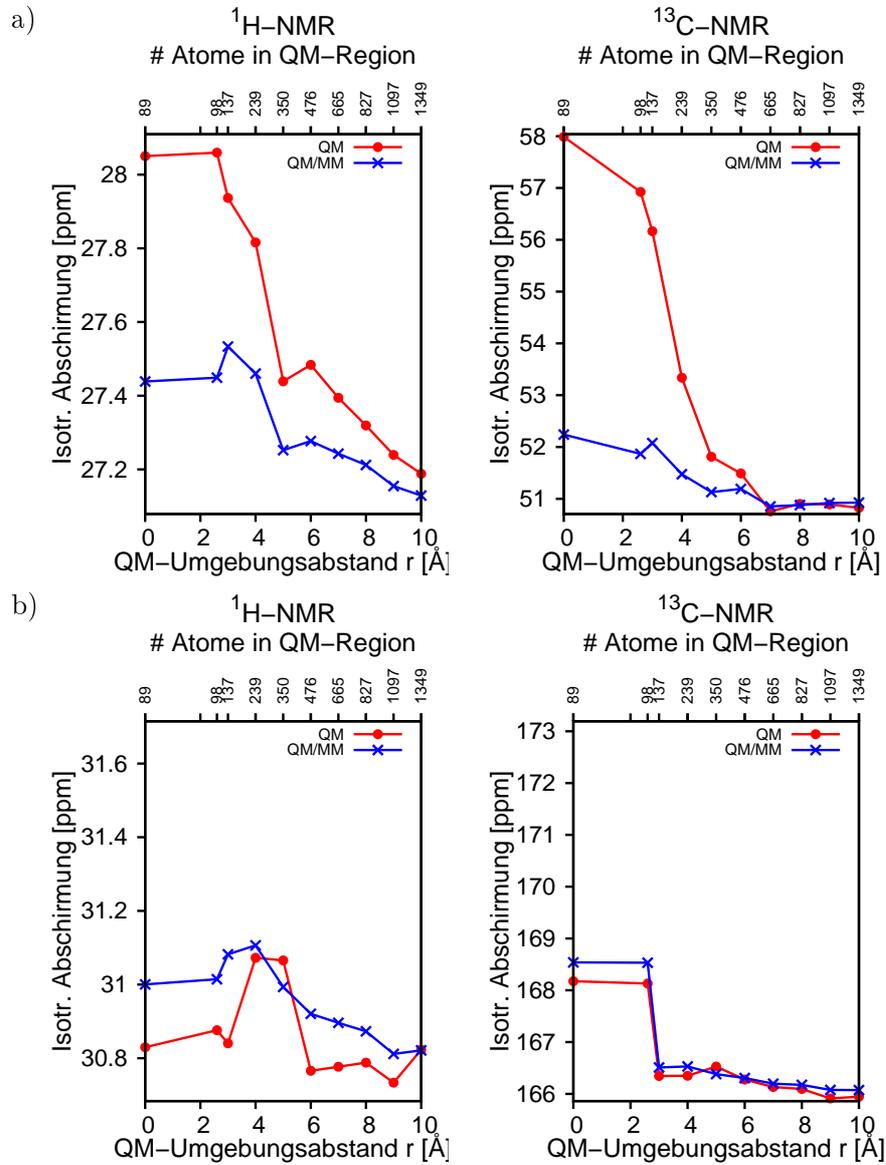


Abb. 5.5: Isotrope Abschirmungen für ausgewählte Kerne im Aminopyrazol-System, bei denen der Vorteil von QM/MM im Vergleich zu reinem QM a) maximal und b) minimal ist, für  $^1\text{H}$  und  $^{13}\text{C}$ , GIAO-B3LYP/SVP.

Die Konvergenz mit der QM-Größe zu erreichen, erfordert (wie nachfolgend gezeigt) recht große QM-Regionen, was die Größe der zu bewältigenden Basissätze einschränkt. Daher zeigt die Abbildung 5.6 zunächst für ein Beispiel, wie der Basissatz die Konvergenzergebnisse (auf HF-Niveau) beeinflusst. Wie sich zeigt treten die Hauptunterschiede bei den kleineren QM-Regionen auf, insbesondere für die kleinsten gewählten Basissätze 3-21G und 6-31G ohne Polarisationsfunktionen. Hingegen geben die kleinsten Basissätze 3-21G und 6-31G das langreichweitige Verhalten bereits sehr gut wieder; die folgenden Untersuchungen beschränken sich auf die Basissätze SVP und 6-31G\*\*. Ein weiterer wichtiger Gesichtspunkt ist der Einfluss durch die verwendete QM-Methode. Die Abbildung 5.7 vergleicht die HF-Ergebnisse mit denen verschiedener DFT-Funktionalen [37, 41–43, 107, 129]. Ein Vergleich mit anderen Korrelationsmethoden, z.B. MP2, wäre wünschenswert, allerdings ist der Rechenaufwand mit den derzeit verfügbaren Methoden in Anbetracht der erforderlichen Systemgrößen zu hoch. Jedoch liefert bereits der Vergleich von HF und verschiedenen DFT-Funktionalen einen Hinweis auf die zu erwartenden Effekte: Wie bei der Basissatz-Abhängigkeit sind die größten Unterschiede für die kleinsten QM-Regionen zu erwarten und für die Heteroatome wie Stickstoff und Sauerstoff. Das allgemeine Bild des Konvergenzverhaltens ist allerdings für alle betrachteten Methoden sehr ähnlich. Demgemäß beschränken die folgenden Untersuchungen sich auf HF- oder B3LYP-Niveau.

Die QM-Regionen für die verschiedenen Systeme nehmen schrittweise bis zu einem QM-Umgebungsabstand von ungefähr 10 Å zu, bei dem die Standardabweichungen typischerweise auf unter 0.1/0.2 ppm für  $^1\text{H}/^{13}\text{C}$  sinkt (siehe nachfolgende Ergebnisse). Die benötigten QM-Größen für jedes weniger strenge Kriterium können den Graphen der Standardabweichung als Funktion der QM-Größe entnommen werden. Die folgende ausführliche Diskussion basiert allerdings nur auf einem Genauigkeitswert. Der Genauigkeitswert bemisst sich an typischen Fehlern der NMR-Verschiebungen auf GIAO-HF/DFT-Niveau, da für die großen Systemgrößen Anwendungsrechnungen momentan hauptsächlich auf diese QM-Niveaus beschränkt sind (obwohl davon ausgegangen werden kann, dass die Erkenntnisse über die benötigten QM-Größen auch für Rechnungen auf höherem QM-Niveau aussagekräftig sind). Als Schlussfolgerung der Abschnitte 5.2 und 5.3 können NMR-Verschiebungen auf GIAO-HF/DFT-Niveau mithilfe des Konzept der intermediären Referenzen auf ungefähr 2-3 ppm genau berechnet werden. Unter der Annahme, dass ein zusätzlicher Fehler von (schlechtestenfalls) 20 % wegen der Vernach-

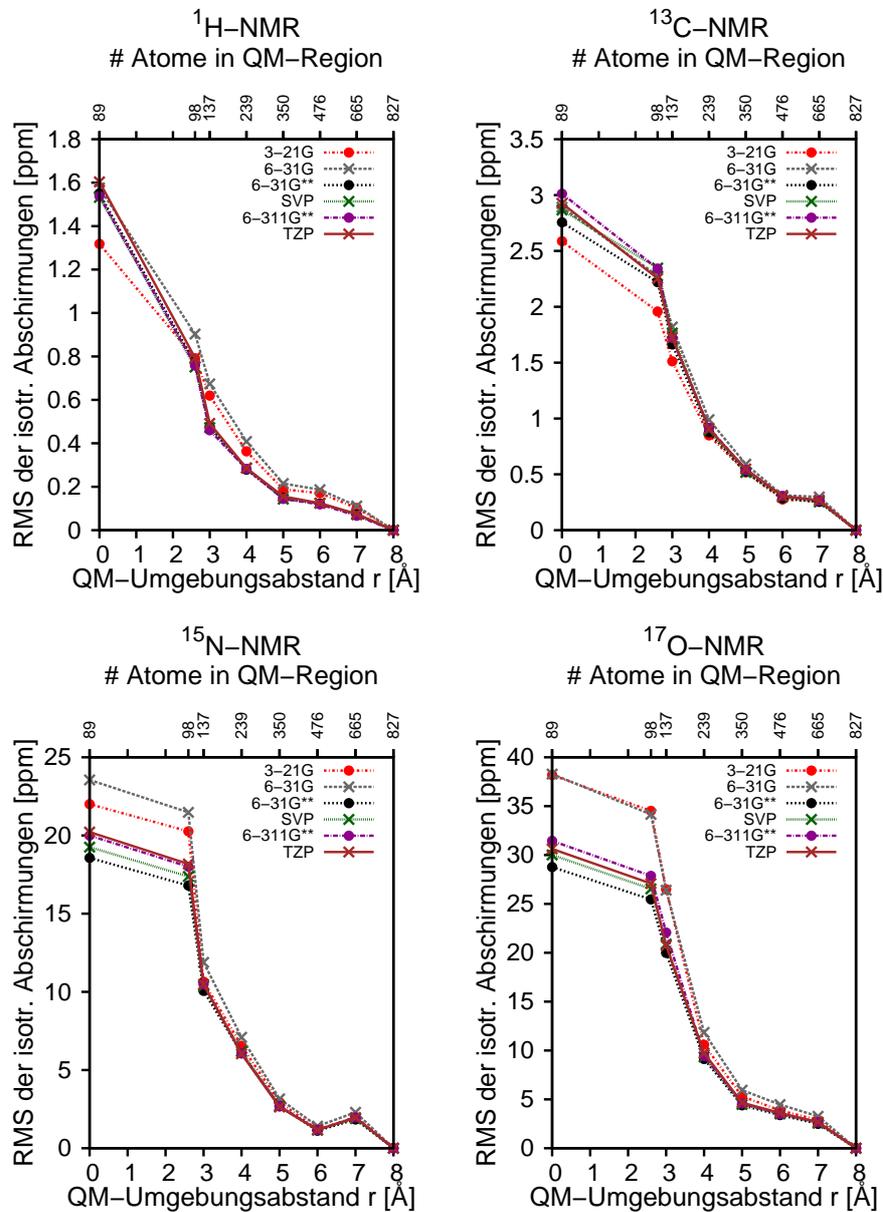


Abb. 5.6: Standardabweichungen (RMS) der isotropen Abschirmungen gegenüber dem QM-Ergebnis des größten QM-Umgebungsabstands (hier 8 Å) für verschiedene Basissätze über 44  $^1\text{H}$ -, 27  $^{13}\text{C}$ -, 9  $^{15}\text{N}$ - bzw. 9  $^{17}\text{O}$ -Kerne, reine QM-Beschreibung: GIAO-HF.

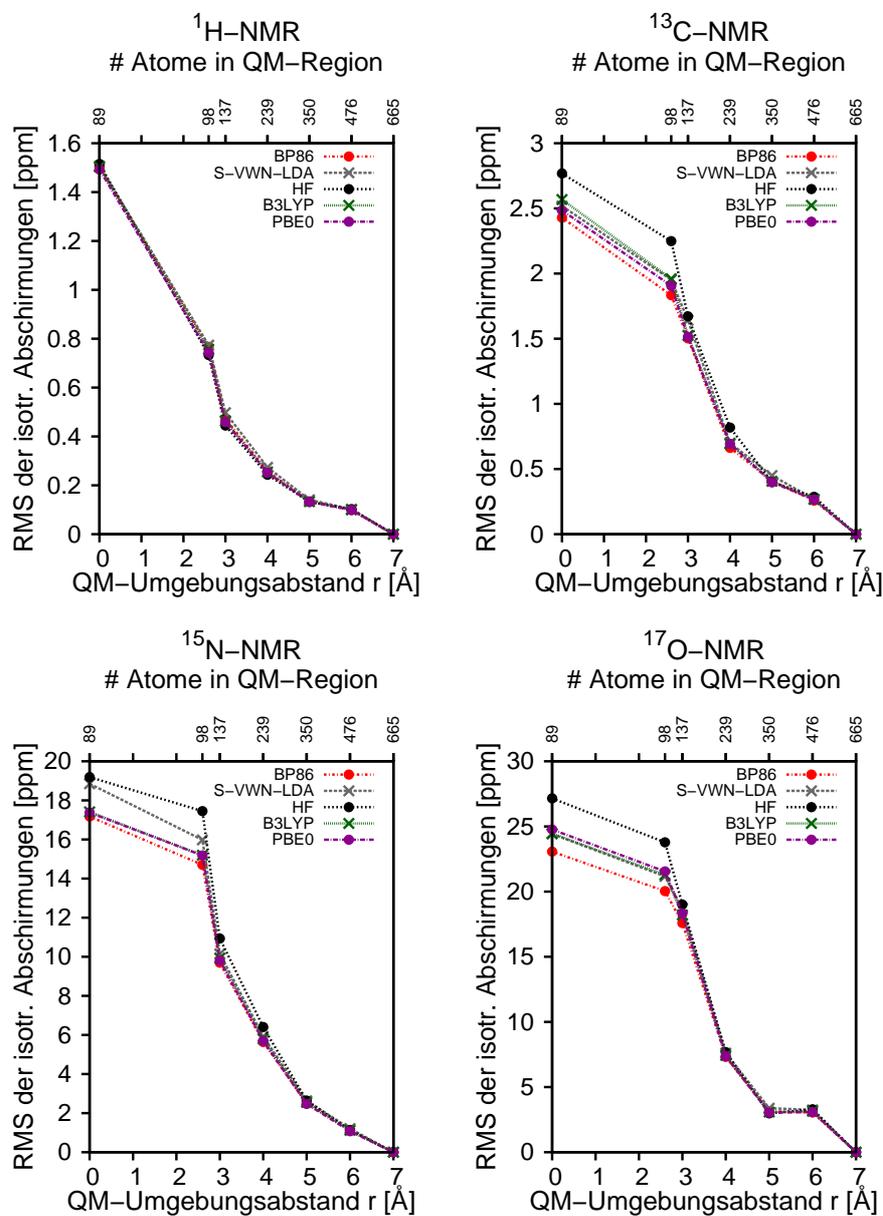


Abb. 5.7: Standardabweichungen (RMS) der isotropen Abschirmungen gegenüber dem QM-Ergebnis des größten QM-Umgebungsabstands (hier 7 Å) für verschiedene QM-Methoden über 44  $^1\text{H}$ -, 27  $^{13}\text{C}$ -, 9  $^{15}\text{N}$ - und 9  $^{17}\text{O}$ -Kerne, reine QM-Beschreibung: GIAO-HF/6-31G\*\*.

lässigung von Umgebungseinflüssen akzeptabel ist, erhält man einen Genauigkeitswert von 0.5 ppm als vernünftiges Beispielkriterium für  $^{13}\text{C}$ . Für  $^1\text{H}$ -Abschirmungen wird im Folgenden ein Genauigkeitswert von 0.1 ppm als Konvergenzschwelle diskutiert.

Neben der Auswertung anhand des Genauigkeitswerts von 0.1/0.5 ppm für  $^1\text{H}/^{13}\text{C}$ -Kerne, werden die Quotienten  $\text{RMS}^{\text{QM}}/\text{RMS}^{\text{QM/MM}}$  für alle QM-Größen berechnet und dann über alle betrachteten QM-Größen gemittelt (außer für die größte QM-Rechnung, um die Division durch null zu vermeiden). Auf diese Weise lässt sich der Gesamtvorteil von QM/MM gegenüber einer reinen QM-Beschreibung vergleichen. Der sich dadurch ergebende Einzelwert wird im Folgenden als mittlere Effizienz Zahl von QM/MM bezeichnet. Die mittlere Effizienz Zahl gibt also an, um welchen Faktor die Standardabweichung durch QM/MM im Vergleich zu einer reinen QM-Beschreibung typischerweise reduziert wird.

#### 5.4.5 Konvergenzergebnisse für die untersuchten Systeme

Die Abbildung 5.8a zeigt die Ergebnisse für das Aminopyrazol-Wasser-System. Die Standardabweichungen fallen unter 0.1/0.5 ppm für  $^1\text{H}/^{13}\text{C}$  bei einem QM-Umgebungsabstand von

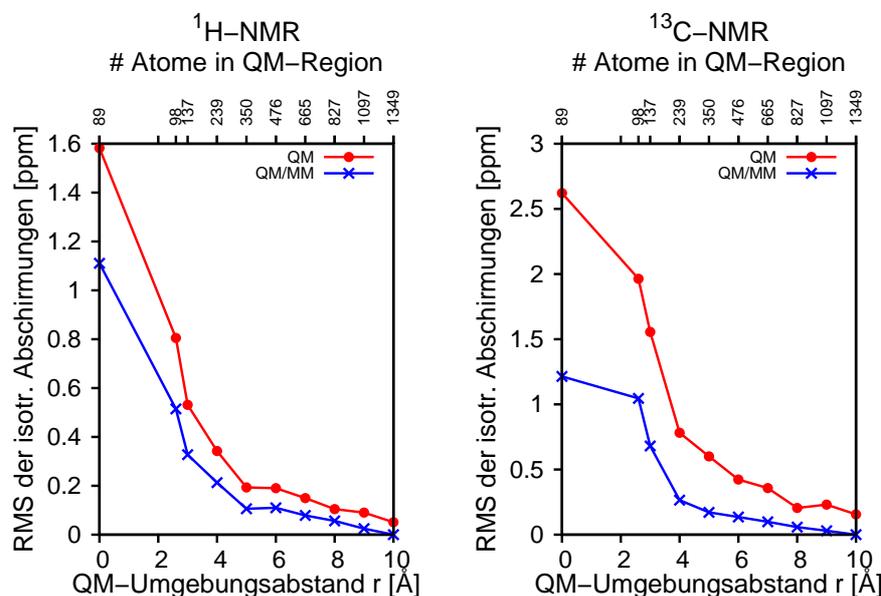


Abb. 5.8a: Standardabweichungen (RMS) der isotropen Abschirmungen gegenüber dem QM/MM-Ergebnis mit dem größtem QM-Umgebungsabstand für das Aminopyrazol-Wasser-System über 44  $^1\text{H}$ - und 27  $^{13}\text{C}$ -Kerne bei einer reinen QM-Beschreibung und QM/MM, GIAO-B3LYP/6-31G\*\*.

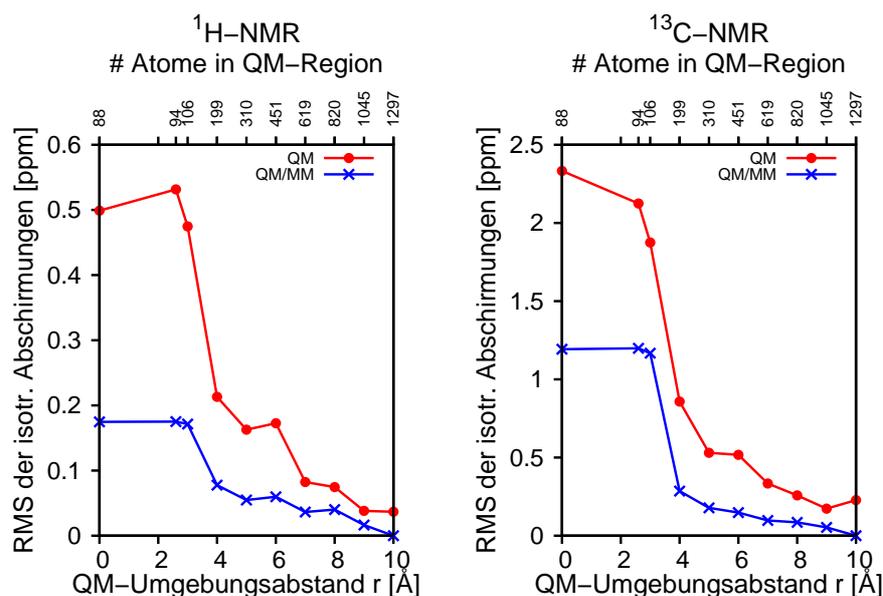


Abb. 5.8b: Standardabweichungen (RMS) der isotropen Abschirmungen gegenüber dem QM/MM-Ergebnis mit dem größtem QM-Umgebungsabstand für das wässrige System des molekularen Clips über 36  $^1\text{H}$ - und 41  $^{13}\text{C}$ -Kerne bei einer reinen QM-Beschreibung und QM/MM, GIAO-HF/6-31G\*\*.

9.0 Å (1097 Atome)/6.0 Å (476 Atome) im Falle einer reinen QM-Beschreibung im Vergleich zu 7.0 Å (665 Atome)/4.0 Å (239 Atome) im Falle von QM/MM. Die mittlere Effizienzzahl von QM/MM ist 1.9/3.4 für  $^1\text{H}/^{13}\text{C}$ . Für das Wirt-Gast-System des Naphthalin-Clips fasst die Abbildung 5.8b die Ergebnisse entsprechend einer analogen Auswertung zusammen. Die Standardabweichungen fallen unter 0.1/0.5 ppm für  $^1\text{H}/^{13}\text{C}$  bei einem QM-Umgebungsabstand von 7.0 Å (619 Atome)/7.0 Å (619 Atome) im Falle einer reinen QM-Beschreibung im Vergleich zu 4.0 Å (199 Atome)/4.0 Å (199 Atome) im Falle von QM/MM. Die mittlere Effizienzzahl von QM/MM ist 2.6/2.7 für  $^1\text{H}/^{13}\text{C}$ . Damit ist das allgemeine Bild für die Konvergenz in der wässrigen Umgebung ähnlich zum ersten Beispiel mit einer verglichen mit dem Aminopyrazol-System geringfügig schnelleren QM-Größenkonvergenz bei den Wasserstoffkernen, wenn man den Wert von 0.1 ppm als Konvergenzschwelle festlegt. Außerdem ist die mittlere Effizienzzahl von QM/MM geringfügig größer.

Die Abbildung 5.8c zeigt die Ergebnisse des Glutathion-Methanol-Systems. Die Standardabweichungen fallen unter 0.1/0.5 ppm für  $^1\text{H}/^{13}\text{C}$  bei einem QM-Umgebungsabstand von 7.0 Å (307 Atome)/5.0 Å (151 Atome) im Falle einer reinen QM-Beschreibung im Vergleich

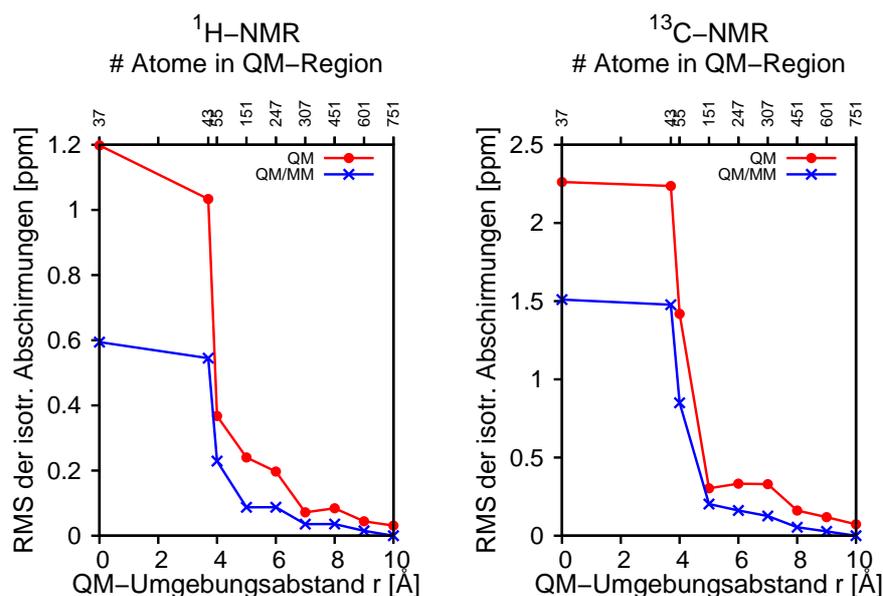


Abb. 5.8c: Standardabweichungen (RMS) der isotropen Abschirmungen gegenüber dem QM/MM-Ergebnis mit dem größtem QM-Umgebungsabstand für Glutathion-System in Methanol über 17  $^1\text{H}$ - und 10  $^{13}\text{C}$ -Kerne bei einer reinen QM-Beschreibung und QM/MM, GIAO-B3LYP/SVP.

zu 5.0 Å (151 Atome)/5.0 Å (151 Atome) im Falle von QM/MM. Die mittlere Effizienzzahl von QM/MM ist 2.2/2.3 für  $^1\text{H}/^{13}\text{C}$ . Betrachtet man die QM-Umgebungsabstände ähneln die Ergebnisse also denjenigen der wässrigen Systeme. Da allerdings die Wasserumgebung dichter als die Methanol-Umgebung ist (bei 300 K), umfasst ein bestimmter QM-Umgebungsabstand in Methanol weniger Atome als bei Wasser.

Die Ergebnisse des DNA-Enzym-Komplexes sind in Abbildung 5.8d zusammengestellt: Die Abweichungen fallen unter 0.1/0.5 ppm für  $^1\text{H}/^{13}\text{C}$  bei einem QM-Umgebungsabstand von 9.0 Å (1449 Atome)/9.0 Å (1449 Atome) im Falle einer reinen QM-Beschreibung im Vergleich zu 8.0 Å (1109 Atome)/3.2 Å (405 Atome) im Falle von QM/MM. Zusätzlich zum im Wasser gelösten DNA-Enzym-Komplex sind auch die Ergebnisse ohne zusätzliches Wasser eingezeichnet (siehe gepunktete Linien in Abbildung 5.8d). Bei dem betrachteten System bleibt das Konvergenzverhalten ähnlich, wenn man eine zusätzliche Wasserumgebung hinzufügt. Lediglich eine geringfügig schnellere Konvergenz der reinen QM-Rechnungen der größten Systeme jenseits von 9 Å ist für den Fall einer zusätzlichen Wasserumgebung zu beobachten. Für den kleineren Basissatz 3-21G wurde die QM-Region noch weiter vergrößert bis zu

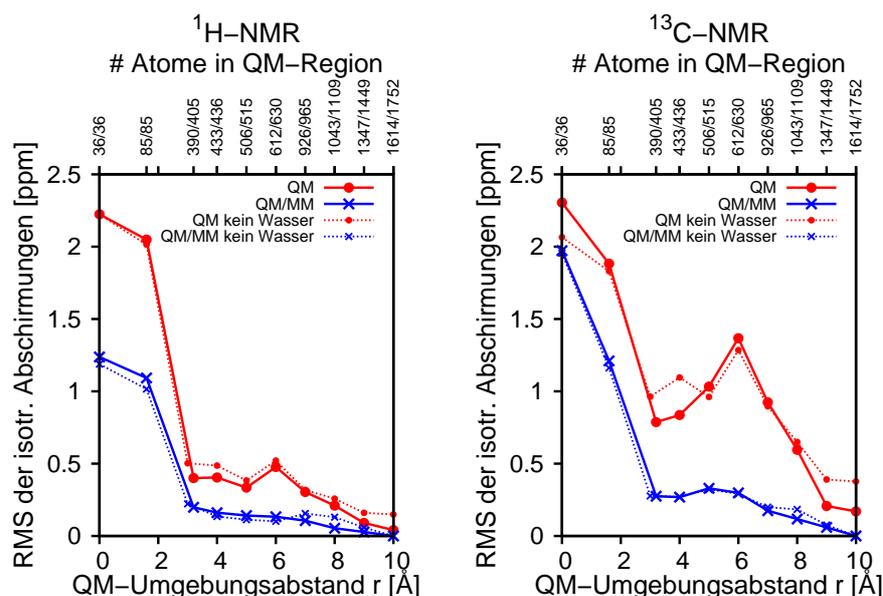


Abb. 5.8d: Standardabweichungen (RMS) der isotropen Abschirmungen gegenüber dem QM/MM-Ergebnis mit dem größtem QM-Umgebungsabstand für den DNA-Enzym-Komplex über 11  $^1\text{H}$ - und 10  $^{13}\text{C}$ -Kerne bei einer reinen QM-Beschreibung und QM/MM, GIAO-HF/6-31G\*\*. Die gepunkteten Linien beziehen sich auf Berechnungen ohne zusätzliche Wasserumgebung (nur Kristallwasser, siehe Abschnitt 5.4.3).

13.0 Å (2379 Atome), wo die  $^1\text{H}$ -Standardabweichung schließlich auch für den Fall einer reinen QM-Beschreibung und keiner Wasserumgebung unter 0.1 Å fällt (siehe Anhang C). Die mittlere Effizienzanzahl von QM/MM ist 2.7/3.4 für  $^1\text{H}/^{13}\text{C}$ .

Die Erkenntnisse mit Hinblick auf die QM/MM-Vorteile verändern sich stark für den letzten untersuchten Testfall (Abbildung 5.8e): Die Einbeziehung der elektrostatischen Wechselwirkungen durch QM/MM beeinflusst hier kaum die Ergebnisse einer reinen QM-Beschreibung. In Anbetracht dieser und der vorherigen Ergebnisse kann man schlussfolgern, dass in dem HBC-System die Umgebungseinflüsse durch andere als elektrostatische Wechselwirkungen bestimmt sind und die man mit QM/MM nicht beschreiben kann (zumindest nicht mit Standard-MM-Verfahren). Im HBC-System sind die Umgebungseinflüsse offensichtlich durch die aromatischen Ringströme bestimmt, dessen langreichweitiges Verhalten bereits in Referenzen [121, 130, 131] beschrieben wurde.

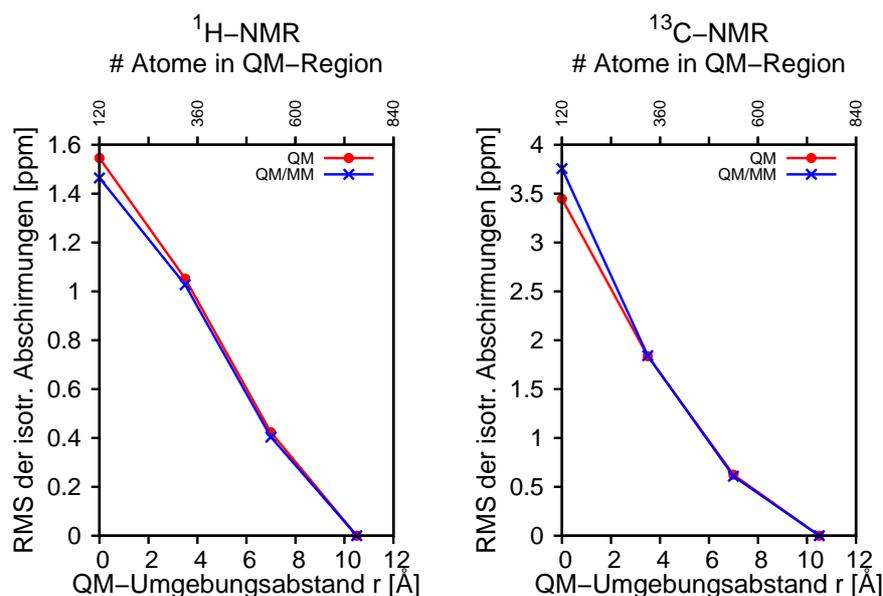


Abb. 5.8e: Standardabweichungen (RMS) der isotropen Abschirmungen gegenüber dem QM/MM-Ergebnis mit dem größtem QM-Umgebungsabstand für das Hexa-peri-hexabenzocoronon-System über 42 <sup>1</sup>H- und 87 <sup>13</sup>C-Kerne bei einer reinen QM-Beschreibung und QM/MM, GIAO-HF/6-31G\*\*. Die Punkte bei 3.5 Å, 7.0 Å and 10.5 Å beziehen sich auf das Trimer, Pentamer bzw. Heptamer.

## 5.5 Zusammenfassung und Ausblick

Die dargestellte Genauigkeitsuntersuchung von NMR-Verschiebungen auf GIAO-SCF- und MP2-Niveau erfasst Methoden- und Basissatzfehler basierend auf einem breit angelegten molekularen Testsatz. Die Genauigkeitsuntersuchung trägt dazu bei, um möglichst umfassend und zugleich zuverlässig NMR-Signale bei künftigen ab-initio Rechnungen zuordnen zu können. Gemäß der Untersuchung liegen die SCF-Methodenfehler (ermittelt als Standardabweichungen gegenüber den genauen CCSD(T)-Referenzen bei einem großen Basissatz) zwischen 0.16/3.2 und 0.29/8.8 ppm für <sup>1</sup>H/<sup>13</sup>C. Insbesondere das Hybrid-GGA-Funktional B97-2 tritt dabei unter den getesteten Funktionalen für <sup>13</sup>C-Verschiebungen hervor, auch für kleinere Basissätze bis hin zum SVP-Niveau. Die MP2-Methode erreicht durchgängig die besten Standardabweichungen von 0.12/2.1 ppm für <sup>1</sup>H/<sup>13</sup>C.

Im Anschluss an die Genauigkeitsuntersuchung entwickelt und prüft das vorangehende Kapitel das Konzept intermediärer Referenzen bei der Berechnung von NMR-Verschiebungen

am Beispiel von Polypeptiden. So stellt das Kapitel Verschiebungswerte gegenüber TMS auf MP2-Niveau und die zugehörigen HF-Abschirmungen basierend auf einem großen Basissatz für alle isolierten Aminosäuren (den intermediären Referenzen) bereit. Darauf aufbauend werden NMR-Verschiebungsrechnungen für Aminosäurereste in realen Polypeptidsysteme mit einer erheblich erhöhten Genauigkeit möglich im Vergleich zu reinen SCF-Rechnungen, wie eine statistisch aussagekräftige Untersuchung zeigt.

Die abschließend vorgestellten Konvergenz-Untersuchungen liefern Einblicke, wie groß die QM-Region gewählt werden muss, um mit einer bestimmten Genauigkeit NMR-Verschiebungen durch quantenchemische Methoden zu berechnen. Ausgehend von einer zentralen Molekülregion und gegebenen Konformationen für verschiedene chemische Systeme zeigen die Untersuchungen, dass:

- reine QM-Berechnungen typischerweise 6–10 Å der Umgebung um den betrachteten Kern einschließen müssen, um mittlere Genauigkeiten unterhalb von 0.1/0.5 ppm für  $^1\text{H}/^{13}\text{C}$  zu erreichen
- deshalb ausgehend von kleinen zentralen Molekülregionen ungefähr 300–1200 Atome für explizite Beschreibungen von Umgebungseinflüssen mittels reinen QM-Rechnungen für  $^1\text{H}/^{13}\text{C}$  eingeschlossen werden müssen
- der mittlere Vorteil von konventionellen QM/MM-Verfahren (elektrostatische Einbettung basierend auf externen Punktladungen) gegenüber einer reinen QM-Beschreibung überwiegend überzeugt und sich in Systemen mit bestimmenden elektrostatischen Einflüssen als ziemlich konstant erweist (mittlere Effizienzzahlen von QM/MM liegen ungefähr zwischen 2 und 4). Trotzdem sind, abhängig vom chemischen System, QM-Regionen mit 200–1000 Atomen für Genauigkeiten von 0.1/0.5 ppm für  $^1\text{H}/^{13}\text{C}$  erforderlich
- deutlich kleinere mittlere Vorteile von QM/MM auftreten, wenn insbesondere andere als elektrostatische Einflüsse dominieren, z.B., aromatische Ringströme
- bei einzelnen Kernen können die QM/MM-Vorteile um einiges größer sein, allerdings wurde auch in sehr wenigen Fällen eine Verschlechterung der Einzelergebnisse aufgrund von QM/MM im Vergleich zu einer reinen QM-Beschreibung für die kleinsten QM-Regionen beobachtet.



## Kapitel 6

### Enzymatische Reparatur oxidativer DNA-Schäden

#### 6.1 Einführung

Die Desoxyribonukleinsäure (DNA), Trägerin des Erbguts in allen Lebewesen, ist als Makromolekül fortwährend Zerfallsreaktionen unterworfen. Insbesondere können Einflüsse wie Wärme, Strahlung oder reaktive Stoffe die DNA schädigen. DNA-Schäden können beispielsweise zur Änderung der Basen-Paarung führen, so dass bei der DNA-Replikation oder -Transkription eine fehlerhafte Information weitergegeben wird. In den Lebewesen sind daher effiziente Reparaturprozesse für DNA-Schäden von zentraler Bedeutung, um mögliche Folgen der Änderung genetischer Informationen (z.B. bis hin zu Krebserkrankungen) möglichst zu verhindern (siehe z.B. Referenz [132]).

Die häufigsten oxidativen und Purin-basierten DNA-Schäden sind in Abbildung 6.1 gezeigt, das 7,8-Dihydro-8-oxoguanin (8-oxoG) und das Formamidopyrimidin-2'-guanosin (FaPyG). Ein wichtiger Reparaturweg für die oxidativen Schäden ist die Basenexzisionsreparatur (BER), die durch spezifische Reparaturenzyme eingeleitet wird, den DNA-Glykosylasen (für

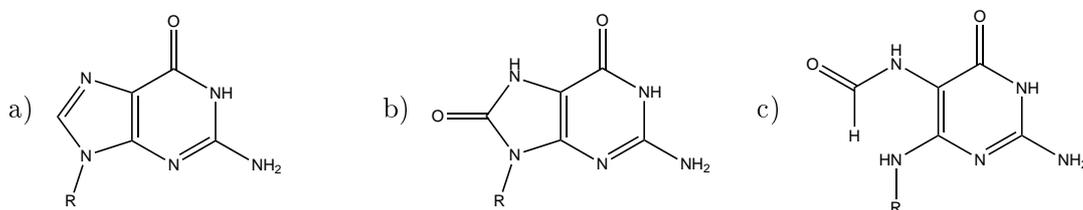


Abb. 6.1: Strukturformeln der DNA-Base Guanin (a) sowie der häufigen oxidativen Schäden 7,8-Dihydro-8-oxoguanin (8-oxoG, b) und Formamidopyrimidin-2'-guanosin (FaPyG, c) (Nukleobase: R=H, Nucleosid: R= Desoxyribose, Nucleotid: R=Desoxyribose-Phosphat).

eine Übersicht über weitere Reparaturwege siehe z.B. Referenz [133]). Bei der Basenexzisionsreparatur werden die oxidativ geschädigten Basen aus der DNA herausgetrennt, indem die glykosidische Bindung gespalten wird. Die Stelle der fehlenden Base wird in Folgeschritten der BER dann an Endonukleasen, einer Polymerase und Ligase weiter aufbereitet und schließlich die Ausgangs-DNA-Sequenz wiederhergestellt.

DNA-Glykosylasen (*Enzyme Commission Number* 3.2.2) können grob in vier strukturelle Superfamilien eingeteilt werden [134], in die Uracil-Glykosylasen (UDG), Alkyladenin-Glykosylasen (AAG), die Helix-Hairpin-Helix-Glycin-Prolin-Aspartat-Glykosylasen (HhH-GPD) und in die MutM/Formamidopyrimidin-Glykosylasen (Fpg), wobei im Fokus der vorgestellten Untersuchungen bakterielles MutM steht.

Schon der erste Schritt innerhalb der BER ist ein sehr komplexer Vorgang, der trotz einer Vielzahl experimenteller sowie theoretischer Untersuchungen (siehe z.B. Referenzen [56, 135–145]) zahlreiche offene Fragestellungen birgt, wovon einige wichtige Aspekte in einem Kooperationsprojekt mit den Gruppen Carell und Zacharias [146] untersucht werden. So ist beispielsweise bislang nicht vollständig geklärt, wie die Glykosylase die geschädigte Base aus einer Vielzahl ungeschädigter Basen erkennt, sie in der aktiven Tasche stabilisiert und nach welchem Mechanismus die Spaltung der glykosidischen Bindung abläuft. Im Rahmen dieser Arbeit soll anhand von QM- und QM/MM-Berechnungen ein Beitrag geliefert werden, um

- den Unterschied der Wechselwirkungsenergie von intakter und geschädigter DNA mit der Protein-Umgebung zu quantifizieren
- Enzymresiduen zu identifizieren, die den Schaden gegenüber der intakten Base bevorzugt binden (oder aber auch benachteiligen)

Dabei liegt für beide Fragestellungen im nachfolgenden Kapitel der Schwerpunkt insbesondere auf Aspekten der theoretischen Beschreibung:

- Wie groß muss die QM-Region innerhalb eines QM/MM-Ansatzes für verlässliche Aussagen gewählt werden?
- Was sind die Einflüsse verschiedener QM-Methoden und Basissätze?

Die theoretischen Untersuchungen sollen insbesondere als Grundlage weiterführender Untersuchungen innerhalb des Kooperationsprojektes [146] dienen.

## 6.2 Einzelheiten zur Strukturauswahl und -vorbereitung

Die nachstehenden QM- und QM/MM-Untersuchungen gründen auf einer experimentellen Kristallstruktur eines MutM-DNA-Komplexes mit einem einzelnen oxidativen Schaden in der aktiven Tasche des Enzyms (PDB code 1R2Y, [56]). Die Anordnung der in der experimentellen Kristallstruktur enthaltenen Atomlagen wurden hier weitestgehend beibehalten. Die Frage nach dem konformationellen Einfluss wird hier also zunächst ausgeklammert und im Rahmen des Kooperationsprojektes [146] weiterverfolgt, einerseits durch Auswahl anderer Kristallstrukturen als Ausgangspunkt [139] und andererseits durch molekulardynamische Rechnungen (siehe Kapitel 2.2 und 4).

Allerdings erfordert die experimentelle Kristallstruktur eine wesentliche Modifikation: Um die Kristallmessung zu ermöglichen wurde experimentell der Protein-Rest Glu-2 zu einem Gln-2 mutiert [56]. Die künstliche Mutation wurde im ersten Schritt der Strukturvorbereitung rückgängig gemacht, also die Gln- wieder durch eine Glu-Seitenkette ersetzt (wie die folgenden Schritte auf der Grundlage der Strukturdaten und Prozeduren des Amber-Kraftfeldpakets mit Parametersatz FF10 [53]). Im zweiten Schritt wurden die Wasserstoffatome hinzugefügt mit den Standardprotonierungs-Zuständen (geladener C- und N-Terminus; Asp, Glu deprotoniert; Arg, Lys protoniert, His: neutral). Außerdem wurden zum Ladungsausgleich  $\text{Na}^+$ -Gegenionen hinzugefügt und schließlich eine Sphäre von 15 Å an  $\text{H}_2\text{O}$ -Lösungsmittelmolekülen. Die Atomlagen der hinzugefügten Atome wurden dann mittels der Amber-Kraftfeld-Beschreibung optimiert (mithilfe des NAMD-Pakets [126], 100000 Optimierungsschritte). Für eine möglichst genaue Beschreibung der geschädigten 8-oxoG-Nukleobase sowie des Glu-2-Rests schloss sich eine QM/MM-Optimierung an (B97-2-D3/SVP [44, 106, 147], aktive Region: Glu-2-Rest und 8-oxoG-Base, QM-Region: Glu-2-Aminosäure und 8-oxo-G-Nukleotid, elektrostatische Einbettung mithilfe der ChemShell-Suite [79], siehe Abschnitte 2.6 und 4.4). Aus der erzielten Struktur mit 8-oxo-G-Schaden wurde schließlich auch die Struktur mit intakter Nukleobase erzeugt. Dazu wurde das O8-Atom des Schadens durch ein Wasserstoffatom ersetzt, das H7-Atom entfernt und anschließend die intakte Guanin-Base mittels einer QM/MM-Rechnung erneut optimiert (Methode s.o., aktive Region: G-Base, QM-Region: G-Nukleotid).

Die Abbildung 6.2 zeigt die Strukturen (ohne Wassermoleküle und  $\text{Na}^+$ -Gegenionen), in denen der Schaden bzw. die intakte Base gemäß der Strukturvorbereitung jeweils in der gleichen Orientierung in der aktiven Bindungstasche des Reparatur-Enzyms MutM liegen. Das Ge-

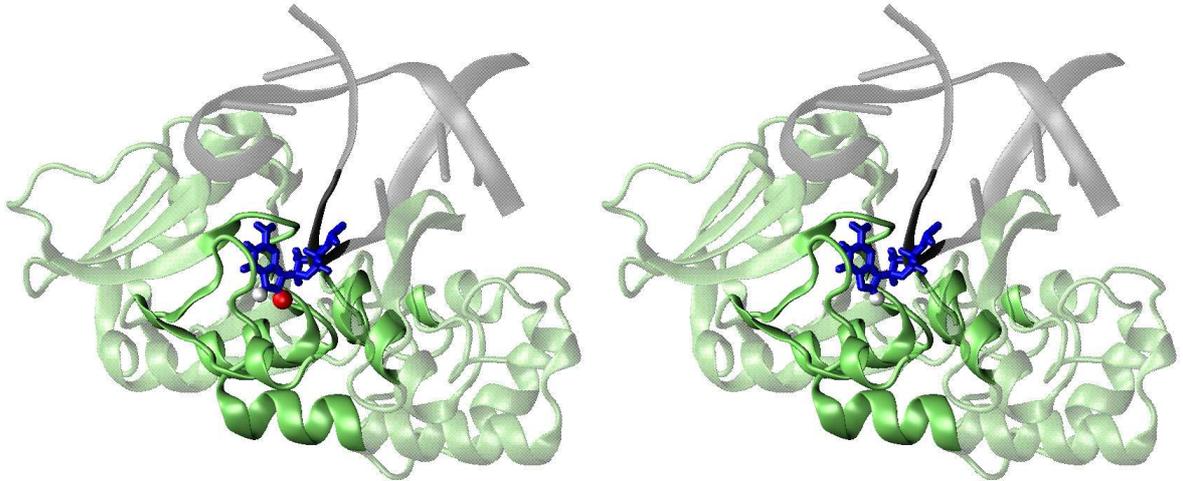


Abb. 6.2: Strukturen des DNA-Enzym-Komplexes [56], links mit 8oxo-G-Schaden (O8-Atom: rot, H7-Atom: weiß) und rechts mit intakter G-Base (H8-Atom: weiß), schematisch gezeigt ohne Wassermoleküle und  $\text{Na}^+$ -Gegenionen.

samtsystem aus Protein, geschädigter DNA und Wasserumgebung besteht aus 62213 Atomen, die Protein-Wasserumgebung aus 61453 Atomen und das DNA-Fragment aus 760 Atomen.

### 6.3 Gesamte Schadensstabilisierung

Gemäß des supermolekularen Ansatz ergibt sich die Wechselwirkung ( $W$ ) zwischen DNA ( $N$ ) und Proteinumgebung ( $P$ ) als Differenz der DNA-Protein-Gesamtenergie ( $NP$ ) und den Fragmentenergien, sowohl für die geschädigte DNA (8OG) als auch die intakte DNA ( $G$ ):

$$\begin{aligned}
 W_{N_{8OG}P} &= E_{N_{8OG}P} - E_{N_{8OG}} - E_P \\
 W_{N_G P} &= E_{N_G P} - E_{N_G} - E_P \\
 \hline
 \Delta W_{NP} &= E_{N_{8OG}P} + E_{N_G} - E_{N_{8OG}} - E_{N_G P}
 \end{aligned} \tag{6.1}$$

Die Wechselwirkungsdifferenz  $\Delta W_{NP} = W_{N_{8OG}P} - W_{N_G P}$  sagt aus, ob oder in welchem Ausmaß das Enzym die geschädigte DNA gegenüber der intakten DNA bevorzugt bindet. Gemäß der Definition in Gleichung 6.1 bedeutet dabei ein negatives Vorzeichen von  $\Delta W_{NP}$  eine bevorzugte Bindung der geschädigten DNA. Die Wechselwirkungen  $W_{N_{8OG}P}$  sowie  $W_{N_G P}$  (und somit auch die Wechselwirkungsdifferenz  $\Delta W_{NP}$ ) können anstatt gemäß des supermolekularen Ansatzes (Gleichung 6.1) prinzipiell auch mithilfe der *Symmetry-Adapted Perturbation Theory* (SAPT,

siehe z.B. Referenz [148]) berechnet werden.

Für eine QM-Beschreibungen liegt die hauptsächliche Herausforderung bei der Quantifizierung von  $\Delta W_{\text{NP}}$  in den Systemgrößen (siehe vorangehender Abschnitt), die eine vollständig quantenmechanische Behandlung der Einzelfragmente in einem machbaren Zeitrahmen ausschließen und auch die Basissatzgröße und methodische Genauigkeit limitieren. Um zu einer möglichst zuverlässigen aber gleichzeitig effizienten Quantifizierung von  $\Delta W_{\text{NP}}$  zu gelangen, werden im Folgenden die unvermeidbaren Näherungen bei der Berechnung von  $\Delta W_{\text{NP}}$  untersucht:

- Basissatzfehler, hier insbesondere der Basissatz-Superpositionsfehler (*basis set superposition error*, BSSE)
- QM-Methoden-Fehler durch Hartree-Fock oder KS-DFT-Methoden
- Fehler bei der Einschränkung der QM-Größe

Ziel der systematischen Untersuchung der einzelnen Fehler ist neben einer Quantifizierung von  $\Delta W_{\text{NP}}$  auch, für das Kooperationsprojekt [146] zuverlässige methodische Eckpunkte für analoge Fragestellungen und Strukturen zu gewinnen.

Am Beispiel eines Systemausschnitts mit QM-Umgebungsabstand von 5 Å (Definition siehe unten) zeigt die Tabelle 6.1 den Einfluss der Basissatzgröße auf die berechnete Wechselwirkungsdifferenz  $\Delta W_{\text{NP}}$ , für den Fall einer SVP-Basis [106] gegenüber der größeren cc-pVTZ-Basis [62], jeweils ohne und mit einer Counterpoise-Korrektur [149] des BSSE-Fehlers. Insbesondere auf SVP-Niveau erweist sich für die Wechselwirkungsdifferenz  $\Delta W_{\text{NP}}$  eine Counterpoise-Korrektur als wichtig, womit man hier auf etwa  $4 \frac{\text{kJ}}{\text{mol}}$  an den Counterpoise-korrigierten cc-pVTZ-Wert gelangt.

Die Tabelle 6.2 führt den methodischen Fehler einer Beschreibung auf HF-Niveau oder mittels B3LYP [41, 42] oder B97-2 [44] auf. Als Referenz dient hier eine SAPT-Rechnung auf Basis der *Scaled-opposite-spin*-Methode (SOS, [150]) und einer neuen linear-skalierenden Implementierung von S. A. Maurer und M. Beer et al. [50, 151]. Darüber hinaus führt die Tabelle die empirische Dispersionskorrektur nach Grimme et al. [147] auf. Die kleinste Abweichung vom SOS-SAPT-Wert ergibt sich für die supermolekulare B97-2-D3-Rechnung ( $5.5 \frac{\text{kJ}}{\text{mol}}$ ), dicht gefolgt von B3LYP-D3 ( $5.6 \frac{\text{kJ}}{\text{mol}}$ ). Die Dispersions-korrigierte HF-Rechnung (HF-D3) liefert einen etwas höheren Fehler ( $7.4 \frac{\text{kJ}}{\text{mol}}$ ).

Tab. 6.1: Einfluss der Basissatzgröße mit und ohne Counterpoise-Korrektur (CP, [149]) auf die Wechselwirkungsdifferenz  $\Delta W_{\text{NP}}$   $\Delta W_{\text{NP}}$  (in  $\frac{\text{kJ}}{\text{mol}}$ ) von intakter und geschädigter DNA mit der Proteinumgebung am Beispiel der Basis SVP [106] und cc-pVTZ [62], reine QM-Beschreibung, HF, QM-Umgebungsabstand 5 Å.

	SVP	cc-pVTZ	Differenz
ohne Korrektur	-25.7	-16.6	9.1
CP-Korrektur	8.3	3.2	-5.1
gesamt	-17.4	-13.4	4.0

Tab. 6.2: Einfluss der QM-Theorie auf die Wechselwirkungsdifferenz  $\Delta W_{\text{NP}}$  von intakter und geschädigter DNA mit der Proteinumgebung (in  $\frac{\text{kJ}}{\text{mol}}$ ) für das Beispiel von HF, B3LYP [41, 42] und B97-2 [44], jeweils mit und ohne Dispersionskorrektur (D3-Korrektur) nach Grimme et al. [147], reine QM-Beschreibung, QM-Umgebungsabstand 5 Å, Basissatz SVP [106], im Falle des supermolekularen Ansatzes mit Counterpoise-Korrektur [149].

	HF	B3LYP	B97-2	
ohne D3-Korrektur	-17.4	-21.9	-18.2	SOS-SAPT
D3-Korrektur	-17.2	-10.9	-14.5	
gesamt	-34.6	-32.8	-32.7	-27.2

Die Fehler bei der Einschränkung der QM-Größe werden bestimmt, indem ausgehend von einem Zentrum die QM-Größe bis zur Konvergenz der Ergebnisse vergrößert wird, wobei einerseits die weitere Umgebung komplett vernachlässigt wird (als reine QM-Beschreibung bezeichnet) oder aber als MM-Teilsystem berücksichtigt wird, gemäß einer elektrostatischen QM/MM-Einbettung (siehe Abbildung 5.2 und Abschnitt 2.6). Die Gesamtwechselwirkungsenergien ( $W_{\text{N}_{\text{SOGP}}}$  bzw.  $W_{\text{N}_{\text{GP}}}$ ) an sich können für das vorliegende System nur mit einem immensen Aufwand mit der QM-Größe konvergiert werden, da die DNA und das Enzym über einen Bereich wechselwirkt, der sich über das gesamte System erstreckt. So führt eine sukzessive Vergrößerung ausgehend vom gesamten Wechselwirkungsbereich sehr schnell zu sehr großen QM-Regionen. Zum Beispiel liegen 2675/3356/4105/4884 Atome sowohl näher als 5/6/7/8 Å zu einem beliebigen DNA-Nukleotid als auch zu einem beliebigen Residu-

um der Proteinumgebung, d.h. an der unmittelbaren Grenzfläche zwischen DNA und Proteinumgebung. Davon abweichend ist zu erwarten, dass man für die Wechselwirkungsdifferenz  $\Delta W_{\text{NP}} = W_{\text{N}_{8\text{oxG}}\text{P}} - W_{\text{N}_{\text{GP}}}$  mit weitaus geringerem Aufwand zur QM-Größen-Konvergenz gelangen kann. Zu diesem Zweck wurde für die folgenden Untersuchungen die QM-Region nach und nach von der lokalen Stelle aus vergrößert, in der sich das 8-oxo-G- und das G-System unterscheiden (der QM-Umgebungsabstand bezieht sich also auf die O8-C8-Bindung des 8-oxoG-Schadens).

Nichtsdestotrotz zeigt sich, dass die Umgebung bis zu etwa 10 Å um die O8-C8-Bindung einen maßgeblichen Einfluss auf die Wechselwirkungsdifferenz  $\Delta W_{\text{NP}}$  hat, wie die schrittweise Vergrößerung der Fragmente und Vernachlässigung der restlichen Umgebung in einer reinen QM-Beschreibung zeigt (siehe Abbildung 6.3). Demgegenüber umfasst die reine MM-Rechnung zwar die Gesamtfragmente (komplettes System bzw. komplette Proteinumgebung und DNA), allerdings macht sich hierbei ein sehr großer methodischer Fehler bemerkbar. Die

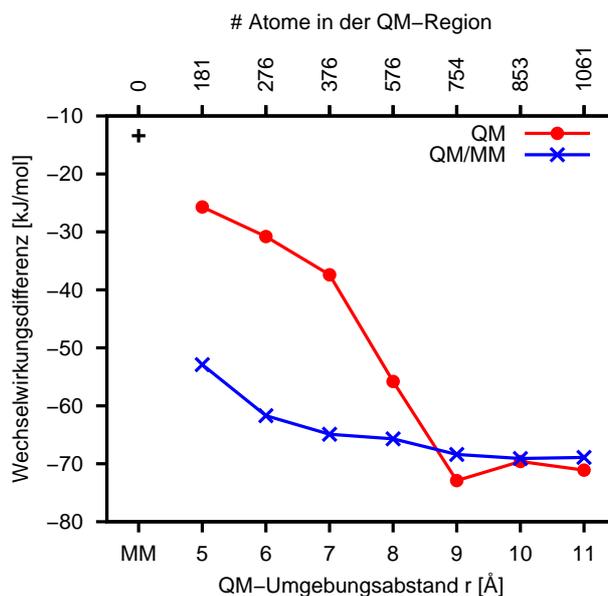


Abb. 6.3: Konvergenz der Wechselwirkungsdifferenz  $\Delta W_{\text{NP}}$  (in  $\frac{\text{kJ}}{\text{mol}}$ ) von intakter und geschädigter DNA (N) mit der Proteinumgebung (P), HF/SVP, ohne Dispersionkorrektur nach Grimme et al. ([147], D3) und ohne Counterpoise-Korrektur ([149], CP). Gezeigt ist die Konvergenz mit der QM-Größe der reinen QM-Beschreibung im Vergleich zu QM/MM. Zudem ist der Wert einer reinen molekularmechanischen Beschreibung (MM) angegeben.

QM/MM-Rechnungen, bei denen stets die komplette Umgebung berücksichtigt wird und sich nur die Zusammensetzung aus QM- und MM-Teilsystem ändert, zeigen gegenüber der reinen QM-Beschreibung ein deutlich schnelleres Konvergenzverhalten. Dennoch sind auch für die QM/MM-Rechnungen große QM-Bereiche nötig (im Bereich von 6 Å für Fehler unter  $10 \frac{\text{kJ}}{\text{mol}}$ ). Bei 10 Å (853 Atome im Gesamtfragment) liegt der Unterschied zwischen der reinen QM-Beschreibung und QM/MM dann unterhalb von  $1 \frac{\text{kJ}}{\text{mol}}$  und die Konvergenz ist nach diesem Kriterium erreicht. Der Anhang D listet die Zusammensetzung der QM-Region bei den verschiedenen QM-Umgebungsabständen auf.

Die abschließende Abbildung 6.4 illustriert die Abhängigkeit der Counterpoise-Korrektur [149] und der Dispersions-Korrektur nach Grimme et al. [147] von der Systemgröße. Im gesamten untersuchten Bereich erweist sich die Counterpoise-Korrektur als nahezu konstant. Das Verhalten der empirischen Dispersions-Korrektur legt nahe, dass Elektronen-Korrelationseffekte bei der Berechnung von  $\Delta W_{\text{NP}}$  im Bereich von 5-6 Å abklingen (eine strengere Überprüfung mittels aufwändigeren SAPT-Rechnungen wird an dieser Stelle zunächst ausgeklammert und im Kooperationsprojekt [146] weiterverfolgt). Die Tabelle 6.3 fasst die endgültige Aussage der vorangehenden Abbildungen und Tabellen zusammen. Darin sind die Abweichung zum Counterpoise-korrigierten cc-pVTZ-Wert und die Abweichung zum SOS-SAPT-Wert basierend auf einem QM-Umgebungsabstand von 5 Å extrapoliert, während die Counterpoise-korrigierte HF-D3-Werte exakt bei einer 10 Å-Umgebung bestimmt wurden (siehe obige Diskussion). Demnach bindet die schadhafte gegenüber der intakten DNA um etwa  $68 \frac{\text{kJ}}{\text{mol}}$  bevorzugt an das Enzym. Dass das Enzym bei einer analogen Bindung die schadhafte DNA so stark bevorzugt, liefert einen möglichen Hinweis, in welcher Weise das Enzym beim Reparaturprozess zwischen 8-oxoG und der intakten Nukleobase differenziert.

## 6.4 Individuelle Beiträge der Enzymresiduen zur Schadensstabilisierung

Im Fokus des vorangehenden Abschnitts steht die Gesamtwechselwirkungsdifferenz der geschädigten bzw. intakten DNA mit der Proteinumgebung. Zwar kann man aus der Abbildung 6.3 in Verbindung mit der detaillierten Auflistung in Anhang D bereits erste Rückschlüsse ziehen, welche Gruppen an Proteinresiduen hauptsächlich zwischen dem Schaden und der intakten Base differenzieren. Nach jedem einzelnen Proteinrest aufzuschlüsseln ist jedoch so nicht direkt möglich. Eine Möglichkeit dazu führt über das Schema 6.5. Darin subtrahiert man im

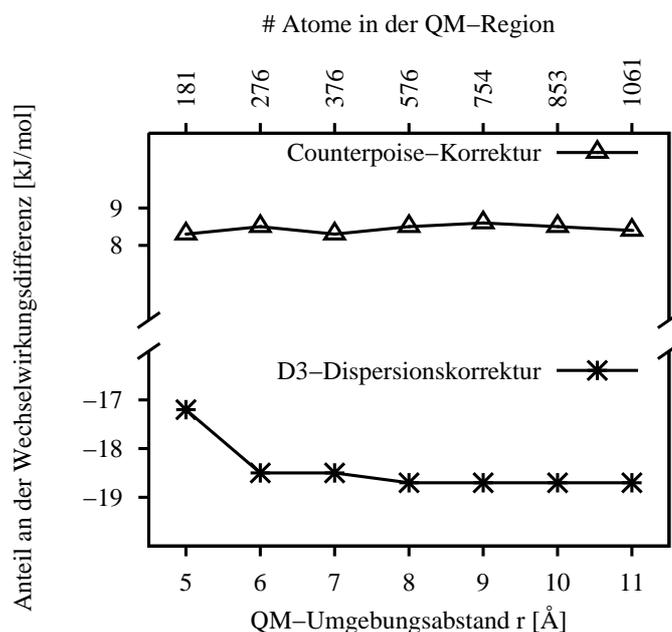


Abb. 6.4: GröÙen-Konvergenz der Counterpoise- und D3-Dispersionskorrektur für die Wechselwirkungsdifferenz  $\Delta W_{\text{NP}}$  (in  $\frac{\text{kJ}}{\text{mol}}$ ) von intakter und geschädigter DNA mit der Proteinumgebung in einer reinen QM-Beschreibung (HF/SVP bzw. D3-Korrektur nach Grimme et al. [147]).

Tab. 6.3: Extrapolierter Gesamtwert (in  $\frac{\text{kJ}}{\text{mol}}$ ) für die Wechselwirkungsdifferenz  $\Delta W_{\text{NP}}$  von intakter und geschädigter DNA mit der Proteinumgebung, die Längenangabe in Å bezeichnet den QM-Umgebungsabstand; CP: Counterpoise-Korrektur [149], D3: Dispersions-Korrektur nach Grimme et al. [147]; SOS-SAPT: *Scaled-opposite-spin Symmetry-Adapted Perturbation Theory* [50, 148, 150, 151].

HF, reine QM-Beschreibung, 10 Å, SVP	siehe Abbildung 6.3	-69.6
CP-Korrektur, 10 Å, SVP	siehe Abbildung 6.4	8.5
D3-Korrektur, 10 Å	siehe Abbildung 6.4	-18.7
Abweichung zu CP/cc-pVTZ, 5 Å	siehe Tabelle 6.1	4.0
Abweichung von HF-D3 zu SOS-SAPT, 5 Å	siehe Tabelle 6.2	7.4
extrapolierter Gesamtwert		-68.4

$W_{N,P}$	$=$	$E_{tot}$	$-E_N$	$-E_P$		gesamt
$- W_{N\setminus L,P}$	$=$	$E_{tot\setminus L}$	$-E_{N\setminus L}$	$-E_P$		ohne Nukleotid( $L$ )
$W_{L,P}$	$=$	$E_{tot}$	$-E_N$	$-E_{tot\setminus L}$	$+E_{N\setminus L}$	Beitrag von $L$
$- W_{L,P\setminus R_i}$	$=$	$E_{tot\setminus R_i}$	$-E_N$	$-E_{tot\setminus(L,R_i)}$	$+E_{N\setminus L}$	ohne Rest $i$ ( $R_i$ )
$W_{L,R_i}$	$=$	$E_{tot}$	$-E_{tot\setminus L}$	$-E_{tot\setminus R_i}$	$+E_{tot\setminus(L,R_i)}$	$R_i - L -$ Beitrag

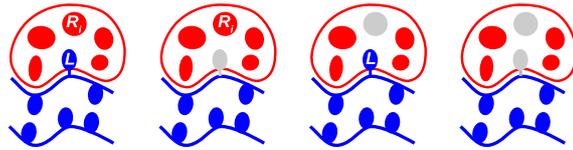


Abb. 6.5: Schema zur Herleitung des Wechselwirkungsbeitrags  $W_{L,R_i}$  eines DNA-Nukleotids  $L$  mit dem Protein-Rest  $R_i$ ;  $N$ : DNA (blau);  $P$ : Protein (rot); die Schreibweise „ $\setminus$ “ bedeutet „in Abwesenheit von“.

ersten Schritt von der tatsächlichen Gesamtwechselwirkung  $W_{N,P}$  zunächst die Wechselwirkungsenergie  $W_{N\setminus L,P}$  bei Abwesenheit des Nukleotids  $L$  und gelangt so zum Beitrag von  $L$  an der Gesamtwechselwirkung. Im zweiten Schritt wird vom verbliebenen Ausdruck analog der Term  $W_{L,P\setminus R_i}$  bei Abwesenheit des Rests  $R_i$  subtrahiert. In der Abbildung 6.5 sind die Systeme schematisch gezeigt, deren absolute Energiewerte für den Wechselwirkungsbeitrag des Rests  $R_i$  mit dem Nukleotid  $L$  schließlich benötigt werden: das gesamte System, das System jeweils ohne  $L$  oder  $R_i$ , sowie das System ohne  $L$  und  $R_i$ .

Da sowohl das Nukleotid  $L$  als auch die Proteinresiduen  $R_i$  kovalent gebunden sind, ist es unausweichlich, beim Entfernen der Gruppen, Bindungen zu durchtrennen. Im Zuge der Untersuchung wurde jeweils das gesamte Nucleosid-Diphosphat des Schadens bzw. der intakten Base und die gesamten Aminosäure-Residuen (inklusive Rückgradatome) herausgeschnitten und die durchtrennten Bindungen mit Wasserstoff abgesättigt (siehe gestrichelte Bindungen in Abbildung 6.6). Die QM-Umgebung umfasst für alle vier Rechnungen eines jeden  $L$ - $R_i$ -Paares jeweils  $4 \text{ \AA}$  der Umgebung um  $L$  und  $R_i$  und der herausgeschnittene Teil wird jeweils im Sinne einer Counterpoise-Korrektur für die Fragmentrechnungen einbezogen. Die Tabelle 6.4 führt die Wechselwirkungsbeiträge zwischen dem Nucleosid-Diphosphat des Schadens bzw. der intakten Base mit den 27 nächstgelegenen Proteinresten auf, sowie die Differenz der

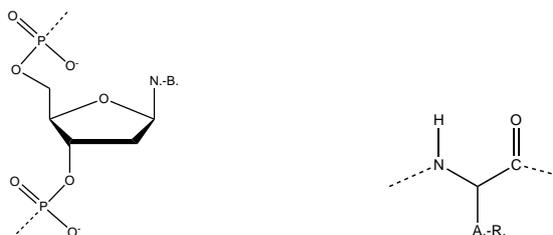


Abb. 6.6: Strukturformeln der herausgeschnittenen Gruppen für die Untersuchung individueller Wechselwirkungsbeiträge; links: Nucleosid-Diphosphat mit geschädigter oder intakter Nucleobase (N.-B.); rechts: gesamtes Aminosäure-Residuum (mit dem Aminosäure-Rest A.-R.); die gestrichelten Bindungen werden durchtrennt und bei den Nachbargruppen mit Wasserstoff abgesättigt.

Wechselwirkungsbeiträge im Falle des Schadens bzw. im Falle der intakten Base.

Eine negative Differenz des Wechselwirkungsbeitrags weist auf eine bevorzugte Stabilisierung des Schadens durch den entsprechenden Proteinrest  $R_i$  hin. Da das Nucleosid-Diphosphat zweifach negativ geladen ist, zeigt die Tabelle für Proteinreste mit negativ geladenen Seitenketten (Glu) eine starke Abstoßung, bzw. bei positiv geladenen Seitenketten (Lys, Arg und der N-Terminus Pro 1) eine starke Anziehung, allerdings in beiden Fällen, dem 8-oxoG und dem G.

Im Rahmen der Strukturauswahl und -vorbereitung (siehe Abschnitt ) zeigt sich, dass (basierend auf der ausgewählten Struktur) die Aminosäureresiduen Pro 1, Glu 5, Tyr 175, Ser 219 und Thr 220 bei der Bindung zwischen Schaden und intakter Nucleobase um mehr als  $10 \frac{\text{kJ}}{\text{mol}}$  differenzieren. Die Residuen Pro 1, Tyr 175, Ser 219 und Thr 220 binden dabei bevorzugt den Schaden. Hingegen präferiert das Residuum Glu 5 die intakte Nucleobase, was allerdings nicht ausschließt, dass Glu 5 oder andere Glutamat-Residuen in einem späteren Reparaturschritt eine wichtige Rolle spielen (siehe Referenz [138]). Die Anordnung für die vier den Schaden am stärksten präferierenden Residuen ist in der Abbildung 6.7 gezeigt. Die starke präferierte Schadensbindung durch Ser 219 kann darin im Wesentlichen als Wasserstoffbrückenbindung zwischen dem Carbonylsauerstoff im Proteinrückgrad mit dem H7-Atom des Schadens identifiziert werden.

Tab. 6.4: Wechselwirkungsbeiträge  $W_{L,R_i}$  des geschädigten DNA-Nukleosid-Diphosphats (8-oxoG) bzw. des intakten Nukleosid-Diphosphats (G)  $L$  mit den nächsten 27 Protein-Residuen  $R_i$ , sowie die Differenz der Wechselwirkungsbeiträge. Neben den Counterpoise-korrigierten QM/MM-Ergebnissen auf HF/SVP-Niveau (QM-Umgebungsabstand 4 Å) sind in Klammern auch D3-Dispersionskorrekturen nach Grimme et al. [147] aufgeführt.

Protein-Residuum $R_i$	8-oxoG		G		Differenz	
Pro 1	-277.5	(-32.8)	-253.6	(-31.9)	-23.9	(-0.9)
Glu 2	+422.2	(-34.4)	+414.7	(-33.4)	+7.5	(-1.0)
Leu 3	-11.5	(-0.2)	-10.3	(-0.2)	-1.2	(+0.0)
Pro 4	-19.5	(-0.4)	-14.6	(-0.3)	-4.9	(-0.1)
Glu 5	+294.6	(-13.5)	+257.2	(-8.7)	+37.4	(-4.8)
Val 6	-14.6	(-0.3)	-12.2	(-0.2)	-2.4	(-0.1)
Lys 59	-417.7	(-15.5)	-408.0	(-15.5)	-9.7	(+0.0)
Arg 75	-220.5	(-1.4)	-218.8	(-1.4)	-1.7	(+0.0)
Met 76	+23.9	(-36.5)	+23.8	(-36.6)	+0.1	(+0.1)
Arg 79	-183.2	(-9.7)	-188.2	(-9.6)	+5.0	(-0.1)
Gly 172	-60.2	(-11.0)	-57.7	(-11.0)	-2.5	(+0.0)
Asn 173	-100.7	(-36.6)	-99.7	(-36.6)	-1.0	(+0.0)
Ile 174	-21.5	(-18.6)	-15.8	(-17.9)	-5.7	(-0.7)
Tyr 175	-45.2	(-1.9)	-32.2	(-1.6)	-13.0	(-0.1)
Ala 213	-1.8	(-0.1)	-0.2	(-0.1)	-1.6	(+0.0)
Gly 218	-15.1	(-0.3)	-14.0	(-0.2)	-1.1	(-0.1)
Ser 219	-39.5	(-12.1)	+31.3	(-6.8)	-70.8	(-5.3)
Thr 220	-30.6	(-4.9)	-15.7	(-4.3)	-14.8	(-0.6)
Val 221	-27.4	(-19.4)	-32.1	(-18.5)	+4.7	(-0.9)
Arg 222	-226.8	(-38.1)	-229.9	(-37.8)	+3.1	(-0.3)
Thr 223	-13.1	(-29.6)	-12.5	(-29.5)	-0.6	(-0.1)
Tyr 224	+9.1	(-30.8)	+2.4	(-28.2)	+6.7	(-2.6)
Val 225	+15.1	(-0.4)	+16.9	(-0.4)	-1.8	(+0.0)
Phe 234	-6.5	(-0.2)	-9.2	(-0.1)	+2.7	(-0.1)
Gln 235	+6.9	(-0.3)	+8.9	(-0.3)	-1.9	(+0.0)
Leu 238	-14.7	(-0.2)	-14.4	(-0.2)	-0.3	(+0.0)
Tyr 241	-78.7	(-25.3)	-77.1	(-25.2)	-1.6	(-0.1)

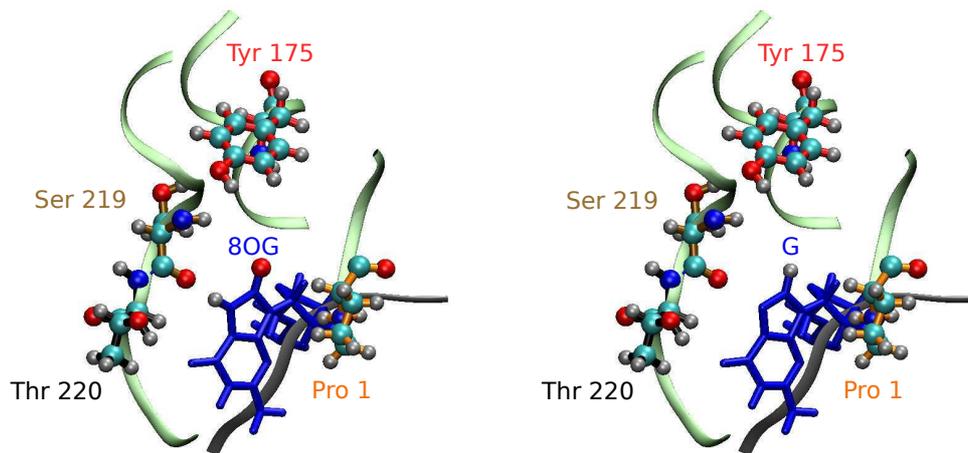


Abb. 6.7: Strukturausschnitt der Proteinresiduen Pro 1, Tyr 175, Ser 219 und Thr 220 und dem Nucleotid des Schadens 8-oxoG (links) bzw. der intakten Nucleobase G (rechts).

## 6.5 Zusammenfassung und Ausblick

Gemäß den voranstehenden QM- und QM/MM-Rechnungen an einer experimentell bestimmten Kristallstruktur bindet das Reparaturenzym MutM die DNA mit einer einzelnen geschädigten Nucleobase 8-oxoG um ungefähr  $68 \frac{\text{kJ}}{\text{mol}}$  stärker als die analoge Struktur mit intakter Nucleobase. Die Enzymresiduen Pro 1, Tyr 175, Ser 219 und Thr 220 haben bei der präferierten Bindung der schadhafte DNA einen wichtigen Einfluss.

Um die durch den lokalen Schaden hervorgerufenen Änderungen der Elektronen-Struktur verlässlich (mit Fehlern im Bereich von  $10 \frac{\text{kJ}}{\text{mol}}$ ) beschreiben zu können, erwiesen sich bei reinen QM-SCF-Beschreibungen große QM-Fragmente als sehr wichtig (QM-Umgebungsabstand um den lokalen Schaden: 9–10 Å). Eine erheblich schnellere Konvergenz lässt sich hier mit einer elektrostatischen QM/MM-Einbettung der Umgebung erreichen (6–7 Å). Auch spielen dabei Korrekturen des Basissatzfehlers und eine Beschreibung der Dispersionswechselwirkungen eine wichtige, wenn auch offenbar etwas kleinere Rolle, wobei künftige SAPT-Rechnungen an größeren Fragmenten das Abklingverhalten der empirischen Dispersions-Korrekturdifferenzen überprüfen müssen. Zukünftige Rechnungen müssen darüber hinaus untersuchen, inwieweit die Ergebnisse auch für andere (z.B. aus anderen experimentellen Strukturanalysen oder Molekulardynamik-Simulationen gewonnenen) Systemkonformationen repräsentativ sind.



## Kapitel 7

### Allgemeine Zusammenfassung

Die vorgestellte Arbeit trägt dazu bei, dynamische molekulare Systeme effizient und verlässlich mit linear-skalierenden QM- und QM/MM-Methoden zu beschreiben. Der Beitrag der Arbeit besteht dabei einerseits in methodischen Weiterentwicklungen zur Reduktion der Vorfaktoren quantenchemischer Berechnungen und andererseits darin, bestehende Methoden gewinnbringend zu kombinieren. Um die gewonnenen theoretischen Beschreibungen verlässlich auszuschöpfen, müssen die Fehlerbereiche der Beschreibungen möglichst genau bekannt sein. Deshalb liegt ein weiterer Schwerpunkt der Arbeit darauf, die methodischen Näherungen systematisch zu untersuchen. Auch in einem Anwendungsteil zur Reparatur oxidativer DNA-Schäden liegt ein besonderes Augenmerk auf Aspekten zu einer zuverlässigen theoretischen Beschreibung. Im Folgenden sind die verschiedenen Beiträge der Arbeit zusammengefasst:

Die Dissertation stellt eine neue Integral-Speichermethode dar, um die Vorteile semidirekter (bzw. indirekter) SCF-Methoden mit denen linear-skalierender direkter Methoden zu vereinen. Für den Austauschteil sind mit dem neuen Verfahren Geschwindigkeitszuwächse von typischerweise 1.5–3.7 möglich, je nach molekularem System, Basissatz und den gewählten Genauigkeiten. Die Geschwindigkeitszuwächse lassen sich dabei gut mit einer parallelen Berechnung der Zweielektronen-Integrale kombinieren. Darüber hinaus wurde im Rahmen dieser Arbeit für das von S. A. Maurer et. al. [60] ausgearbeitete neue QQR-Screening ein umfangreicher molekularer Testsatz entwickelt. Anhand des Testsatzes zeigt sich umfassend, wie leistungsstark das neue Screening ist, das für Zweielektronen-Integrale den Abfall mit zunehmendem Bra-Ket Abstand der Ladungsverteilungen berücksichtigt.

Neben den methodischen Weiterentwicklungen für SCF-Verfahren bei einer gegebenen Molekülgeometrie werden zusätzliche Ansätze vorgestellt, um Informationen aus vergangenen

Zeitschritten einer Born-Oppenheimer-Molekulardynamik zu nutzen und auf diese Weise im aktuellen Schritt die SCF- oder auch die CPSCF-Lösung zu beschleunigen. Ein Ansatz liegt darin, die Zweielektronen-Integrale aus vergangenen MD-Schritten zur Lösung des aktuellen SCF-Problems wiederzuverwenden, wenn die Integrale gemäß einer Abschätzung numerisch invariant sind. Die präsentierte Abschätzung der Integraländerungen basierend auf gespeicherten ersten Integralableitungen erweist sich als vielversprechend, auch wenn die Abschätzung für optimale Resultate weiterentwickelt werden muss. Außerdem gelingt es, CPSCF-Testrechnungen in aufeinander folgenden Zeitschritten zu beschleunigen, indem eine bestehende Methode zur Fockmatrix-Extrapolation beim SCF-Verfahren auf das CPSCF-Verfahren übertragen wird. Als weiteren Beitrag beschreibt die Arbeit eine erweiterte ChemShell-Schnittstelle [79] zwischen dem QM-Programmpaket Q-Chem [64] und MM-Programmpaketen, um große molekulare Umgebungen bei linear-skalierenden SCF-Energie- und Gradientenberechnungen effizient einbeziehen zu können.

Darüber hinaus widmet sich die Arbeit der Ab-initio-Berechnung von NMR-Verschiebungen. Eine Genauigkeitsuntersuchung von NMR-Verschiebungen auf GIAO-SCF und -MP2-Niveau erfasst Methoden- und Basissatzfehler basierend auf einem breit angelegten molekularen Testsatz. Im Anschluss an die Genauigkeitsuntersuchung etabliert die Dissertation am Beispiel von Polypeptiden ein weiterführendes Konzept, um die Genauigkeit berechneter NMR-Verschiebungen gegenüber dem GIAO-SCF-Niveau zu steigern. Dafür werden intermediäre Werte auf HF- und MP2-Niveau für alle isolierten Aminosäuren bereitgestellt und das deutlich verbesserte Fehlerverhalten durch den Vergleich mit MP2-Referenzwerten aufgezeigt. Wie groß die QM-Region (ausgehend von einer zentralen Molekülregion und gegebenen Konformation) gewählt werden muss, um mit einer bestimmten Genauigkeit  $^1\text{H}$ - bzw.  $^{13}\text{C}$ - NMR-Verschiebungen zu berechnen, zeigen die anschließend vorgestellten Konvergenz-Untersuchungen für eine breite Systemauswahl: Typischerweise müssen für mittlere Genauigkeiten unterhalb von 0.1/0.5 ppm für  $^1\text{H}$ -/ $^{13}\text{C}$ - Verschiebungen 6–10 Å der Umgebung um die betrachteten Kerne berücksichtigt werden, was in den betrachteten molekularen Systemen etwa 300–1200 Atome einschließt. Elektrostatische Umgebungseffekte lassen sich mittels einer elektrostatischen QM/MM-Einbettung überwiegend überzeugend berücksichtigen, allerdings unter der Voraussetzung, dass auch hier die QM-Größen mit typischerweise 200–1000 Atome nicht zu klein gewählt werden.

---

Der letzte Teil der Arbeit greift bei einer Studie zur enzymatischen Reparatur oxidativer DNA-Schäden auf die beschriebenen QM- und QM/MM-Methoden zurück. Die Studie quantifiziert an einer experimentell bestimmten Kristallstruktur mit einem DNA-Schaden (8oxoG) die bevorzugte Bindung des Schadens durch das Reparaturenzym MutM (rund  $68 \frac{\text{kJ}}{\text{mol}}$ ) und identifiziert als wichtige Proteinresiduen bei der Schadensbindung die Gruppen Pro 1, Tyr 175, Ser 219 und Thr 220. Für zukünftige vergleichbare Fragestellungen werden zudem verlässliche Aussagen über die benötigten QM-Größen, den Dispersionseffekten und Basissatzfehlern erarbeitet. Insbesondere zeigt sich, dass bei reinen QM-Beschreibungen große QM-Fragmente (700-900 QM-Atome) um den Schaden ausgewählt werden müssen, um die durch den lokalen Schaden hervorgerufenen Änderungen der Elektronenstruktur verlässlich zu beschreiben. Eine wesentlich schnellere Konvergenz lässt sich hier mit einer elektrostatischen QM/MM-Einbettung der Umgebung erreichen (400-700 QM-Atome).



## Anhang A

### Q-Chem-ChemShell-Interface

Das im Rahmen dieser Arbeit erweiterte ChemShell-Q-Chem-Interface ermöglicht Kopplungen zwischen der ChemShell-Programmsuite [79] und dem QM-Programmpaket Q-Chem [64]. Das Interface stellt grundlegende Funktionalität des Q-Chem-Programm-Pakets bereit; momentan ausführlich getestet sind Grundzustands-Energien und Gradienten auf *Restricted-closed-shell*-HF-, -DFT (Funktionale BLYP, BP86, B3LYP, B3PW91, BPW91, S-VWN-LDA, PBE0, B97-2 etc.) und -RI-MP2-Niveau. Eine ausführliche Aufstellung enthält die im Rahmen der Arbeit erstellte Handbuchseite vor [152].

Für das Beispiel einer QM/MM-Rechnung zeigt die Abbildung schematisch den Programmablauf, ausgehend von einer PDB-Datei und MM-Parametrisierung mittels der Amber-Tools [53] (die Parametrisierung liefert die System-Parameter/Topologie-Datei und System-Koordinaten-Datei). Basierend auf einem Benutzer-Input erstellt das Interface die benötigten Input-Dateien für den Q-Chem-Programmdurchlauf. ChemShell führt das MM-Programm (z.B. das integrierte Programm DL-Poly) aus, sowie das QM-Programm Q-Chem und ggf. ergänzend das Programm DFT-D3 von Grimme et al. [147] für empirische Dispersionskorrekturen. Nach Ablauf der separaten Rechnungen extrahiert das Q-Chem-Interface aus dem Q-Chem-Output die benötigten Größen. Sie werden von ChemShell dann mit den MM-Ergebnissen zusammengeführt. Für Kraftrechnungen und einer elektrostatischen QM/MM-Einbettung steuert das Interface die Berechnung der QM-Beiträge für die Kräfte auf die äußeren Atome (siehe Gleichung 2.81). Bei Geometrieoptimierungen oder MD-Simulationen generiert ChemShell im Anschluss die neue Molekülgeometrie und das Q-Chem-Interface den neuen Q-Chem-Input.

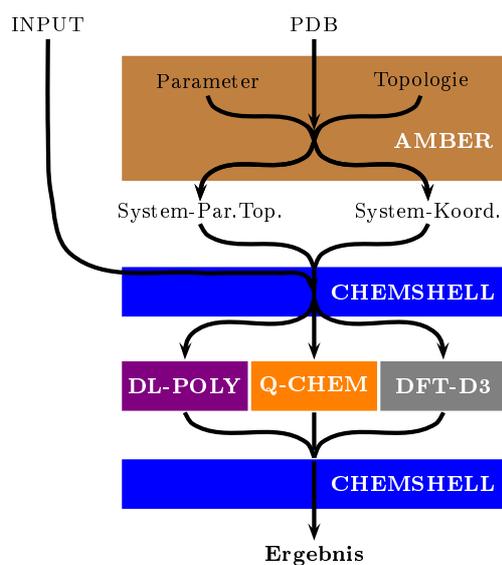


Abb. A.1: Schema des Programmablaufs ausgehend von einer PDB-Datei und Benutzer-Input: Das MM-Programm Amber [53] zur Parametrisierung, ChemShell [79] zur Erzeugung der Input-Dateien für DL-Poly, Q-Chem [64] und DFT-D3 [147] und dem Ausführen der Programme sowie Zusammenführen der Ergebnisse, DL-Poly zur Berechnung der MM-Terme, Q-Chem zur Berechnung der QM-Terme und DFT-D3 ggf. zur Berechnung der Dispersionskorrektur nach Grimme et al.

## Anhang B

### Molekularer Testsatz

Tab. B.1: Die entsprechenden Strukturen können inklusive weiterer Informationen und Quellenangaben im xyz-Dateiformat heruntergeladen werden unter <http://www.cup.uni-muenchen.de/pc/ochsenfeld/download.html>.

Systeme	Beschreibung
<b>Isolatoren:</b>	
Alkane	n-Alkane
Amyloseketten	1-4-glykosidisch verknüpfte D-Glucose-Einheiten
DNA-Fragmente	Nukleotidpaare (ohne endständige Phosphatgruppen)
Diamant-Fragmente	Ausschnitte aus der Kristallstruktur (bis zu 470 Atome)
<b>Delokale Systeme:</b>	
Polyene	konjugiert, all-trans
Polyine	konjugiert
Graphit-Fragmente	mit 1-4 Graphitlagen
Kohlenstoff-Nanoröhren	verschieden lange Ausschnitte aus einer (6,3)- u. (5,0)-Röhre
<b>Intermolekulare Wechselwirkungen:</b>	
Benzol-Dimer	Benzol-Dimer ( $D_{6h}$ ) mit verschiedenen Abständen der Ebenen
Wasser	kugelförmige Ausschnitte aus flüssigem Wasser (TIP3)
<b>Ionische Systeme:</b>	
LiF	Ausschnitt aus der Kristallstruktur (bis zu 288 LiF-Einheiten)
Angiotensin	deprotoniert und als Zwitterion
<b>Systeme mit Elementen höherer Perioden:</b>	
Zeolith	Zeolithfragmente (hier reine Silicate)
Schwefel	Strukturen aus der Kristallstruktur von $S_8$
Kupfer-Komplex	Phthalocyanin-Komplex ( $CuPcF_{16}$ )
<b>Radikalische Systeme:</b>	
Triphenyl-Methyl-Radikal	basierend auf Literaturdaten von Triphenylmethan
<b>Biomolekulare Systeme:</b>	
$\beta$ -Carotine	Konjugiertes $\Pi$ -System, leicht verzerrt



## Anhang C

### Details zur QM-Größenkonvergenz von NMR-Abschirmungen

Die Standardabweichungen werden berechnet gemäß,

$$\text{RMS}_{\text{QM/MM}}(r) = \sqrt{\frac{\sum_A^{N_{\text{nuclei}}} \left\{ \sigma_A^{\text{QM/MM}}(r) - \sigma_A^{\text{QM/MM}}(r_{\text{max}}) \right\}^2}{N_{\text{nuclei}}}},$$

für QM/MM und für eine reine QM-Beschreibung gemäß:

$$\text{RMS}_{\text{QM}}(r) = \sqrt{\frac{\sum_A^{N_{\text{nuclei}}} \left\{ \sigma_A^{\text{QM}}(r) - \sigma_A^{\text{QM/MM}}(r_{\text{max}}) \right\}^2}{N_{\text{nuclei}}}}.$$

Die maximalen Abweichungen (MAX) werden analog berechnet als:

$$\text{MAX}_{\text{QM/MM}}(r) = \max_{\text{over } N_{\text{nuclei}}} \left| \sigma_A^{\text{QM/MM}}(r) - \sigma_A^{\text{QM/MM}}(r_{\text{max}}) \right| \quad (\text{C.1})$$

für QM/MM und für eine reine QM-Beschreibung als:

$$\text{MAX}_{\text{QM}}(r) = \max_{\text{over } N_{\text{nuclei}}} \left| \sigma_A^{\text{QM}}(r) - \sigma_A^{\text{QM/MM}}(r_{\text{max}}) \right| \quad (\text{C.2})$$

Die mittlere Effizienzanzahl von QM/MM (ME) wird berechnet, indem die Brüche  $\text{RMS}_{\text{QM/MM}} / \text{RMS}_{\text{QM}}$  über alle Resultate mit unterschiedlichen QM-Größen  $r$  gemittelt werden, d.h.,

$$\text{ME} = \frac{1}{N_{\text{sizes}} - 1} \sum_{r=r_0, \dots, r_{\text{max}-1}} \frac{\text{RMS}_{\text{QM}}(r)}{\text{RMS}_{\text{QM/MM}}(r)} \quad (\text{C.3})$$

Um die Kerne auszuwählen, bei denen der Vorteil von QM/MM gegenüber einer reinen QM/MM-Beschreibung am größten bzw. am kleinsten sind, werden die Fehler der QM/MM- und der reinen QM-Beschreibung über alle QM-Größen für jeden Kern aufsummiert. Auf diese Weise erhält man die Summe der Fehler (SE).

$$\text{SE}_A^{\text{QM/MM}} = \sum_{r=r_0, \dots, r_{\text{max}}} \left| \sigma_A^{\text{QM/MM}}(r) - \sigma_A^{\text{QM/MM}}(r_{\text{max}}) \right|. \quad (\text{C.4})$$

$$\text{SE}_A^{\text{QM}} = \sum_{r=r_0, \dots, r_{\max}} \left| \sigma_A^{\text{QM}}(r) - \sigma_A^{\text{QM/MM}}(r_{\max}) \right|. \quad (\text{C.5})$$

Der QM/MM-Vorteil kann nun entweder gemäß der relativen Summe der Fehler (RSE),

$$\text{RSE}_A = \frac{\text{SE}_A^{\text{QM/MM}}}{\text{SE}_A^{\text{QM}}}, \quad (\text{C.6})$$

oder der Differenz der Summe der Fehler (DSE),

$$\text{DSE}_A = \text{SE}_A^{\text{QM/MM}} - \text{SE}_A^{\text{QM}}. \quad (\text{C.7})$$

ausgewertet werden. Die vorgestellten Ergebnisse gründen auf einem Kompromiss beider Kriterien: Von den 40 % der Kerne mit dem kleinsten RSE werden diejenigen mit dem kleinsten DSE für die Diagramme des größten QM/MM-Vorteils ausgewählt. Entsprechend sind für die Diagramme des kleinsten Vorteils von den 40 % der Kerne mit größtem RSE diejenigen mit größtem DSE ausgewählt.

Aminopyrazol-Wasser-System:

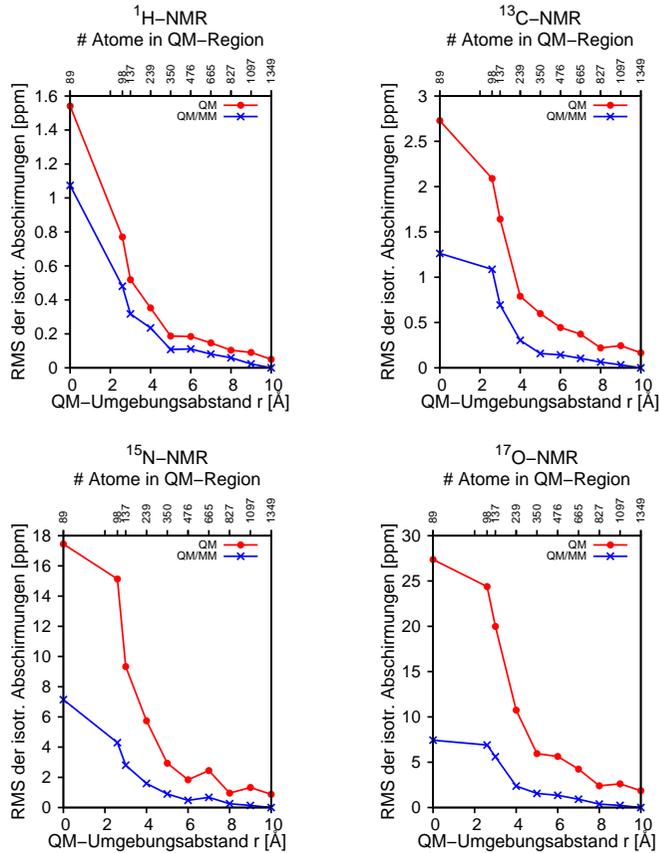


Abb. C.1: Standardabweichung der isotropen Verschiebung gegenüber den QM/MM-Ergebnissen mit dem größten QM-Umgebungsabstand für das Aminopyrazol-System über 44 <sup>1</sup>H (links oben), 27 <sup>13</sup>C (rechts oben), 9 <sup>15</sup>N (links unten) und 9 <sup>17</sup>O (rechts unten) Kerne für eine reine QM-Beschreibung und QM/MM, GIAO-B3LYP/SVP.

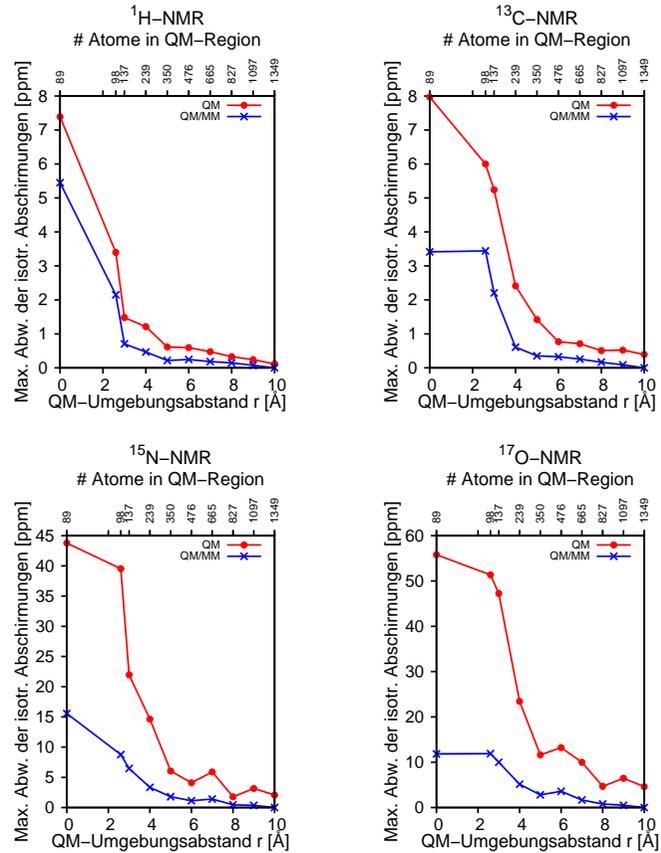


Abb. C.2: Maximale Abweichung der isotropen Verschiebungen gegenüber den QM/MM-Ergebnissen mit dem größten QM-Umgebungsabstand für das Aminopyrazol-System über 44 <sup>1</sup>H (links oben), 27 <sup>13</sup>C (rechts oben), 9 <sup>15</sup>N (links unten), 9 und <sup>17</sup>O (rechts unten) Kerne für eine reine QM-Beschreibung und QM/MM, GIAO-B3LYP/SVP.

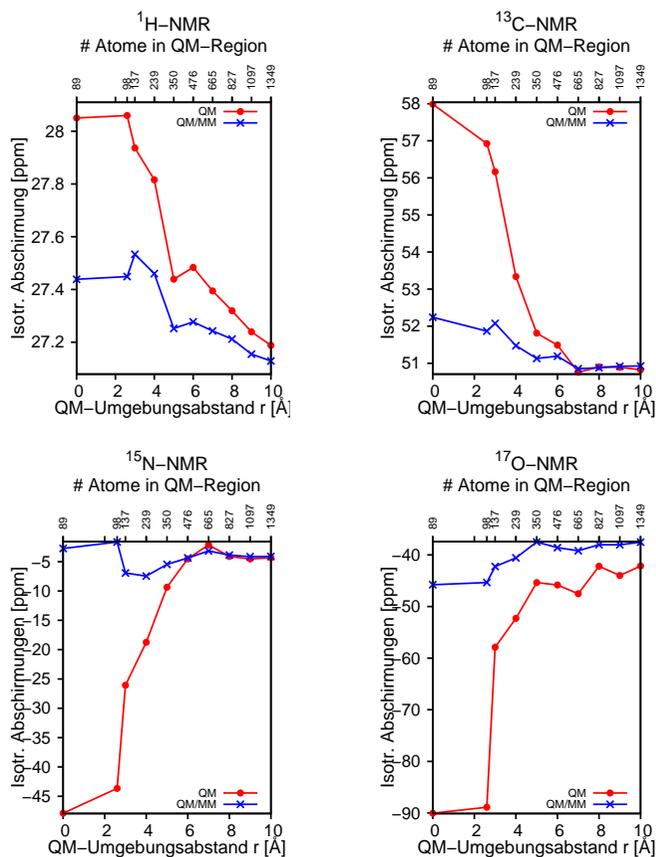


Abb. C.3: Isotrope Abschirmungen für ausgewählte Kerne im Aminopyrazol-System, bei denen der Vorteil von QM/MM gegenüber einer reinen QM-Beschreibung am größten ist, für  $^1\text{H}$  (links oben),  $^{13}\text{C}$  (rechts oben),  $^{15}\text{N}$  (links unten) und  $^{17}\text{O}$  (rechts unten), GIAO-B3LYP/SVP.

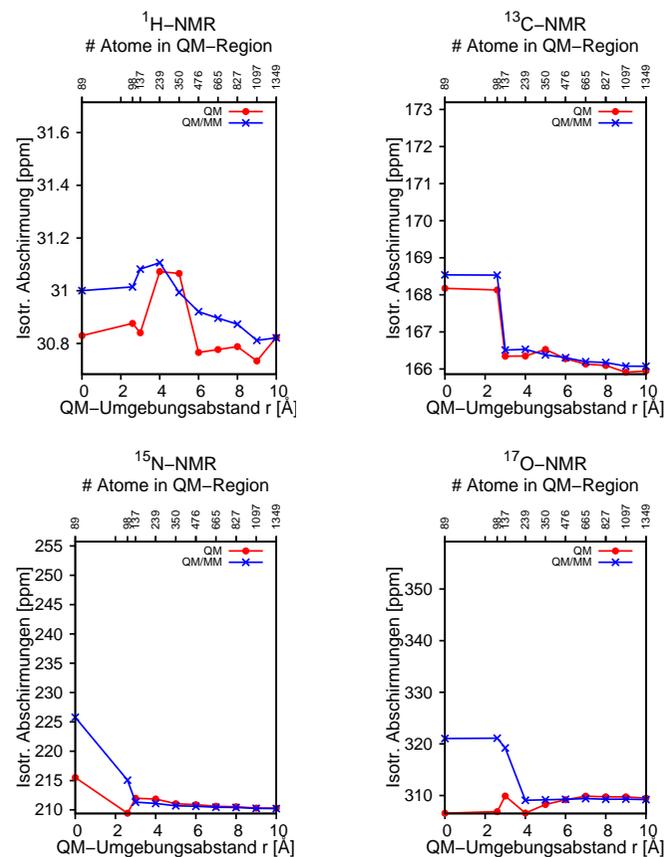


Abb. C.4: Isotrope Abschirmungen für ausgewählte Kerne im Aminopyrazol-System, bei denen der Vorteil von QM/MM gegenüber einer reinen QM-Beschreibung am kleinsten ist, für  $^1\text{H}$  (links oben),  $^{13}\text{C}$  (rechts oben),  $^{15}\text{N}$  (links unten) und  $^{17}\text{O}$  (rechts unten), GIAO-B3LYP/SVP.

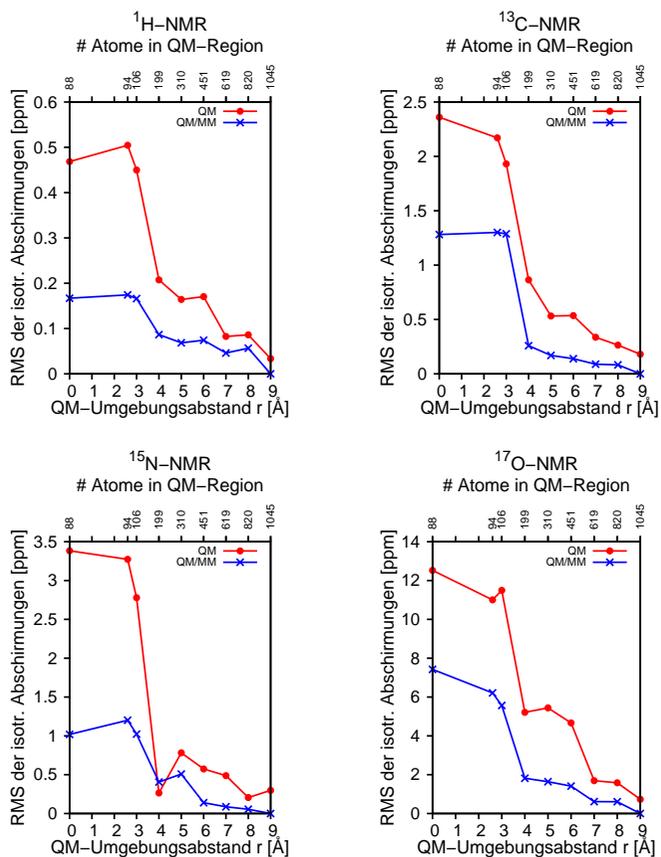


Abb. C.5: Standardabweichung der isotropen Verschiebung gegenüber den QM/MM-Ergebnissen mit dem größten QM-Umgebungsabstand für den molekularen Clip über 36 <sup>1</sup>H (links oben), 41 <sup>13</sup>C (rechts oben), 2 <sup>15</sup>N (links unten) und 7 <sup>17</sup>O (rechts unten) Kerne für eine reine QM-Beschreibung und QM/MM, GIAO-HF/SVP.

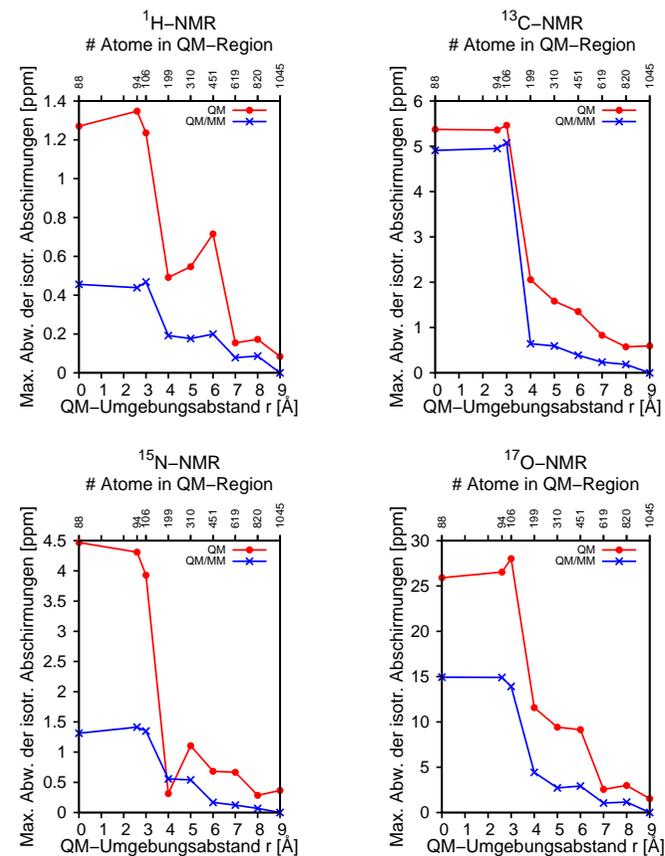


Abb. C.6: Maximale Abweichung der isotropen Verschiebungen gegenüber den QM/MM-Ergebnissen mit dem größten QM-Umgebungsabstand für den molekularen Clip über 36 <sup>1</sup>H (links oben), 41 <sup>13</sup>C (rechts oben), 2 <sup>15</sup>N (links unten) und 7 <sup>17</sup>O (rechts unten) Kerne für eine reine QM-Beschreibung und QM/MM, GIAO-HF/SVP.

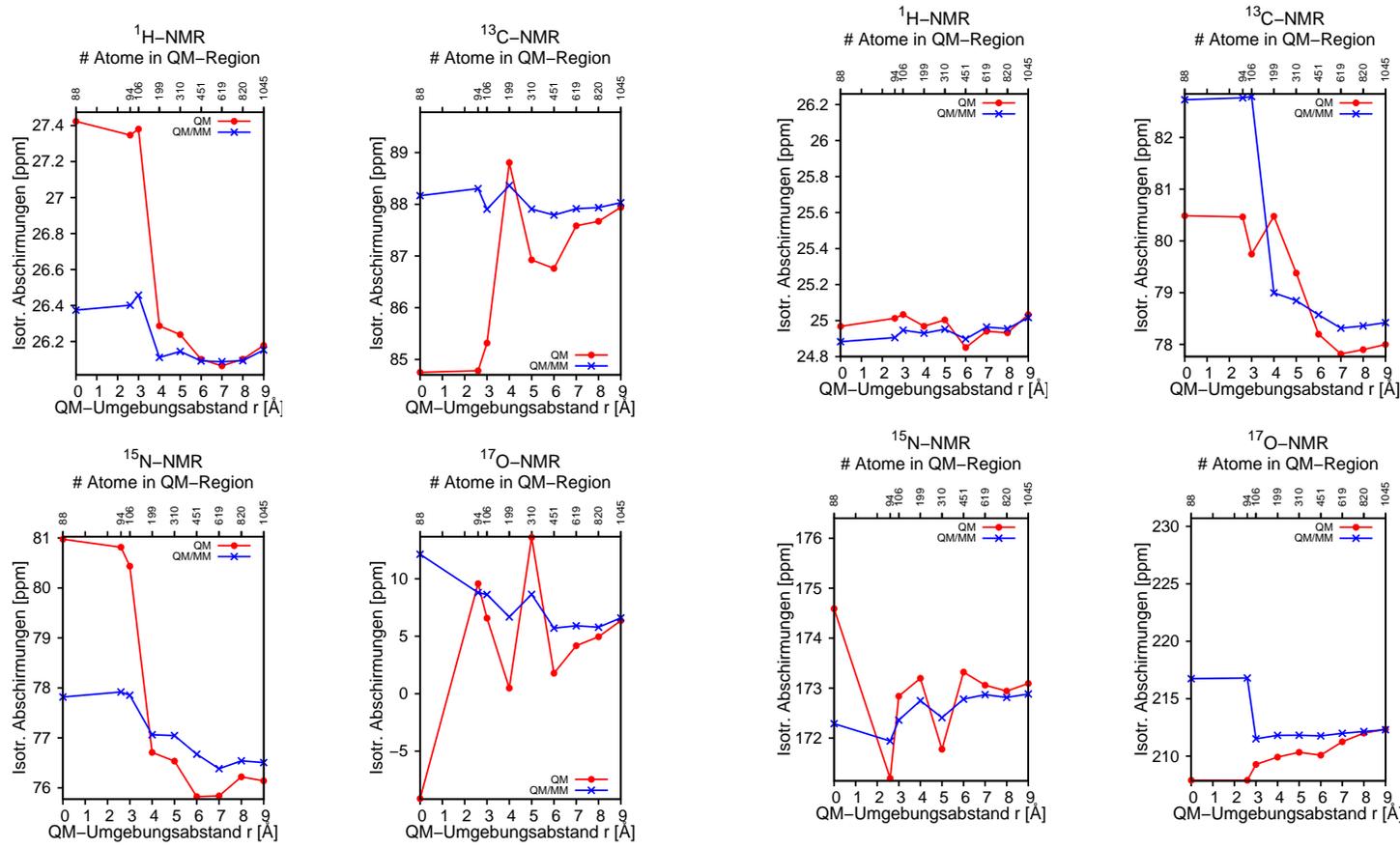


Abb. C.7: Isotrope Abschirmungen für ausgewählte Kerne im molekularen Clip, bei denen der Vorteil von QM/MM gegenüber einer reinen QM-Beschreibung am größten ist, für <sup>1</sup>H (links oben), <sup>13</sup>C (rechts oben), <sup>15</sup>N (links unten) und <sup>17</sup>O (rechts unten), GIAO-HF/SVP.

Abb. C.8: Isotrope Abschirmungen für ausgewählte Kerne im molekularen Clip, bei denen der Vorteil von QM/MM gegenüber einer reinen QM-Beschreibung am kleinsten ist, für <sup>1</sup>H (links oben), <sup>13</sup>C (rechts oben), <sup>15</sup>N (links unten) und <sup>17</sup>O (rechts unten), GIAO-HF/SVP.

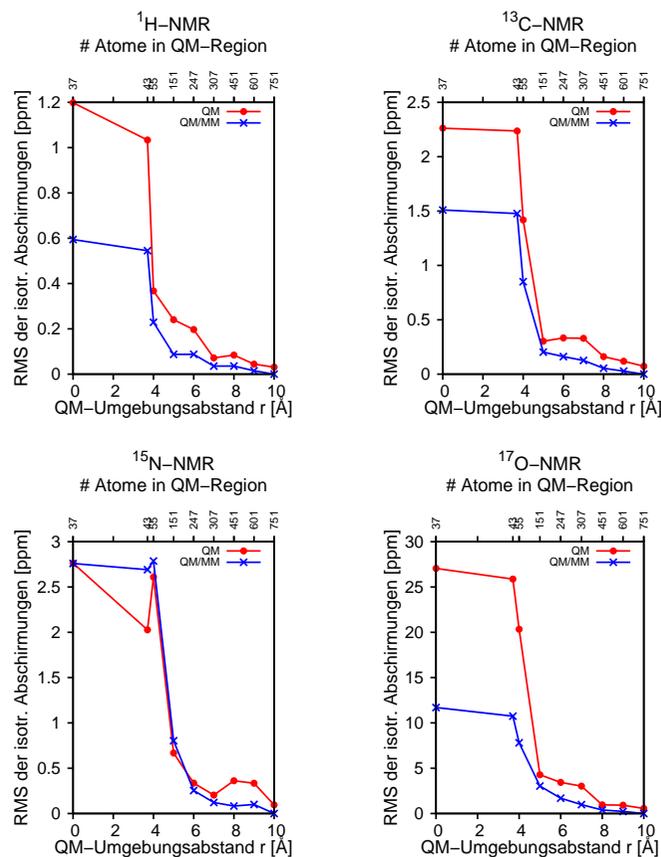


Abb. C.9: Standardabweichung der isotropen Verschiebung gegenüber den QM/MM-Ergebnissen mit dem größten QM-Umgebungsabstand für das Glutathion-System über 17  $^1\text{H}$  (links oben), 10  $^{13}\text{C}$  (rechts oben), 3  $^{15}\text{N}$  (links unten) und 6  $^{17}\text{O}$  (rechts unten) Kerne für eine reine QM-Beschreibung und QM/MM, GIAO-B3LYP/SVP.

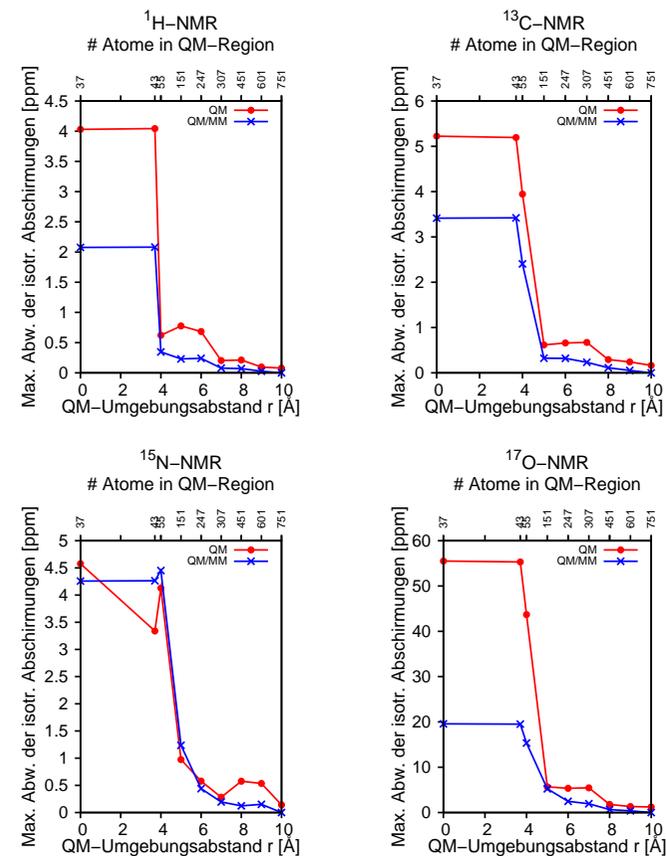


Abb. C.10: Maximale Abweichung der isotropen Verschiebungen gegenüber den QM/MM-Ergebnissen mit dem größten QM-Umgebungsabstand für das Glutathion-System über 17  $^1\text{H}$  (links oben), 10  $^{13}\text{C}$  (rechts oben), 3  $^{15}\text{N}$  (links unten) und 6  $^{17}\text{O}$  (rechts unten) Kerne für eine reine QM-Beschreibung und QM/MM, GIAO-B3LYP/SVP\*\*.

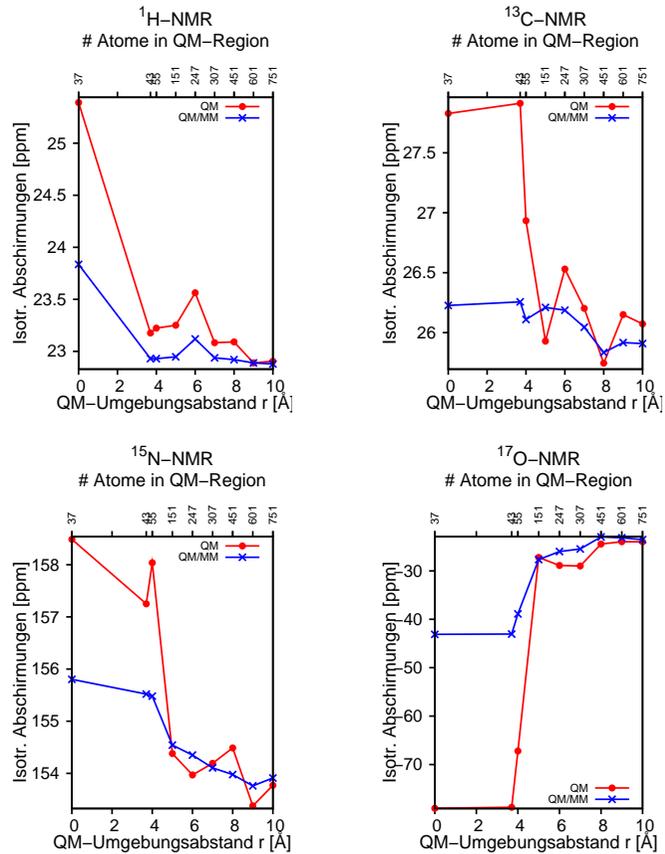


Abb. C.11: Isotrope Abschirmungen für ausgewählte Kerne im Glutathion-System, bei denen der Vorteil von QM/MM gegenüber einer reinen QM-Beschreibung am größten ist, für  $^1\text{H}$  (links oben),  $^{13}\text{C}$  (rechts oben),  $^{15}\text{N}$  (links unten), und  $^{17}\text{O}$  (rechts unten), GIAO-B3LYP/SVP.

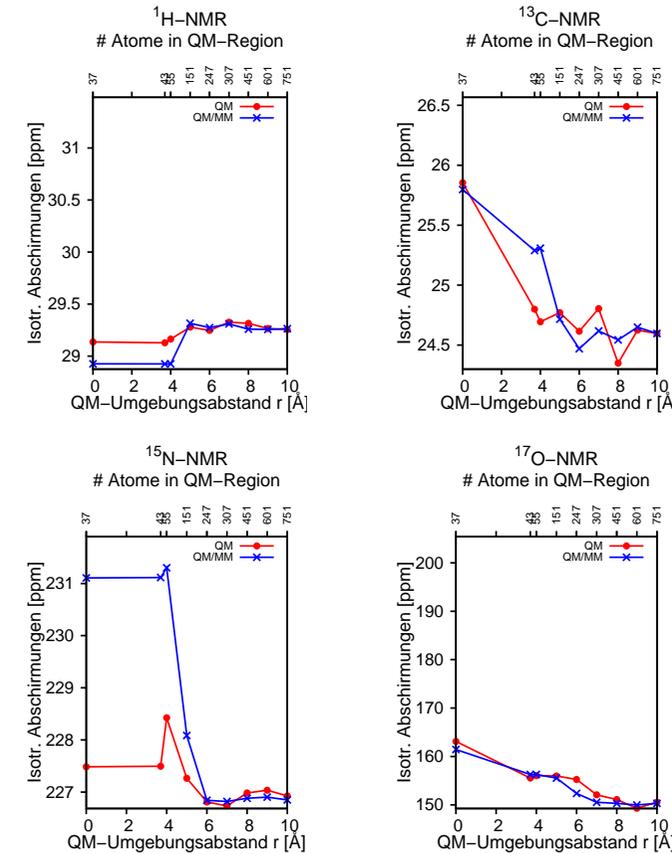


Abb. C.12: Isotrope Abschirmungen für ausgewählte Kerne im Glutathion-System, bei denen der Vorteil von QM/MM gegenüber einer reinen QM-Beschreibung am kleinsten ist, für  $^1\text{H}$  (links oben),  $^{13}\text{C}$  (rechts oben),  $^{15}\text{N}$  (links unten), und  $^{17}\text{O}$  (rechts unten), GIAO-B3LYP/SVP.

## HBC-System:

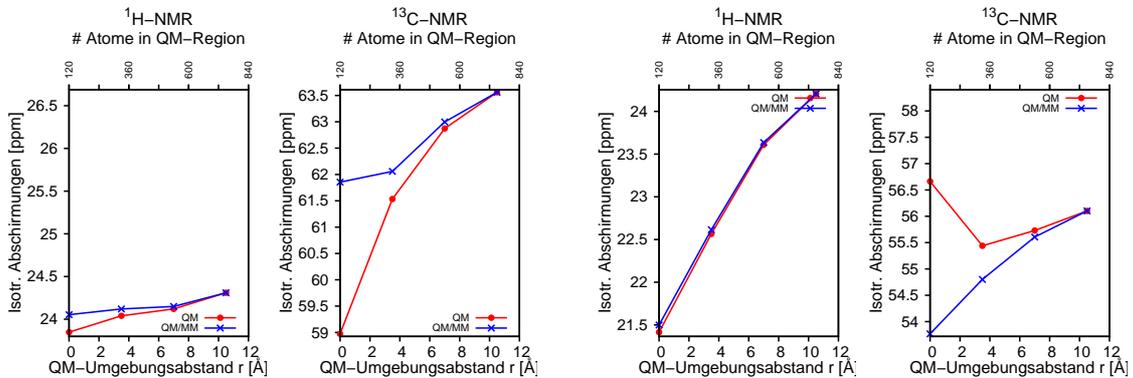


Abb. C.13: Isotrope Abschirmungen für ausgewählte Kerne im HBC-System, bei denen der Vorteil von QM/MM verglichen mit einer reinen QM-Beschreibung am größten ist, für  $^1\text{H}$  und  $^{13}\text{C}$ , GIAO-HF/6-31G\*\*.

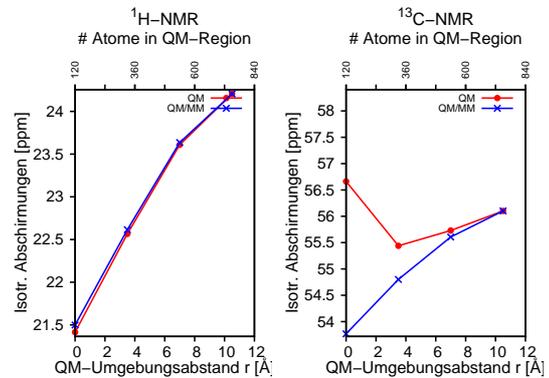


Abb. C.14: Isotrope Abschirmungen für ausgewählte Kerne im HBC-System, bei denen der Vorteil von QM/MM verglichen mit einer reinen QM-Beschreibung am kleinsten ist, für  $^1\text{H}$  und  $^{13}\text{C}$ , GIAO-HF/6-31G\*\*.

## DNA-Enzym-Komplex:

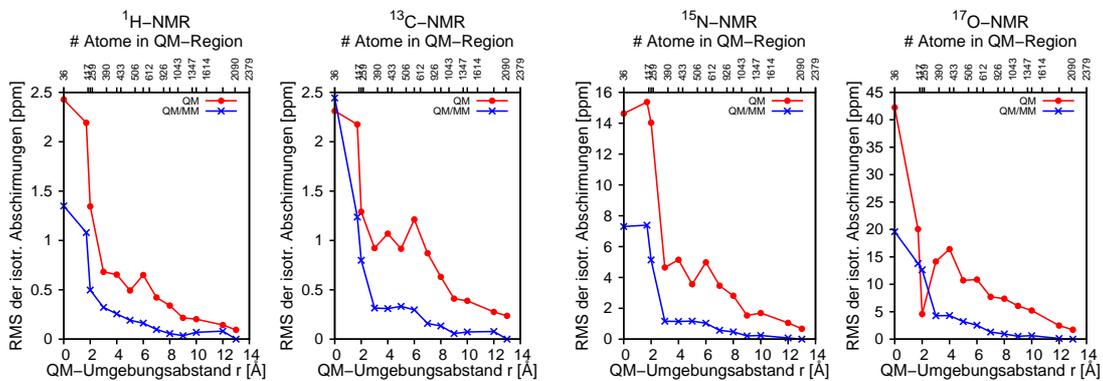


Abb. C.15: Standardabweichung der isotropen Verschiebung gegenüber den QM/MM-Ergebnissen mit dem größten QM-Umgebungsabstand für den DNA-Protein-Komplex (kein Lösungsmittel) über 11  $^1\text{H}$  (ganz links), 5  $^{15}\text{N}$  (Mitte links), 10  $^{13}\text{C}$  (Mitte rechts), 7  $^{17}\text{O}$  (ganz rechts) Kerne für eine reine QM-Beschreibung und QM/MM, GIAO-HF/3-21G.

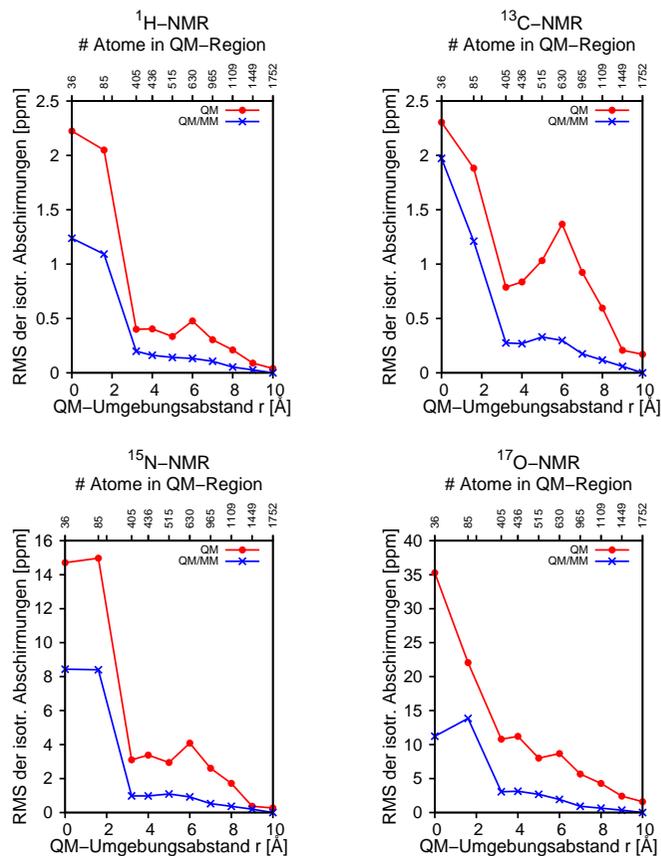


Abb. C.16: Standardabweichung der isotropen Abschirmung gegenüber den QM/MM-Ergebnissen mit dem größten QM-Umgebungsabstand für den DNA-Protein-Komplex in Wasser über 11  $^1\text{H}$  (links oben), 10  $^{13}\text{C}$  (rechts oben), 5  $^{15}\text{N}$  (links unten) und 7  $^{17}\text{O}$  (rechts unten) Kerne für eine reine QM-Beschreibung und QM/MM, GIAO-HF/6-31G\*\*.

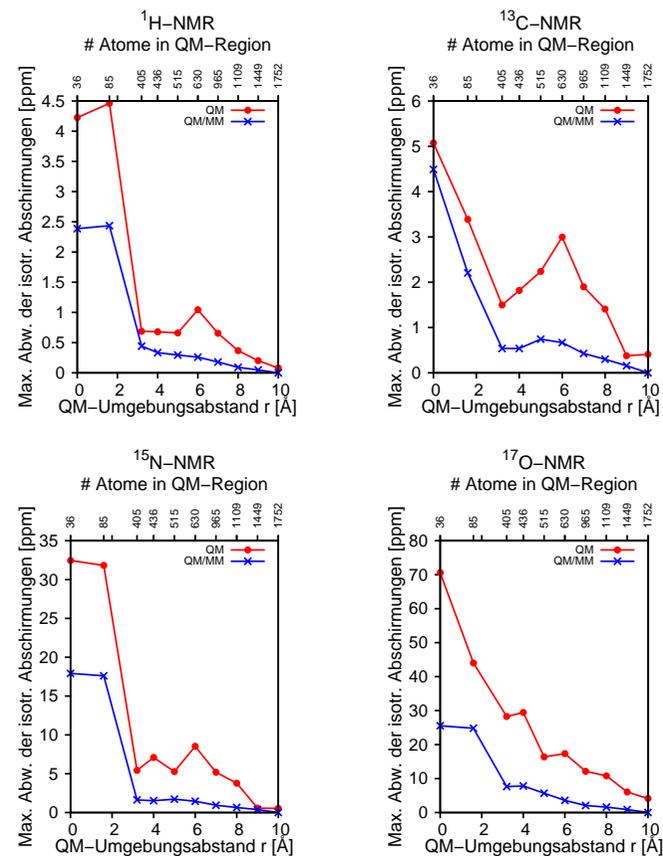


Abb. C.17: Maximale Abweichung der isotropen Abschirmung gegenüber den QM/MM-Ergebnissen mit dem größten QM-Umgebungsabstand für den DNA-Protein-Komplex in Wasser über 11  $^1\text{H}$  (links oben), 10  $^{13}\text{C}$  (rechts oben), 5  $^{15}\text{N}$  (links unten) und 7  $^{17}\text{O}$  (rechts unten) Kerne für eine reine QM-Beschreibung und QM/MM, GIAO-HF/6-31G\*\*.

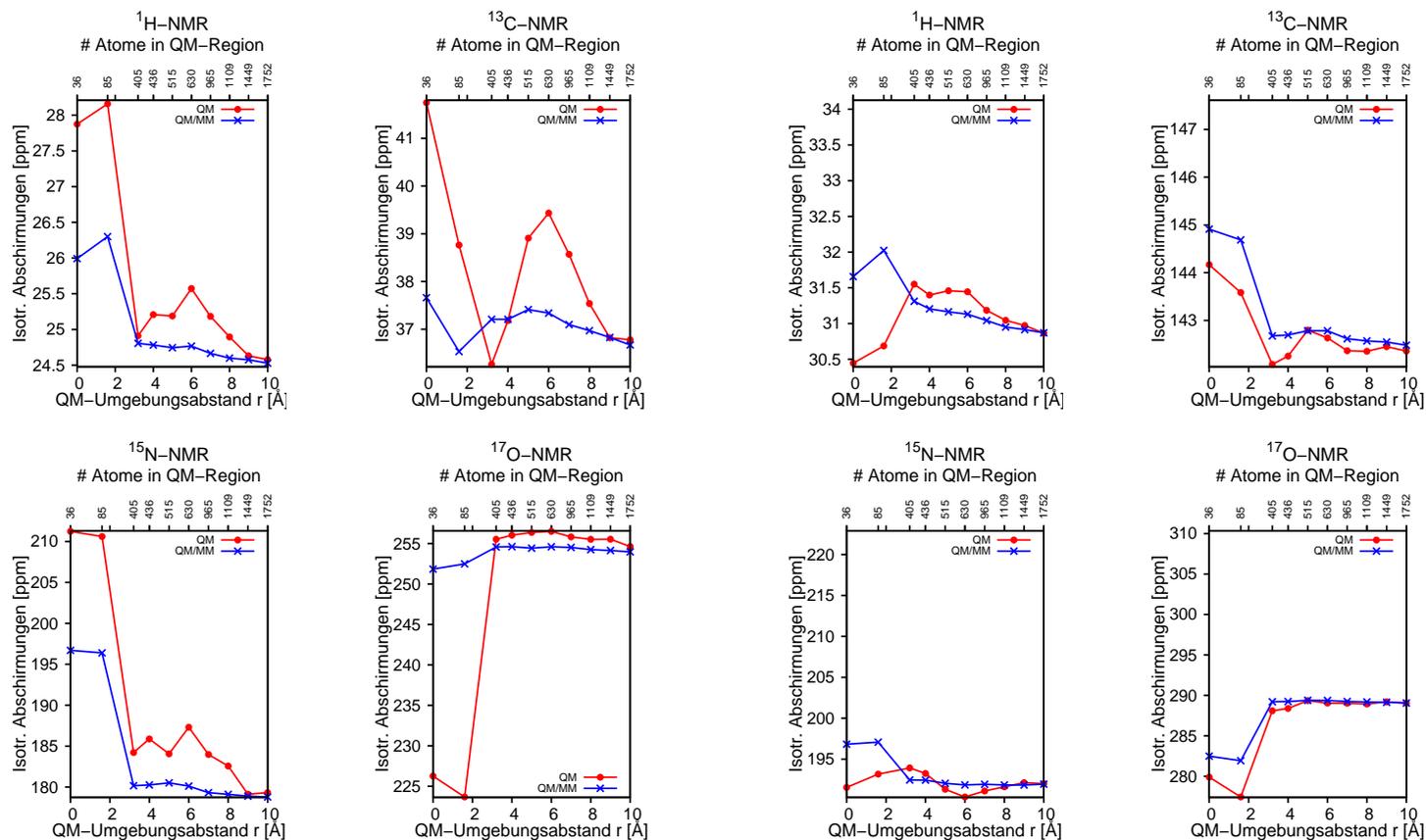


Abb. C.18: Isotrope Abschirmungen für ausgewählte Kerne im DNA-Protein-Komplex in Wasser, bei denen der Vorteil von QM/MM gegenüber einer reinen QM-Beschreibung am größten ist, für  $^1\text{H}$  (links oben),  $^{13}\text{C}$  (rechts oben),  $^{15}\text{N}$  (links unten) und  $^{17}\text{O}$  (rechts unten), GIAO-HF/6-31G\*\*.

Abb. C.19: Isotrope Abschirmungen für ausgewählte Kerne im DNA-Protein-Komplex in Wasser, bei denen der Vorteil von QM/MM gegenüber einer reinen QM-Beschreibung am kleinsten ist, für  $^1\text{H}$  (links oben),  $^{13}\text{C}$  (rechts oben),  $^{15}\text{N}$  (links unten) und  $^{17}\text{O}$  (rechts unten), GIAO-HF/6-31G\*\*.



## Anhang D

### Details zur Untersuchung der DNA-Reparatur

Die QM-Region bei der Berechnung der Gesamtwechselwirkung zwischen DNA und Protein umfasst bei den verschiedenen QM-Umgebungsabständen folgende Residuen:

5.0 Å: Pro 1, Glu 2, Glu 5, Ile 174, Tyr 175, Ser 219, Val 221, Tyr 224, 8OG/DG 292

jeweils zusätzlich:

6.0 Å: Pro 4, Arg 79, Thr 220, Arg 222, Thr 223

7.0 Å: Val 6, Met 76, Gly 172, Tyr 241, DG 293

8.0 Å: Leu 3, Lys 59, Arg 75, Asn 173, Ala 213, Val 225, Phe 234, Gln 235, Leu 238, H<sub>2</sub>O 335, DA 291

9.0 Å: Thr 8, Ile 9, Leu 74, Glu 77, Gly 170, Phe 171, Val 176, Ile 210, Gly 217, Gly 232, Arg 263, H<sub>2</sub>O 340

10.0 Å: Glu 7, Gly 78, Glu 178, Thr 209, Val 214, Ala 231, Thr 233, H<sub>2</sub>O 316, H<sub>2</sub>O 355, H<sub>2</sub>O 384, H<sub>2</sub>O 13698

11.0 Å: Hie 73, Arg 111, Leu 163, Asp 177, Lys 216, Asn 226, Glu 230, Hie 236, Gly 242, H<sub>2</sub>O 316, H<sub>2</sub>O 317, H<sub>2</sub>O 350, H<sub>2</sub>O 363, H<sub>2</sub>O 387, DT 294, H<sub>2</sub>O 392, H<sub>2</sub>O 13661, H<sub>2</sub>O 13760, H<sub>2</sub>O 13840, H<sub>2</sub>O 13862



## Tabellenverzeichnis

2.1	Übersicht über verschiedene grundlegende molekulardynamische Methoden . . .	14
2.2	Berechnung ausgewählter molekularer Eigenschaften als Energieableitung . . . .	25
3.1	CPU-Zeit und Prozentsatz verschiedener Teilschritte bei der Bildung der Austauschmatrix . . . . .	43
3.2	Einfluss des Integral-Grenzwerts und des SCF-Konvergenzkriteriums auf die Geschwindigkeitszuwächse zur Bildung des Austauschteils . . . . .	44
3.3	Geschwindigkeitszuwächse bei der Bildung des Austauschteils bei verschiedenen molekularen Systemen und Basissätzen . . . . .	45
3.4	Prozentsatz gespeicherter Zweielektronen-Integrale in der zweiten Iteration und zusätzlich berechnete Integrale in späteren Iterationen . . . . .	46
3.5	Untersuchung des Skalenverhaltens mit der Systemgröße für lineare Alkane, helikale Peptide und Amylose-Fragmente . . . . .	48
4.1	Mittlere Anzahl der CPSCF-Iterationen, die ohne Extrapolation und mit unterschiedlichen Extrapolationsschemata benötigt werden sowie Standardabweichungen und maximale Fehler der NMR-Verschiebungen für Tetrafluorethylen .	60
4.2	Mittlere Anzahl der CPSCF-Iterationen, die ohne Extrapolation und mit unterschiedlichen Extrapolationsschemata benötigt werden sowie Standardabweichungen der NMR-Verschiebungen für weitere Testsysteme . . . . .	61
5.1	<sup>1</sup> H-NMR-Verschiebungen (in ppm) auf GIAO-CCSD(T), -MP2-, -DFT- und -HF-Niveau und mittlere vorzeichenbehaftete Abweichung, Wurzel der mittleren quadratischen Abweichung und Standardabweichung gegenüber den CCSD(T)-Ergebnissen . . . . .	77

5.2	$^{13}\text{C}$ -NMR-Verschiebungen (in ppm) auf GIAO-CCSD(T), -MP2-, -DFT- und -HF-Niveau und mittlere vorzeichenbehaftete Abweichung, Wurzel der mittleren quadratischen Abweichung und Standardabweichung gegenüber den CCSD(T)-Ergebnissen . . . . .	79
5.3	Abhängigkeit der mittleren vorzeichenbehafteten Abweichung und mittleren absoluten Abweichung von der Wahl der Referenz . . . . .	81
5.4	Abhängigkeit der Standardabweichung von der Wahl des Basissatzes bei der Berechnung von $^{13}\text{C}$ -Verschiebungen . . . . .	81
5.5	$^{13}\text{C}$ -Verschiebungen der isolierten Aminosäuren mit ACE- und NME-Termini gegenüber TMS . . . . .	85
5.6	$^{13}\text{C}$ -Abschirmungen der isolierten Aminosäuren mit ACE- und NME-Termini . . . . .	86
5.7	Mittlere vorzeichenbehaftete Abweichung, mittlere absolute Abweichung und Standardabweichung gegenüber MP2-Verschiebungen für Hartree-Fock-Verschiebungen im Vergleich zu intermediär bestimmten Verschiebungen . . . . .	88
6.1	Einfluss der Basissatzgröße mit und ohne Counterpoise-Korrektur auf die Wechselwirkungsdifferenz intakter und geschädigter DNA mit der Proteinumgebung . . . . .	114
6.2	Einfluss der QM-Theorie auf die Wechselwirkungsdifferenz von intakter und geschädigter DNA mit der Proteinumgebung . . . . .	114
6.3	Extrapolierter Gesamtwert für die Wechselwirkungsdifferenz von intakter und geschädigter DNA (N) mit der Proteinumgebung (P) . . . . .	117
6.4	Wechselwirkungsbeiträge des geschädigten bzw. intakten DNA-Nukleosid-Diphosphats mit den nächsten 25 Protein-Resten . . . . .	120

## Abbildungsverzeichnis

2.1	Schematische Darstellung zu Kern-Elektron-Separationswegen . . . . .	12
2.2	Skalenverhalten Wellenfunktions-basierter Elektronenstruktur-Methoden . . . . .	17
2.3	Schema und Beispiel einer inneren und äußeren Molekülregion als Grundlage des QM/MM-Ansatzes . . . . .	32
3.1	Übersicht über die Kernbestandteile der quasi-indirekten bzw. semidirekten Me- thode . . . . .	40
3.2	Maximaler theoretischer Geschwindigkeitszuwachs für die Bildung der HF-Aus- tauschmatrix abhängig vom Prozentsatz der Integralberechnung . . . . .	42
3.3	Korrelation zwischen der CPU-Zeit der Integralberechnung und der Anzahl der benötigten Gleitkomma-Operationen . . . . .	47
3.4	Untersuchung des Skalenverhaltens mit der Systemgröße am Beispiel von Ket- tensilicaten . . . . .	50
3.5	Kombination der Integralspeicher-Methode mit einer parallelen Berechnung der Zweielektronen-Integrale . . . . .	51
4.1a	Vergleich der Vernachlässigung von Zweielektronen-Integralen mit der Wieder- verwendung von Zweielektronen-Integralen für ein DNA-Basenpaars . . . . .	64
4.1b	Vergleich der Vernachlässigung von Zweielektronen-Integralen mit der Wieder- verwendung von Zweielektronen-Integralen für das Beispiel eines Kohlenstoff- nanoröhren-Fragments . . . . .	64
4.2	Schema zur Ableitung von Schätzwerten für die Änderung eines Zweielektronen- Integrals zwischen zwei aufeinanderfolgenden Molekülgeometrien . . . . .	66

4.3	Exakte und abgeschätzte Größenverteilung der maximalen Schalenquartettänderung und Verteilung des Quotienten aus der abgeschätzten und exakten maximalen Schalenquartettänderung . . . . .	69
5.1	Schema für das Konzept intermediärer Referenzen für Polypeptide . . . . .	84
5.2	Schema der Untersuchung zur Konvergenz von NMR-Abschirmungen mit der Größe der QM-Region . . . . .	91
5.3a	Aminopyrazol-Peptid und molekularer Clip innerhalb einer Wasserumgebung von 5 Å . . . . .	93
5.3b	Glutathion innerhalb einer Methanol-Umgebung von 5 Å . . . . .	93
5.3c	Heptamer-Ausschnitt aus der gestapelten Struktur des Phenyl-substituierten Hexa-peri-hexabenzocoronens . . . . .	93
5.3d	Komplex des Enzyms MutM und DNA mit 8-oxoG-Schaden . . . . .	94
5.4	Differenzbetrag der isotropen Abschirmung gegenüber den Ergebnissen der größten QM-Rechnung für den am meisten und am wenigsten beeinflussten $^1\text{H}$ - und $^{13}\text{C}$ -Kern im Aminopyrazol-Wasser-System . . . . .	97
5.5	Isotrope Abschirmungen für ausgewählte Kerne im Aminopyrazol-System, bei denen der Vorteil von QM/MM im Vergleich zu reinem QM maximal und minimal ist . . . . .	98
5.6	Standardabweichungen der isotropen Abschirmungen gegenüber dem QM-Ergebnis des größten QM-Umgebungsabstands für verschiedene Basissätze . . . . .	100
5.7	Standardabweichungen der isotropen Abschirmungen gegenüber dem QM-Ergebnis des größten QM-Umgebungsabstands für verschiedene QM-Methoden . . . . .	101
5.8a	Standardabweichungen der isotropen Abschirmungen gegenüber dem QM/MM-Ergebnis mit dem größtem QM-Umgebungsabstand für das Aminopyrazol-Wasser-System . . . . .	103
5.8b	Standardabweichungen der isotropen Abschirmungen gegenüber dem QM/MM-Ergebnis mit dem größtem QM-Umgebungsabstand für das wässrige System des molekularen Clips . . . . .	104

5.8c	Standardabweichungen der isotropen Abschirmungen gegenüber dem QM/MM-Ergebnis mit dem größtem QM-Umgebungsabstand für das Glutathion-System in Methanol . . . . .	105
5.8d	Standardabweichungen der isotropen Abschirmungen gegenüber dem QM/MM-Ergebnis mit dem größtem QM-Umgebungsabstand für den DNA-Enzym-Komplex . . . . .	106
5.8e	Standardabweichungen der isotropen Abschirmungen gegenüber dem QM/MM-Ergebnis mit dem größtem QM-Umgebungsabstand für das Hexaperi-hexabenzocoronen-System . . . . .	107
6.1	Strukturformeln der DNA-Base Guanin sowie der häufigen oxidativen Schäden 8-oxoG und FaPyG . . . . .	109
6.2	Strukturen des DNA-Enzym-Komplexes mit 8oxo-G-Schaden und mit intakter G-Base . . . . .	112
6.3	QM-Größenkonvergenz der Wechselwirkungsdifferenz intakter und geschädigter DNA mit der Proteinumgebung . . . . .	115
6.4	Größenkonvergenz der Counterpoise- und D3-Dispersionskorrektur für die Wechselwirkungsdifferenz intakter und geschädigter DNA mit der Proteinumgebung .	117
6.5	Schema zur Herleitung des Wechselwirkungsbeitrags eines DNA-Nukleotids mit einem Protein-Rest . . . . .	118
6.6	Strukturformeln der herausgeschnittenen Gruppen für die Untersuchung individueller Wechselwirkungsbeiträge . . . . .	119
6.7	Strukturausschnitt der Proteinresiduen Pro 1, Tyr 175, Ser 219 und Thr 220 und dem Nukleotid des Schadens 8-oxoG (links) bzw. der intakten Nukleobase G (rechts). . . . .	121



## Literaturverzeichnis

- [1] F. Jensen, *Introduction to Computational Chemistry*, Wiley-VCH, Weinheim, (2007).  
[1](#), [2.4](#), [2.5](#)
- [2] C. Ochsenfeld, J. Kussmann und D. S. Lambrecht, *Linear-Scaling Methods in Quantum Chemistry*, Reviews in Computational Chemistry, Wiley-VCH, Weinheim, (2007). [1](#),  
[2.3.1](#)
- [3] S. Goedecker und G. E. Scuseria, *Computing in Science and Engineering* **5**, 14 (2003).  
[1](#)
- [4] Q. Cui und M. Karplus, *J. Phys. Chem. B* **104**, 3721 (2000). [1](#), [5.4](#)
- [5] C. Ochsenfeld, J. Kussmann und F. Koziol, *Angew. Int. Ed.* **43**, 4485 (2004). [1](#), [5.1](#), [5.4](#)
- [6] J. Kussmann und C. Ochsenfeld, *J. Chem. Phys.* **127**, 054103 (2007). [1](#), [2.4.2](#), [5.1](#), [5.4](#)
- [7] M. Beer und C. Ochsenfeld, *J. Chem. Phys.* **128**, 221102 (2008). [1](#), [2.4.2](#), [4.2](#), [5.1](#), [5.4](#),  
[5.4.1](#), [5.4.2](#)
- [8] D. Marx und J. Hutter, *Ab Initio Molecular Dynamics*, Basic Theory and Advanced Methods, Cambridge University Press, Cambridge, (2009). [2.1](#), [2.1](#), [2.1](#), [2.2](#)
- [9] M. Born und K. Huang, *Dynamical Theory of Crystal Lattices*, Clarendon Press, Oxford, (1954). [2.1](#)
- [10] W. Kutzelnigg, *Mol. Phys.* **90**, 909 (1997). [2.1](#)
- [11] P. T. von den Hoff, *Gekoppelte Kern- und Elektronendynamik: Molekulare Systeme und deren Kontrolle durch die Bewegung der Elektronen*, Dissertation, Arbeitsgruppe de Vivie-Riedle, Ludwig-Maximilians-Universität München, (2011). [2.1](#)

- [12] A. Hofmann und R. de Vivie-Riedle, *Chem. Phys. Lett.* **346**, 299 (2001). 2.1
- [13] M. Born und B. Oppenheimer, *Annalen der Physik* **84**, 457 (1927). 2.1
- [14] L. Verlet, *Phys. Rev.* **159**, 98 (1967). 2.2, 2.2
- [15] J.-I. Choe und B. Kim, *Bull. Korean Chem. Soc.* **21**, 419 (2000). 2.2
- [16] T. Schlick, E. Barth und M. Mandziuk, *Annu. Rev. Biophys. Biomol. Struct.* **26**, 181 (1997). 2.2
- [17] A. Szabo und N. S. Ostlund, *Modern Quantum Chemistry*, Dover, New York, (1996). 2.3.1, 2.3.2
- [18] P. Pulay, *Chem. Phys. Letters* **73**, 393 (1980). 2.3.1, 2.4.2, 4.2, 4.3
- [19] M. Häser und R. Ahlrichs, *J. Comp. Chem.* **10**, 104 (1989). 2.3.1, 3.1, 3.2, 3.2.1
- [20] P. E. Maslen, C. Ochsenfeld, C. A. White, M. S. Lee und M. Head-Gordon, *J. Phys. Chem. A* **102**, 2215 (1998). 2.3.1
- [21] D. Cremer und J. Gauss, *J. Comput. Chem.* **7**, 274 (1986). 2.3.1, 3.1, 3.2.1
- [22] C. Ochsenfeld, C. A. White und M. Head-Gordon, *J. Chem. Phys.* **109**, 1663 (1998). 2.3.1, 3.1, 3.2
- [23] C. Ochsenfeld, *Chem. Phys. Lett.* **327**, 216 (2000). 2.3.1, 3.1, 3.2
- [24] C. A. White, B. G. Johnson und M. Head-Gordon, *J. Phys. Lett.* **230**, 8 (1994). 2.3.1, 3.2, 3.3
- [25] J. Čížek, *J. Chem. Phys.* **45**, 4256 (1966). 2.3.2
- [26] J. Čížek und J. Paldus, *Int. J. Quant. Chem.* **5**, 359 (1971). 2.3.2
- [27] J. Noga und R. J. Bartlett, *J. Chem. Phys.* **86**, 7041 (1987). 2.3.2
- [28] C. Møller und M. S. Plesset, *Phys. Rev.* **46**, 618 (1934). 2.3.2
- [29] K. Raghavachari, G. W. Trucks, J. A. Pople und M. Head-Gordon, *Chem. Phys. Lett.* **157**, 479 (1989). 2.3.2

- [30] G. E. Scuseria, *J. Chem. Phys.* **94**, 442 (1991). 2.3.2
- [31] P. Hohenberg und W. Kohn, *Phys. Rev.* **136**, B864 (1964). 2.3.3
- [32] W. Kohn und L. J. Sham, *Phys. Rev.* **140**, A1133 (1965). 2.3.3
- [33] G. E. Scuseria, *J. Phys. Chem. A* **103**, 4782 (1999). 2.3.3
- [34] R. G. Parr und W. Yang, *Density-Functional Theory of Atoms and Molecules*, Band 16 der International Series of Monographs on Chemistry, Oxford Science Publications, (1989). 2.3.3
- [35] S. Kurth, J. P. Perdew und P. Blaha, *Int. J. Quant. Chem.* **75**, 889 (1999). 2.3.3
- [36] J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria und G. I. Csonka, *J. Chem. Phys.* **123**, 062201 (2005). 2.3.3
- [37] S. H. Vosko, L. Wilk und M. Nusair, *Can. J. Phys.* **58**, 1200 (1980). 2.3.3, 5.4.4
- [38] J. P. Perdew und Y. Wang, *Phys. Rev. B* **45**, 13244 (1992). 2.3.3
- [39] J. P. Perdew, K. Burke und M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996). 2.3.3, 5.2, 5.1, 5.2, 5.3, 5.4
- [40] A. D. Becke und M. R. Roussel, *Phys. Rev. A* **39**, 3761 (1989). 2.3.3
- [41] P. J. Stephens, F. J. Devlin, C. F. Chabalowski und M. J. Frisch, *J. Phys. Chem.* **98**, 11623 (1994). 2.3.3, 5.2, 5.1, 5.2, 5.3, 5.4, 5.4.4, 6.3, 6.2
- [42] A. D. Becke, *J. Chem. Phys.* **98**, 1372 (1993). 2.3.3, 5.2, 5.1, 5.2, 5.3, 5.4, 5.4.4, 6.3, 6.2
- [43] C. Adamo, G. E. Scuseria und V. Barone, *J. Chem. Phys.* **111**, 2889 (1999). 2.3.3, 5.2, 5.1, 5.2, 5.3, 5.4, 5.2, 5.4.4
- [44] P. J. Wilson, T. J. Bradley und D. J. Tozer, *J. Chem. Phys.* **115**, 9233 (2001). 2.3.3, 5.2, 5.1, 5.2, 5.3, 5.4, 5.2, 6.2, 6.3, 6.2
- [45] J. Gauss, *Molecular Properties*, in *Modern Methods and Algorithms of Quantum Chemistry*, Hrsg.: J. Grotendorst, Band 3, S. 541–592, John von Neumann Institute for Computing, Jülich, (2000). 2.4, 2.4.2, 5.1

- [46] L. Brillouin, *J. Phys. Radium* **3**, 373 (1932). [2.4.1](#)
- [47] F. London, *J. Phys. Radium* **8**, 397 (1937). [2.4.1](#)
- [48] R. Ditchfield, *Mol. Phys.* **27**, 789 (1974). [2.4.1](#)
- [49] K. Wolinski, J. F. Hinton und P. Pulay, *J. Am. Chem. Soc.* **112**, 8251 (1990). [2.4.1](#)
- [50] M. Beer, *Effiziente ab-initio Berechnung molekularer Eigenschaften großer Systeme*, Dissertation, Arbeitsgruppe Ochsenfeld, Ludwig-Maximilians-Universität München, (2011). [2.4.1](#), [6.3](#), [6.3](#)
- [51] C. Saravanan, Y. Shao, R. Baer, P. N. Ross und M. Head-Gordon, *J. Comput. Chem.* **24**, 618 (2003). [2.4.2](#)
- [52] J. Kussmann, *Entwicklung linear skalierender ab initio Methoden basierend auf Elektronendichtematrizen*, Dissertation, Arbeitsgruppe Ochsenfeld, Universität Tübingen, (2006). [2.4.2](#)
- [53] D. A. Case, T. A. Darden, T. E. Cheatham(III), C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossvai, K. F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko und P. Kollman, (2010). [2.5](#), [4.4](#), [4.5](#), [5.3](#), [5.5](#), [5.6](#), [5.3](#), [6.2](#), [A](#), [A.1](#)
- [54] J. E. Jones, *Proc. R. Soc. London. Series A* **106**, 463 (1924). [2.5](#)
- [55] P. Rzepecki, K. Hochdörffer, T. Schaller, J. Zienau, K. Harms, C. Ochsenfeld, X. Xie und T. Schrader, *J. Am. Chem. Soc.* **130**, 586 (2008). [2.3](#), [5.4.3](#), [5.4.3](#), [5.3a](#)
- [56] J. C. Fromme und G. L. Verdine, *J. Biol. Chem.* **278**, 51543 (2003). [2.3](#), [5.4.3](#), [5.3d](#), [5.4.3](#), [6.1](#), [6.2](#), [6.2](#)
- [57] P. Sherwood, *Hybrid Quantum Mechanics/Molecular Mechanics Approaches*, in *Modern Methods and Algorithms of Quantum Chemistry*, Hrsg.: J. Grotendorst, Band 3, S. 285–305, John von Neumann Institute for Computing, Jülich, (2000). [2.6](#), [2.6](#)

- 
- [58] D. Bakowies und W. Thiel, *J. Phys. Chem.* **100**, 10580 (1996). 2.6
- [59] J. Almlöf, K. Faegri Jr. und K. Korsell, *J. Comput. Chem.* **3**, 385 (1982). 3.1
- [60] S. A. Maurer, D. S. Lambrecht, D. Flaig und C. Ochsenfeld, *J. Chem. Phys.* **136**, 144107 (2012). 3.1, 3.3, 3.3, 3.4, 7
- [61] B. Doser, D. S. Lambrecht, J. Kussmann und C. Ochsenfeld, *J. Chem. Phys.* **130**, 064107 (2009). 3.2.2
- [62] T. H. Dunning, *J. Chem. Phys.* **90**, 1007 (1989). 3.2.4, 5.2, 6.3, 6.1
- [63] J. Kussmann, Arbeitsgruppe Ochsenfeld, *openMP-Implementierung im Programmpaket Q-Chem*, unveröffentlicht (2011). 3.2.4
- [64] Entwicklungsversion des Programmpakets Q-CHEM, [www.q-chem.com](http://www.q-chem.com). 3.2.4, 4.1, 4.2, 4.3, 4.4, 4.5, 5.2, 5.4.2, 7, A, A.1
- [65] Open Multiprocessing (OpenMP), <http://openmp.org/wp/about-openmp/>. 3.2.4
- [66] D. S. Lambrecht und C. Ochsenfeld, *J. Chem. Phys.* **123**, 184101 (2005). 3.3
- [67] D. S. Lambrecht und C. Ochsenfeld, *J. Chem. Phys.* **136**, 149901 (2012). 3.3
- [68] R. Car und M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985). 4.1
- [69] P. Tangney, *J. Chem. Phys.* **124**, 044111 (2006). 4.1
- [70] P. Tangney und S. Scandolo, *J. Chem. Phys.* **116**, 14 (2002). 4.1
- [71] V. Wathelet, B. Champagne, D. H. Mosley, É. A. Perpiète und J.-M. André, *J. Mol. Struct. (THEOCHEM)* **425**, 95 (1998). 4.1
- [72] P. Pulay und G. Fogarasi, *Chem. Phys. Lett.* **386**, 272 (2004). 4.1, 4.2
- [73] J. M. Herbert und M. Head-Gordon, *J. Chem. Phys.* **122**, 3269 (2005). 4.1, 4.2
- [74] A. M. N. Niklasson, C. J. Tymczak und M. Challacombe, *Phys. Rev. Lett.* **97**, 123001 (2006). 4.1
- [75] A. M. N. Niklasson, *Phys. Rev. Lett.* **100**, 123004 (2008). 4.1

- [76] A. M. N. Niklasson, P. Steneteg, A. Odell, N. Bock, M. Challacombe, C. J. Tymczak, E. Holmström, G. Zheng und V. Weber, *J. Chem. Phys.* **130**, 214109 (2009). 4.1
- [77] P. Steneteg, I. A. Abrikosov, V. Weber und A. M. N. Niklasson, *Phys. Rev. B* **82**, 075110 (2010). 4.1
- [78] A. M. N. Niklasson und M. J. Cawkwell, *ArXiv e-prints 1203.6836* (2012). 4.1
- [79] P. Sherwood, A. H. de Vries, M. F. Guest, G. Schreckenbach, C. R. A. Catlow, S. A. French, A. A. Sokol, S. T. Bromley, W. Thiel, A. J. Turner, S. Billeter, F. Terstegen, S. Thiel, J. Kendrick, S. C. Rogers, J. Casci, M. Watson, F. King, E. Karlsen, M. Sjøvoll, A. Fahmi, A. Schäfer und C. Lennartz, *J. Mol. Struct. (THEOCHEM)* **632**, 1 (2003). 4.1, 4.4, 4.5, 5.4.3, 6.2, 7, A, A.1
- [80] M. Bühl und V. G. Malkin, *Calculation of NMR and EPR Parameters, Theory and Applications*, Wiley-VCH, Weinheim, (2004). 4.1, 5.1
- [81] C. J. Jameson, *Annu. Rev. Phys. Chem.* **47**, 135 (1996). 4.1
- [82] C. W. Swalina, E. P. O'Brien und G. Moyna, *Magn. Reson. Chem.* **40**, 195 (2002). 4.1
- [83] M. Robinson und P. D. Haynes, *J. Chem. Phys.* **133**, 084109 (2010). 4.1
- [84] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan und M. Karplus, *J. Comput. Chem.* **4**, 187 (1983). 4.4, 4.5
- [85] C. V. Sumowski, *Berechnung von NMR-chemischen Verschiebungen biomolekularer Systeme und QM/MM-Studien an Peptiden und RNA*, Dissertation, Arbeitsgruppe Ochsenfeld, Universität Tübingen, (2009). 4.4, 5.1
- [86] M. Hesse, H. Meier und B. Zeeh, *Spektroskopische Methoden in der organischen Chemie*, Thieme, Stuttgart, (1991). 5.1
- [87] T. Zuschneid, H. Fischer, T. Handel, K. Albert und G. Häfelinger, *Z. Naturforsch.* **59b**, 1153 (2004). 5.1
- [88] J. Gauss, *J. Chem. Phys.* **99**, 3629 (1993). 5.1

- 
- [89] J. R. Cheeseman, G. W. Trucks, T. A. Keith und M. J. Frisch, *J. Chem. Phys.* **104**, 5497 (1996). 5.1
- [90] L. Olsson und D. Cremer, *J. Chem. Phys.* **105**, 8995 (1996). 5.1
- [91] C. Adamo und V. Barone, *Chem. Phys. Lett.* **298**, 113 (1998). 5.1
- [92] P. R. Rablen, S. A. Pearlman und J. Finkbiner, *J. Phys. Chem. A* **103**, 7357 (1999). 5.1
- [93] H. Sun, L. K. Sanders und E. Oldfield, *J. Am. Chem. Soc.* **124**, 5486 (2002). 5.1
- [94] A. A. Auer, J. Gauss und J. F. Stanton, *J. Chem. Phys.* **118**, 10407 (2003). 5.1, 5.2, 5.1, 5.2, 5.2, 5.4, 5.5, 5.6
- [95] F. Jensen, *J. Chem. Theor. Comput.* **4**, 719 (2008). 5.1
- [96] Y. Zhao und D. G. Truhlar, *J. Phys. Chem. A* **112**, 6794 (2008). 5.1
- [97] J. Kongsted, K. Aidas, K. V. Mikkelsen und S. P. A. Sauer, *J. Chem. Theor. Comput.* **4**, 267 (2008). 5.1
- [98] F. A. A. Mulder und M. Filatov, *Chem. Soc. Rev.* **39**, 578 (2010). 5.1
- [99] A. Gregušová, S. A. Perera und R. J. Bartlett, *J. Chem. Theor. Comput.* **6**, 1228 (2010). 5.1
- [100] T. Kupka, M. Stachów, M. Nieradka, J. Kaminsky und T. Pluta, *J. Chem. Theor. Comput.* **6**, 1580 (2010). 5.1
- [101] C. Ochsenfeld, F. Koziol, S. P. Brown, T. Schaller, U. P. Seelbach und F.-G. Klärner, *Solid State Nucl. Magn. Reson.* **22**, 128 (2002). 5.1
- [102] J. Zienau, J. Kussmann und C. Ochsenfeld, *Mol. Phys.* **108**, 333 (2010). 5.1
- [103] K. Braunger, L. Kick und M. Thubauville, Bachelor-Arbeiten, Arbeitsgruppe Ochsenfeld, Ludwig-Maximilians-Universität München, (2012). 5.1
- [104] A. K. Jameson und C. J. Jameson, *Chem. Phys. Lett.* **134**, 461 (1987). 5.2

- [105] CFOUR, a quantum chemical program package written by J. F. Stanton, J. Gauss, M. E. Harding, P. G. Szalay with contributions from A. A. Auer, R. J. Bartlett, U. Benedikt, C. Berger, D. E. Bernholdt, Y. J. Bomble, O. Christiansen, M. Heckert, O. Heun, C. Huber, T.-C. Jagau, D. Jonsson, J. Jusélius, K. Klein, W. J. Lauderdale, D. A. Matthews, T. Metzroth, D. P. O'Neill, D. R. Price, E. Prochnow, K. Ruud, F. Schiffmann, S. Stopkowicz, A. Tajti, J. Vázquez, F. Wang, J. D. Watts and the integral packages MOLECULE (J. Almlöf and P. R. Taylor), PROPS (P. R. Taylor), ABACUS (T. Helgaker, H. J. Aa. Jensen, P. Jørgensen, and J. Olsen), and ECP routines by A. V. Mitin and C. van Wüllen. For the current version, see <http://www.cfour.de>. 5.2
- [106] A. Schäfer, H. Horn und R. Ahlrichs, *J. Chem. Phys.* **97**, 2571 (1992). 5.2, 5.1, 5.2, 5.2, 5.4, 5.5, 5.6, 5.4.2, 6.2, 6.3, 6.1, 6.2
- [107] A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988). 5.2, 5.1, 5.2, 5.3, 5.4, 5.4.4
- [108] T. W. Keal und D. J. Tozer, *J. Chem. Phys.* **119**, 3015 (2003). 5.2, 5.1, 5.2, 5.3, 5.4, 5.2
- [109] T. Helgaker, P. J. Wilson, R. D. Amos und N. C. Handy, *J. Chem. Phys.* **113**, 2983 (2000). 5.2, 5.1, 5.2, 5.3, 5.4, 5.2
- [110] J. Kussmann, Arbeitsgruppe Ochsenfeld, *Entwicklungsversion eines QM-Programmpackages*, unveröffentlicht (2012). 5.2
- [111] Bibliothek von Austausch-Korrelationsfunktionalen für Dichtefunktionaltheorie (libxc), [www.tddft.org/programs/octopus/wiki/index.php/Libxc](http://www.tddft.org/programs/octopus/wiki/index.php/Libxc). 5.2
- [112] W. J. Hehre, R. F. Stewart und J. A. Pople, *J. Chem. Phys.* **51**, 2657 (1969). 5.2, 5.4
- [113] IUPAC Comm. Nomenclature Org. Chem. and IUPAC-IUB Comm. Biochem. Nomenclature, *Biochemistry* **14**, 449 (1975). 5.3
- [114] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey und M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983). 5.3, 5.4.3
- [115] N. Ferré und J. G. Ángyán, *Chem. Phys. Lett.* **356**, 331 (2002). 5.4.1
- [116] H. Lin und D. Truhlar, *Theor. Chem. Acc.* **117**, 185 (2007). 5.4.1

- [117] W. J. Hehre, R. Ditchfield und J. A. Pople, *J. Chem. Phys.* **56**, 2257 (1972). 5.4.2
- [118] P. C. Hariharan und J. A. Pople, *Theor. Chim. Acta* **28**, 213 (1973). 5.4.2
- [119] F.-G. Klärner, B. Kahlert, A. Nellesen, J. Zienau, C. Ochsenfeld und T. Schrader, *J. Am. Chem. Soc.* **128**, 4831 (2006). 5.4.3, 5.3a, 5.4.3
- [120] S. Ware, W. F. Anderson, J. P. Donahue und J. Hawiger, *J. Prot. Sci.* **8**, 2663 (1999). 5.4.3, 5.3b, 5.4.3
- [121] W. Pisula, v. Tomović, M. D. Watson, K. Müllen, J. Kussmann, C. Ochsenfeld, T. Metzroth und J. Gauss, *J. Phys. Chem. B* **111**, 7481 (2007). 5.4.3, 5.3c, 5.4.3, 5.4.5
- [122] T. A. Halgren, *J. Comput. Chem.* **17**, 490 (1996). 5.4.3
- [123] F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson und W. C. Still, *J. Comput. Chem.* **11**, 440 (1990). 5.4.3
- [124] Desmond Molecular Dynamics System, D. E. Shaw Research, Maestro-Desmond Interoperability Tools, Version 2.4, New York, (2010). 5.4.3
- [125] X. Cheng, C. Kelso, V. Hornak, C. de los Santos, A. P. Grollman und C. Simmerling, *J. Am. Chem. Soc.* **127**, 13906 (2005). 5.4.3
- [126] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale und K. Schulten, *J. Comput. Chem.* **26**, 1781 (2005). 5.4.3, 6.2
- [127] A. H. de Vries, P. Sherwood, S. J. Collins, A. M. Rigby, M. Rigutto und G. J. Kramer, *J. Phys. Chem. B* **103**, 6133 (1999). 5.4.3
- [128] C. I. Bayly, P. Cieplak, W. Cornell und P. A. Kollman, *J. Phys. Chem.* **97**, 10269 (1993). 5.4.3
- [129] P. A. M. Dirac, *Camb. Phil. Soc.* **26**, 376 (1930). 5.4.4
- [130] C. Ochsenfeld, *Phys. Chem. Chem. Phys.* **2**, 2153 (2000). 5.4.5
- [131] C. Ochsenfeld, S. P. Brown, I. Schnell, J. Gauss und H. W. Spiess, *J. Am. Chem. Soc.* **123**, 2597 (2001). 5.4.5

- [132] D. E. Metzler, *Biochemistry, The Chemical Reactions of Living Cells, Vol. 2*, Elsevier Academic Press, Amsterdam, (2003). [6.1](#)
- [133] D. J. Voet, J. G. Voet und C. W. Pratt, *Principles of Biochemistry*, Wiley, Weinheim, (2010). [6.1](#)
- [134] J. C. Fromme, A. Benerjee und G. L. Verdine, *Curr. Opin. Struct. Biol.* **14**, 43 (2004). [6.1](#)
- [135] R. Strasser, *Studien zur Erkennung geschädigter DNA in der Nukleotid- und Basenexzisionsreparatur*, Dissertation, Arbeitsgruppe Carell, Ludwig-Maximilians-Universität München, (2010). [6.1](#)
- [136] M. Ober, *Synthese eines carbozyklischen Analogons des Formamidopyrimidin-dG DNA-Schadens: Untersuchung zur Replikation und Reparatur*, Dissertation, Arbeitsgruppe Carell, Ludwig-Maximilians-Universität München, (2005). [6.1](#)
- [137] D. O. Zharkov, R. A. Rieger, C. R. Iden und A. P. Grollman, *J. Biol. Chem.* **272**, 5335 (1997). [6.1](#)
- [138] O. V. Lavrukhin und R. S. Lloyd, *Biochem.* **39**, 15266 (2000). [6.1](#), [6.4](#)
- [139] F. Coste, M. Ober, T. Carell, S. Boiteux, C. Zelwer und B. Castaing, *J. Biol. Chem.* **279**, 44074 (2004). [6.1](#), [6.2](#)
- [140] F. Coste, M. Ober, Y.-V. Le Bihan, M. A. Izquierdo, N. Hervouet, H. Mueller, T. Carell und B. Castaing, *Chem. Biol.* **15**, 706 (2008). [6.1](#)
- [141] F. Büsch, J. Pieck, M. Ober, J. Gierlich, G. Hsu, L. Beese und T. Carell, *Chem. Eur. J.* **14**, 2125 (2008). [6.1](#)
- [142] K. Song, C. Kelso, C. de los Santos, A. P. Grollman und C. Simmerling, *J. Am. Chem. Soc.* **129**, 14536 (2007). [6.1](#)
- [143] K. Nam, G. L. Verdine und M. Karplus, *J. Am. Chem. Soc.* **131**, 18208 (2009). [6.1](#)
- [144] Y. Qi, M. C. Spong, K. Nam, A. Banerjee, S. Jiralerspong, M. Karplus und G. L. Verdine, *Nature* **462**, 762 (2009). [6.1](#)

- 
- [145] Y. Qi, M. C. Spong, K. Nam, M. Karplus und G. L. Verdine, *J. Biol. Chem.* **285**, 1468 (2010). 6.1
- [146] Kooperationsprojekt innerhalb des Sonderforschungsbereichs 749, Arbeitsgruppen von Prof. Carell (LMU), Prof. Zacharias (TU München) und Prof. Ochsenfeld (LMU). 6.1, 6.2, 6.3, 6.3
- [147] S. Grimme, J. Antony, S. Ehrlich und H. Krieg, *J. Chem. Phys.* **132**, 154104 (2010). 6.2, 6.3, 6.2, 6.3, 6.4, 6.3, 6.4, A, A.1
- [148] B. Jeziorski, R. Moszynski und K. Szalewicz, *Chem. Rev.* **94**, 1887 (1994). 6.3, 6.3
- [149] S. F. Boys und F. Bernardi, *Mol. Phys.* **19**, 553 (1970). 6.3, 6.1, 6.2, 6.3, 6.3
- [150] Y. Jung, R. C. Lochan, A. D. Dutoi und M. Head-Gordon, *J. Chem. Phys.* **121**, 9793 (2004). 6.3, 6.3
- [151] S. A. Maurer, M. Beer und C. Ochsenfeld, *in Vorbereitung* (2012). 6.3, 6.3
- [152] Q-Chem-ChemShell-Handbuch,  
<http://www.cse.scitech.ac.uk/ccg/software/chemshell/manual/qchem.html>. A



## Veröffentlichungen

D. Flaig and C. Ochsenfeld,

„Combining the Advantages of Semi-Direct Schemes and Linear-Scaling Self-Consistent Field Methods“,  
*Mol. Phys.* **108**, 2725 (2010).

D. Flaig, M. Beer, and C. Ochsenfeld,

„Convergence of Electronic Structure with the Size of the QM region: Example of QM/MM NMR Shieldings“,  
*J. Chem. Theory Comput.* **8**, 2260 (2012).

S. A. Maurer, D. S. Lambrecht, D. Flaig, and C. Ochsenfeld,

„Distance-Dependent Schwarz Based Integral Estimates for Two-Electron Integrals - Reliable Tightness vs. Rigorous Upper Bounds“,  
*J. Chem. Phys.* **136**, 144107 (2012).

## In Vorbereitung

D. Flaig, M. Hanni, M. Beer, K. Braunger, M. Thubauville, L. Kick, and C. Ochsenfeld,

„Intermediate References in Ab-initio Calculations of NMR shifts for Peptidic Systems“,  
*in preparation* (2012).

D. Flaig and C. Ochsenfeld,

„Extrapolation method for CPSCF calculations in Born-Oppenheimer trajectories“,  
*in preparation* (2012).