Entwicklung, Anwendung und Vergleich von Methoden zur Berechnung von Infrarotspektren einzelner Moleküle in polaren Lösungsmitteln

Matthias Schmitz

München 2004

Entwicklung, Anwendung und Vergleich von Methoden zur Berechnung von Infrarotspektren einzelner Moleküle in polaren Lösungsmitteln

Matthias Schmitz

Dissertation an der
Fakultät für Physik
Ludwig-Maximilians-Universität
München

vorgelegt von
Matthias Schmitz
aus Königswinter

München, im Oktober 2004

Erstgutachter: Prof. Paul Tavan
Zweitgutachter: Prof. Regina de Vivie-Riedle
Tag der mündlichen Prüfung: 30. 11. 2004

Zusammenfassung

Die Absorptionsspektroskopie im mittleren infraroten (MIR, etwa 500–4000 cm^{-1}) Spektralbereich ist ein wichtiges Hilfsmittel in der biomolekularen Forschung. Mit ihrer Hilfe können z.B. strukturelle Eigenschaften von Proteinen und enzymatisch katalysierte Reaktionen sichtbar gemacht werden. Zur Interpretation solcher Spektren benötigt man jedoch Rechenmethoden, vermittels derer MIR Spektren mit hoher Genauigkeit vorhergesagt werden können.

Im ersten Teil dieser Arbeit betrachte ich Cyclopyrimindimere (CPD), die durch ultraviolette Strahlung in dem Erbgutmolekül DNS entstehen und potentiell mutagene oder lethale Folgen für die Zelle haben. Um zukünftige Experimente zu leiten, welche die licht-induzierte Reparaturreaktion dieser Defekte durch das Enzym Photolyase mit zeitaufgelöster Spektroskopie verfolgen wollen, habe ich mit Hilfe der Dichtefunktionaltheorie (DFT) die MIR Spektren von Modellstrukturen einzelner Intermediate der CPD Reparatur berechnet. Die Ergebnisse zeigen, dass die Intermediate der CPD Reparatur anhand der spezifischen MIR Absorption ihrer Carbonylmoden identifiziert werden können und dass der Weg der Spaltung eines CPD in native DNS Basen auf diese Weise aufgeschlüsselt werden kann.

Im zweiten Teil dieser Arbeit untersuche ich jene Methoden, die in der Literatur zur Berechnung von MIR Spektren aus Molekulardyamik (MD) Simulationen, bei denen eine DFT Beschreibung eines Moleküls mit einer molekülmechanischen (MM) Beschreibung der Umgebung kombiniert wird, vorgeschlagen worden sind. Dazu leite ich die Vorschriften der verschiedenen Methoden für den Fall eines einzelnen Moleküls in einem polaren Lösungsmittel aus der linearen, quantenmechanischen Störungstheorie her. Anhand dieser Herleitung und der Ergebnisse einer exemplarischen Anwendung der Methoden auf eine DFT/MM-MD Simulation eines Formaldehydmoleküls in Wasser diskutiere ich die den jeweiligen Methoden zugrundeliegenden Annahmen und Näherungen sowie mögliche neue Verfahren zur Korrektur der durch die Näherungen induzierten Fehler bei der Berechnung von MIR Spektren.

Ferner entwickle ich aus dieser Analyse ein neues Verfahren zur Berechnung von MIR Spektren kleiner Moleküle in polaren Lösungsmitteln, mit dessen Hilfe sich die Frequenzfluktuationen des gelösten Moleküls, die durch die Wechselwirkung mit der fluktuierenden Lösungsmittelumgebung entstehen, mit einer Auflösung von etwa 10–30 fs berechnen lassen. Anhand einer exemplarischen Anwendung zeige ich, dass es diese Methode erlaubt, die Ursachen der Frequenzfluktuationen im Detail zu untersuchen und deren Beiträge zu MIR Linienbreiten zu ermitteln.

Inhaltsverzeichnis

Abkürzungsverzeichnis **v**

1 Einleitung **1**
1.1 Struktur von Proteinen 1
1.2 Experimentelle Methoden zur Strukturaufklärung von Proteinen . . . 3
1.3 Reparatur defekter DNS durch Photolyase 5
1.4 Theoretische Methoden 7
1.5 Infrarotspektroskopie und Ramanspektroskopie 8
1.6 Dekodieren von Infrarotspektren von Proteinen 9
1.7 Berechnung von Schwingungsspektren – Stand der Technik 12
1.8 Ziel und Gliederung dieser Arbeit 13

2 Schwingungsanalyse von Intermediaten der DNS Reparatur **15**
2.1 Schwingungsspektren neutraler und anionischer Modellstrukturen . . 15
2.2 Schwingungsspektren isotopenmarkierter Modellstrukturen 25

3 Berechnung von Infrarotspektren einzelner Moleküle in Lösungsmitteln **27**
3.1 Infrarotspektren aus Molekulardynamiksimulationen 27
3.2 Zeitaufgelöste Infrarotspektren und Vergleich verschiedener Methoden 43

4 Ausführliche Zusammenfassung und Ausblick **57**

A Über die in der DFT/MM Trajektorie beobachteten Gradientensprünge **65**

Literaturverzeichnis **69**

Danksagung **77**

Lebenslauf **79**

Abkürzungsverzeichnis

ATP	Adenosintriphosphat 1
CPD	Cyclopyrimidindimer 5
DFT	Dichtefunktionaltheorie 12
DNS	Desoxyribonukleinsäure 2
FADH	Flavin-Adenin-Dinukleotid 5
FTTCF	Fourier transform of the time correlation function 12
GV	Generalisiertes Virial 27, 58
INMA	Instantane Normalmodenanalyse 12
IR	Infrarot 2
MD	Molekulardynamik 7
MIR	Mittleres Infrarot (500–4000 cm^{-1}) 13
MM	Molekülmechanisch 7
NMA	Normalmodenanalyse 12
NMR	Nuclear Magnetic Resonance 4
PDB	Protein data bank 4
PMA	Principal Mode Analysis 12
RNS	Ribonukleinsäure 2
TF-GV	Zeitaufgelöstes, fehlerbereinigtes generalisiertes Virial 43, 59
UV	Ultraviolett 5

1 Einleitung

"I WILL NOT COMPLAIN ABOUT THE SOLUTION WHEN I HEAR IT"
(*B. Simpson* [4])

Proteine sind hochspezialisierte molekulare Maschinen und erfüllen in den Zellen der Lebewesen die verschiedensten Aufgaben. Abhängig von ihrem Aufbau leisten sie zum Beispiel unter ATP Verbrauch die für die Muskelkontraktion notwendige mechanische Arbeit (Myosin in der Muskelzelle [5]), wandeln Lichtenergie in chemische Energie (Bacteriorhodopsin in der Membram des *Halobacter Salinarium* [6]) oder in ein biochemisches Signal um (Rhodopsin in der Retina des Auges [5]). Ferner synthetisieren oder spalten sie als Glieder in der Kette des zellulären Stoffwechsels verschiedene Moleküle (z.B. ATP Synthase [7]), reparieren defektes Erbmaterial (z.B. DNS Photolyase [8]) oder prozessieren zelluläre Informationen (z.B. Ras p21 [9]).

Diese Funktionalitäten hat die Natur im Laufe eines Jahrmillionen dauernden Evolutionsprozesses hervorgebracht, während dessen die Größe, der Energieverbrauch und die Lebensdauer der Proteine für ihre spezifischen Aufgaben optimiert wurden. Daher motiviert neben einem grundsätzlichen, naturwissenschaftlichen Interesse auch die Vorbildfunktion, die Proteine bei der Entwicklung kleinster Bauteile für technische Anwendungen übernehmen können, dazu, diese elementaren Bausteine des Lebens und die durch sie vermittelten Prozesse zu untersuchen. Aber auch ein gezieltes Bekämpfen von all jenen Krankheiten, welche durch Fehlfunktionen von Proteinen verursacht werden, setzt ein genaues Verständnis der Funktionsweise der beteiligten Proteine voraus.

1.1 Struktur von Proteinen

Proteine bestehen aus Aminosäuren, deren Aufbau in Abbildung 1.1 schematisch gezeigt ist. Charakteristisch für Aminosäuren sind die Aminogruppe (NH_2), die Carboxylgruppe (COOH) und die Seitengruppen (R). Die zwanzig verschiedenen Sorten von Aminosäuren, aus denen ein Protein zusammengesetzt ist, unterscheiden sich durch diese Seitengruppen, die jeweils verschiedene chemische Eigenschaften haben. So kann man sie z.B. als polare, apolare, saure oder basische Seitengruppen klassifizieren [5].

Abbildung 1.1 veranschaulicht ferner die Peptidsynthese, bei der zwei Aminosäuren unter Abspaltung eines Wassermoleküls miteinander verbunden werden. Das

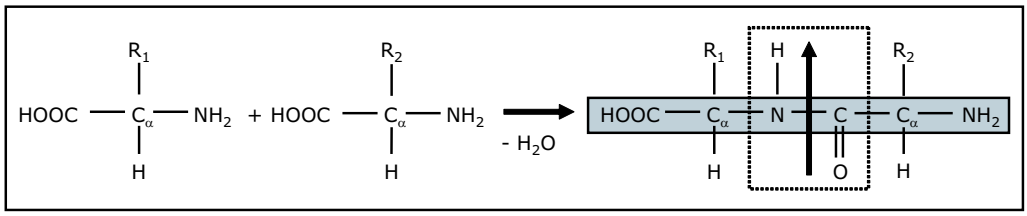

Abbildung 1.1: Chemischer Aufbau und Synthese von Aminosäuren: Eine Aminosäure (links) besteht aus einem zentralen Kohlenstoffatom (C_α), an das eine Aminogruppe (NH_2), eine Carboxylgruppe (COOH) und eine Seitengruppe (R) gebunden ist. Bei der Peptidsynthese werden zwei Aminosäuren unter Abspaltung eines Wassermoleküls (H_2O) miteinander verbunden. Dabei entsteht ein torsionsstabiles Peptidplättchen (gepunktet umrahmt), welches ein ausgeprägtes Dipolmoment (senkrechter Pfeil) besitzt. Das sogenannte Rückgrat (engl.: „backbone“) der Aminosäurekette ist grau hinterlegt.

dabei entstehende, in Abbildung 1.1 durch einen gestrichelten Rahmen hervorgehobene Peptidplättchen zeichnet sich durch eine hohe Polarität aus, da sowohl der Stickstoff (N) als auch der Sauerstoff (O) eine hohe Elektronegativität besitzen und demzufolge Elektronen von dem angrenzenden Wasserstoff (H) bzw. Kohlenstoff (C) abziehen.

Bei der Synthese eines Proteins im Ribosom, einem großen zellulären RNS/Protein-Komplex [5], werden typischerweise zwischen einigen 100 und einigen 1000 Aminosäuren durch Wiederholungen des in Abbildung 1.1 gezeigten Syntheseschrittes zu langen Ketten verbunden. Die Aminosäuresequenz dieser Ketten, die sogenannte „Primärstruktur“, ist dabei im genetischen Code der DNS verschlüsselt, der Zelle in Form der RNS verfügbar und für jedes Protein eindeutig definiert. Abhängig von der Primärstruktur nehmen die Aminosäureketten unter natürlichen Bedingungen spontan eine für jedes Protein spezifische dreidimensionale Struktur an, die man als „Tertiärstruktur“ bezeichnet. Der genaue Ablauf dieser sogenannten „Proteinfaltung“ ist zur Zeit noch Gegenstand aktueller Diskussion [10–13]. Ein wichtiger Schritt während des Faltungsprozesses ist aber unzweifelhaft die Ausbildung sogenannter „Sekundärstrukturmotive“, die sich in fast jedem Protein wiederfinden.

Abbildung 1.2 zeigt schematisch die beiden wichtigsten Sekundärstrukturmotive, die α-Helix und das β-Faltblatt. Wie diese Abbildung verdeutlicht, werden beide Strukturmotive durch die elektrostatische Wechselwirkung zwischen den Dipolen gegenüberliegender Peptidplättchen stabilisiert. Man beachte jedoch, dass sich die beiden Motive hinsichtlich der relativen Ausrichtung nebeneinanderliegender Dipole unterscheiden. Dieser Unterschied bewirkt, dass die Peptidplättchen in den jeweiligen Strukturen elektromagnetische Strahlung aus dem infraroten (IR) Spektralbereich unterschiedlich absorbieren, und kann daher in entsprechenden spektroskopischen Experimenten sichtbar gemacht werden [14–17]. Auf derartige Zusammenhänge zwischen strukturellen Eigenschaften und dem Absorptionsverhalten von IR Strahlung

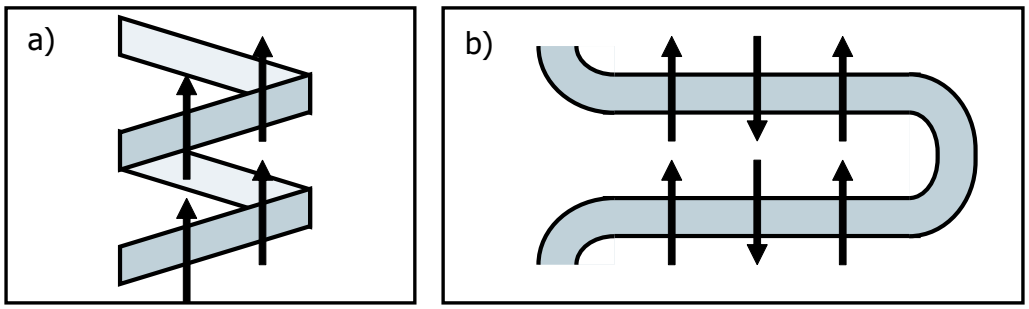

Abbildung 1.2: Die Sekundärstrukturmotive a) der spiralförmig gewundenen α-Helix und b) des parallelen β-Faltblatts. Der Verlauf der Aminosäurekette ist wie in Abb. 1.1 grau eingezeichnet. Beide Strukturen werden durch die elektrostatische Wechselwirkung zwischen den Dipolen (schwarze Pfeile) gegenüberliegender Peptidplättchen stabilisiert. Die Dipole nebeneinanderliegender Peptidplättchen sind in der α-Helix parallel und im β-Faltblatt antiparallel ausgerichtet.

werde ich später noch ausführlicher eingehen.

Zunächst sei aber noch festgehalten, dass die α-Helix und das β-Faltblatt vergleichsweise rigide Strukturen sind, die einem Protein in seinem funktionsfähigen („nativen“) Zustand seine definierte Struktur verleihen. Andere, flexiblere Teile eines Proteins, die häufig als offene Schleifen (engl. „loops“) vorliegen, ermöglichen dynamische Prozesse. Die spezifische Funktion eines Proteins wird also unmittelbar von seiner Struktur und von seiner Beweglichkeit determiniert [7, 18–20], die ihrerseits im physiologischen Milieu aus der Aminosäuresequenz folgen.

Während die Sequenz der Aminosäuren in einem Protein, abgesehen von posttranslationalen biochemischen Modifikationen, in der Regel aus der Basenfolge der DNS bestimmbar ist, ist es bis heute nicht möglich, aus dieser Sequenz die dreidimensionale Struktur eines gefalteten Proteins oder seine dynamischen Eigenschaften und damit seine Funktionsweise vorherzusagen. Um an diese Informationen zu gelangen, muss man daher das Protein in seiner gefalteten Struktur untersuchen. Da das eigentliche biologisch aktive Zentrum eines Proteins häufig nur einige wenige 10 Atome umfasst, muss man sich experimenteller und theoretischer Mittel bedienen, die eine atomare Auflösung haben, um solche Zentren genau zu charakterisieren. Die wichtigsten Methoden, die dieses Kriterium erfüllen, werde ich im folgenden Abschnitt vorstellen.

1.2 Experimentelle Methoden zur Strukturaufklärung von Proteinen

Die älteste und verbreitetste Methode zur Strukturaufklärung ist die Röntgenstrukturanalyse von Proteinkristallen, bei der die dreidimensionale Anordnung der Atome

des untersuchten Moleküls aus dem Beugungsmuster gestreuter Röntgenstrahlung berechnet wird [21, 22]. Die auf diese Weise gewonnenen Strukturen sind in der weltweit zugänglichen *protein data bank* (PDB) [23] gesammelt und haben einen bedeutenden Beitrag zum Verständnis der Funktionsweise von Proteinen geliefert (siehe z.B. [12, 20, 24]).

Der Röntgenstrukturanalyse sind aber durch drei wichtige Aspekte Grenzen gesetzt: Erstens muss das Protein für eine solche Analyse aus seiner ursprünglichen Umgebung entnommen, kristallisiert und bei tiefen Temperaturen bestrahlt werden. Da viele Proteine außerhalb ihrer ursprünglichen Umgebung und der dort vorhandenen thermodynamischen Bedingungen nicht funktionsfähig sind, ist nicht klar, ob sie bei dieser Präparation nicht auch ihre Struktur verändern („denaturieren“) und die ermittelte Röntgenstruktur damit nicht mehr die native Struktur widerspiegelt.

Der zweite Punkt betrifft die endliche experimentelle Auflösung, aufgrund derer die Positionen nicht aller Atome des Proteins aus dem Beugungsmuster eindeutig ermittelbar sind. Dies gilt insbesondere für die Wasserstoffatome, die einen kleinen Streuquerschnitt für Röntgenstrahlen haben und deswegen in den Beugungsmustern in der Regel unsichtbar sind. Die Information über den Protonierungszustand basischer bzw. saurer Seitengruppen des Proteins ist aber in vielen Fällen entscheidend für das Verständnis von Reaktionen in Proteinen, bei denen z.B. Protontransferprozesse stattfinden (z.B. bei Bacteriorhodopsin [6, 25–27]). Aber auch stark bewegliche Teile eines Proteins sind in Röntgenstrukturen häufig nicht sichtbar, obwohl gerade diese beweglichen Bereiche in vielen Fällen eine funktionale Bedeutung haben (z.B. die kupferbindende Oktarepeatsequenz des Prionproteins [28]).

Der dritte Punkt folgt aus der Tatsache, dass zum derzeitigen Stand der Technik hochauflösende Röntgenbeugungsmuster nur von statischen Strukturen möglich sind. Methoden zur Bestimmung zeitaufgelöster Röntgenstrukturen befinden sich erst in der Entwicklung [29]. Dynamische Prozesse, z.B. enzymatisch katalysierte Reaktionen oder Faltungsprozesse eines Proteins, können mit dieser Methode demnach nicht zeitaufgelöst verfolgt werden.

Als weitere experimentelle Methode zur Strukturaufklärung von Proteinen steht die Kernspinresonanz- (NMR, von engl. *nuclear magnetic resonance*) Spektroskopie zur Verfügung. Mittels dieser Methode kann der räumliche Abstand z.B. zwischen zwei Wasserstoffatomen aus der experimentell messbaren Stärke der Dipol-Kopplung ihrer jeweiligen Kernspins ermittelt werden, wenn der Abstand kleiner als 5 Å ist [30, 31]. Die Schwierigkeit besteht nun darin herauszufinden, welches Paar von Wasserstoffatomen des untersuchten Moleküls den so ermittelten Abstand voneinander hat. Diese Zuordnung wird mit zunehmender Anzahl von Wasserstoffatomen, also mit zunehmender Größe des Moleküls immer schwieriger, so dass diese Methode hauptsächlich für kleinere Proteine oder Peptide verwendet wird [32–34]. Jedoch reichen auch bei erfolgreicher Zuordnung aller gemessenen paarweisen Abstände die so gewonnenen Informationen nicht aus, um die dreidimensionale Struktur eines Proteins eindeutig bestimmen zu können. In einem anschließenden Strukturverfeinerungsprozess (engl. *NMR structure refinement*) werden daher zusätzliche Informationen,

z.B. die Topologie der Aminosäuren und empirisch bekannte Längen und Winkel chemischer Bindungen, dazu verwendet, eine oder mehrere Strukturen zu finden, die jeweils gleichzeitig die gemessenen paarweisen Abstände und die empirisch bekannten Eigenschaften des Proteins aufweisen [35]. Auf die molekülmechanischen Kraftfelder, die bei diesem *refinement* verwendet werden, werde ich später noch genauer eingehen.

Wie aus der Diskussion der Röntgenstrukturanalyse und der NMR Spektroskopie in diesem Abschnitt folgt, ist keine der beiden Methoden in der Lage, die Struktur und die Funktionsweise eines Proteins allein aus den jeweiligen experimentellen Daten abzuleiten. Als konkretes Beispiel dafür, welche interessanten Fragen diese Methoden offen lassen, und wie ergänzende Methoden bei der Klärung dieser Fragen hilfreich sein können, werde ich im folgenden die Reparatur geschädigter DNS durch das Enzym Photolyase anführen.

1.3 Reparatur defekter DNS durch Photolyase

Abbildung 1.3a zeigt die dreidimensionale Struktur eines kurzen Ausschnitts eines DNS Moleküls. Man erkennt, dass die DNS aus zwei sich umeinander windenden Strängen besteht. Diese Struktur nennt man Doppelhelix. Sie wurde im Jahre 1953 von Watson und Crick durch Röntgenstrukturanalyse entdeckt [36]. Jeder einzelne Strang besteht aus einer Kette von Desoxyribosen die über Phosphodiester miteinander verbunden sind [5]. An jeder Ribose befindet sich eine von vier Basen, entweder eine der Pyrimidinbasen Thymin (T) und Cytosin (C), oder eine der Purinbasen Adenin (A) und Guanin (G). In einem Doppelstrang liegen sich im Normalfall die Basen A und T, bzw. G und C gegenüber. Die Reihenfolge dieser Basen innerhalb eines Strangs kodiert die genetische Information, die in der DNS enthalten ist [5].

Ultraviolette (UV) Strahlung kann in der DNS Dimerisierungsreaktionen zwischen zwei benachbarten Pyrimidinbasen induzieren [38]. Die am häufigsten [39] auftretende, durch UV Strahlung induzierte Dimerisierungsreaktion ist in Abb. 1.3 abgebildet: Abb. 1.3b zeigt zwei Pyrimidinbasen im ursprünglichen, nativen Zustand als Teil der Sequenz A-T-T-A. Unter UV Bestrahlung bilden sich zwei chemische Bindungen, die in Abb. 1.3c sichtbar sind und die Pyrimidinbasen zu einem sogenannten Cyclopyrimidindimer (CPD) verbinden.

Dieser CPD Defekt kann für die Zelle mutagene oder lethale Folgen haben [40]. Um diese Gefahr zu beheben, hat die Natur im Laufe der Evolution eine Klasse von Proteinen namens Photolyase hervorgebracht, die man in verschiedenen Gattungen aller biologischen Domänen (Eukaryoten, Bakterien, Archaebakterien) findet, beim Menschen zum Beispiel jedoch nicht [8]. Dieses Photolyase Protein kann CPD Defekte mit hoher Effizienz entdecken und reparieren [8, 38, 41]. Dazu überträgt die Photolyase in einem lichtinduzierten Prozess von seinem Flavin Kofaktor (FADH) ein Elektron auf das Dimer und löst so die Spaltung des Dimers aus [42, 43]. Anhand der Kristallstrukturen der Photolyasen von *Escherichia Coli* (PDB Code 1dnp

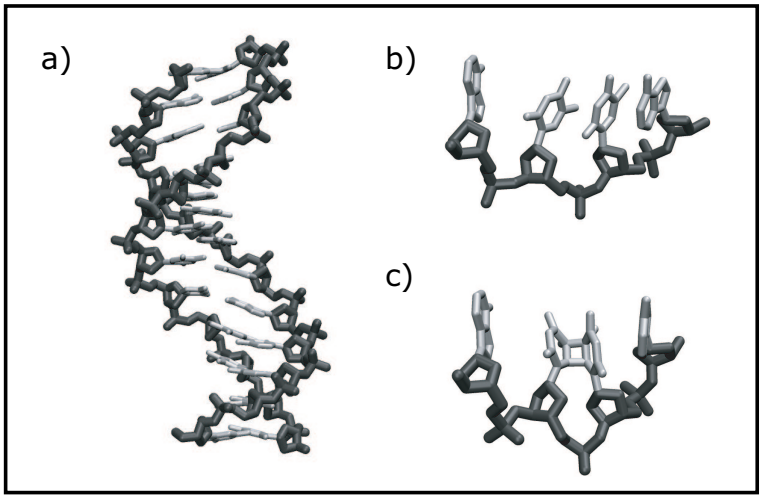

Abbildung 1.3: Dreidimensionale Struktur von DNS in nativer und UV geschädigter Form [37]: a) Doppelsträngige DNS mit 12 Basenpaaren in nativer Form (PDB Eintrag 1coc), b) Ausschnitt der Basensequenz A-T-T-A aus der nativen DNS, c) derselbe Ausschnitt mit dem UV induzierten CPD Defekt (PDB 1ttd).

[24]) und von *Anacystis nidulans* (PDB Code 1qnf [44]) konnte vorausgesagt werden, dass für einen direkten Elektrontransfer von dem FADH Kofaktor auf das Dimer ein Herausklappen des Dimers aus der Doppelhelix in die Bindungstasche der Photolyase notwendig ist. Die Existenz des DNS/Photolyase Komplex mit ausgeklappten DNS Basen in dieser Bindungstasche ist tatsächlich kürzlich in einer Röntgenstruktur nachgewiesen worden [45]. Nach der Spaltung des Dimers in der Bindungstasche der Photolyase und nach Rücktransfer des Elektrons an den FADH Kofaktor löst sich die DNS von dem Protein und stellt so den nativen Zustand sowohl der DNS als auch der Photolyase wieder her.

Die Röntgenstrukturen geben jedoch keine Auskunft darüber, wie die Spaltung des Dimers in der Bindungstasche genau abläuft. Fraglich ist, ob die zwei Bindungen, welche die beiden Pyrimidinbasen zusammenhalten, nach dem Elektrontransfer gleichzeitig [46, 47] oder nacheinander [8, 48] aufbrechen. In einem neueren Vorschlag wird sogar zusätzlich zu dem erwähnten Elektrontransfer auch ein Protontransfer von dem Protein auf das Dimer für möglich gehalten [49]. Unklar ist weiterhin, auf welchen Zeitskalen die Spaltung abläuft und ob das Protein die Spaltung des Dimers außer durch das erwähnte Bereitstellen eines Elektrons (möglicherweise auch eines Protons) zusätzlich z.B. durch sterische oder elektrostatische Einflüsse begünstigt.

Die Funktionsweise der Photolyase ist ein typisches Beispiel dafür, dass eine eindeutige Bestimmung eines dynamischen Vorgangs in einem Protein neben der Röntgen- oder NMR Strukturanalyse weitere Methoden erfordert. Zwei Klassen von Methoden, welche die Strukturanalyse ergänzen können, werde ich nun näher vorstellen.

1.4 Theoretische Methoden

Proteine lassen sich näherungsweise durch eine molekülmechanische (MM) Energiefunktion $V_K(\vec{r}_K)$ (auch „Kraftfeld“ genannt) beschreiben, die von den Kernkoordinaten $\vec{r}_K$ aller Atome des Proteins abhängt. Mit solchen Kraftfeldern lassen sich experimentelle Röntgen- bzw. NMR Daten durch Informationen über lokale chemische Strukturen (z.B. Bindungslängen und -winkel) ergänzen. Zum Beispiel können die Positionen von Atomen, die durch eine Röntgenstrukturanalyse nicht eindeutig bestimmt wurden, durch einfaches Minimieren von $V_K(\vec{r}_K)$ nachträglich näherungsweise ermittelt werden [50]. Bei dem oben schon angesprochenen NMR Struktur *refinement* werden *simulated annealing* Verfahren zur Minimierung der Potentialfunktion $V_K(\vec{r}_K)$ unter Verwendung zusätzlicher experimenteller Randbedingungen (engl. *constraints*) durchgeführt [35]. Diese *constraints* gewährleisten dabei die Einhaltung der durch NMR experimentell gemessenen paarweisen Abstände zwischen den Atomen.

Das molekülmechanische Kraftfeld $V_K(\vec{r}_K)$ wird darüberhinaus zusammen mit den Newtonschen Bewegungsgleichungen in sogenannten „Molekulardynamik“ (MD) Simulationen dafür verwendet, ausgehend von statischen Strukturen Informationen über die Dynamik der Proteine auf atomarer Ebene zu gewinnen. Für Proteine ist diese Methode erstmals für das „bovine pancreatic trypsin inhibitor“ (BPTI) Protein angewandt worden [51]. Darauffolgende Anwendungen haben zahlreiche wertvolle Informationen zur Funktionsweise von Proteinen aufgedeckt. Einige herausragende Beispiele finden sich in den Literaturzitaten [13, 34, 52–55].

Allerdings sind auch dieser theoretischen Methode Grenzen gesetzt. Heutige MD Simulationsprogramme [50, 56–58] verwenden in der Regel Kraftfelder [57–59], die zum Zwecke einer schnellen Berechenbarkeit zahlreiche Näherungen enthalten und dadurch erhebliche Artefakte produzieren können [60–62]. Außerdem können selbst moderne Computer trotz dieser Näherungen lediglich etwa 10^7 Integrationsschritte innerhalb eines Jahres Rechenzeit für ein Protein in realistischer Lösungsmittelumgebung bewältigen. Zur genauen Simulation chemischer Reaktionen (z.B. die oben erwähnte Spaltung des CPD Defektes) sind zudem aufwendige quantenmechanische Rechnungen erforderlich, welche die Anzahl der Integrationsschritte noch weiter auf etwa 10^5 reduziert.

Bei einer Integrationsschrittweite von typischerweise einer Femtosekunde (1 fs $= 10^{-15}$ s), die benötigt wird, um auch die Bewegung der schnellsten Freiheitsgrade des Systems, der Streckschwingungen kovalent gebundener Atome, hinreichend genau aufzulösen, entspricht das einer Simulationszeit von 0.1 bis 10 Nanosekunden (1 ns $= 10^{-9}$ s). Diese Simulationszeit ist um einige Größenordnungen kleiner als die Zeitskala von Millisekunden (1 ms $= 10^{-3}$ s) bis Minuten, auf der Faltungsprozesse oder Funktionszyklen von Proteinen in der Regel stattfinden [63–66]. Solche Prozesse sind MD Simulationen gegenwärtig also noch nicht zugänglich.

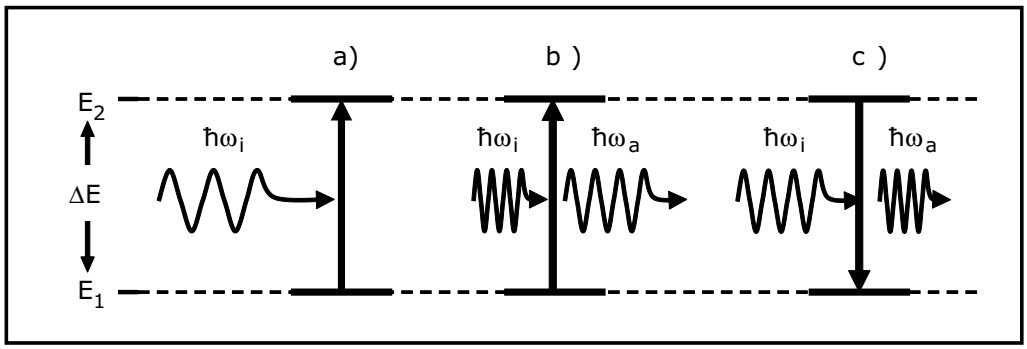

Abbildung 1.4: a) Absorptionsspektroskopie: Das diskrete Energiespektrum eines Moleküls ist hier durch zwei Zustände mit den Energien E_1 und E_2 repräsentiert. Eine Absorption eines Photons findet nur statt, wenn die Energie $\hbar\omega_i$ des Photons gleich der Energiedifferenz $\Delta E = E_2 - E_1$ ist. b+c) Ramanspektroskopie: b) Ist die Energie $\hbar\omega_i$ des einfallenden Photons größer als ΔE, so kann das Molekül die Energie ΔE in einem inelastischen Stoß aufnehmen. Die restliche Energie $\hbar\omega_a = \hbar\omega_i - \Delta E$ übernimmt ein gestreutes Photon niedrigerer Frequenz. c) Von angeregten Molekülen kann ein Photon die Energie ΔE aufnehmen und als höherfrequentes Photon der Energie $\hbar\omega_a = \hbar\omega_i + \Delta E$ gestreut werden.

1.5 Infrarotspektroskopie und Ramanspektroskopie

Um die Struktur und die Dynamik von Proteinen auf experimentelle Weise genauer aufzuklären, verwendet man häufig spektroskopische Methoden, bei denen entweder die Absorption oder die Streuung elektromagnetischer Strahlung (z.B. die „Ramanstreuung“) an Molekülen beobachtet wird. Beide Prozesse werden durch die quantenmechanische Natur der Molekülen bestimmt, derzufolge Moleküle ein diskretes Spektrum von Zuständen verschiedener Energien haben.

Abb. 1.4a veranschaulicht, dass ein Molekül ein Photon nur dann vollständig absorbieren kann, wenn die Energie $\hbar\omega_i$ des Photons, die über das Planck'sche Wirkungsquant $\hbar$ direkt mit der Frequenz ω_i des Photons verbunden ist, genau mit der Energiedifferenz ΔE zweier Zustände des Moleküls übereinstimmt. Über die frequenzabhängige Absorption von Licht kann man also das Spektrum der Anregungsenergien ΔE bestimmen. Anhand der Wellenlänge $\lambda = 2\pi c/\omega$ der absorbierten elektromagnetischen Welle, die im Vakuum über die Lichtgeschwindigkeit c mit seiner Frequenz zusammenhängt, unterscheidet man dabei grundsätzlich verschiedene Absorptionsmechanismen: Im ultravioletten (UV, 10 nm $< \lambda <$ 400 nm) und dem sichtbaren (vis, von engl. „visible“, 400 nm $< \lambda <$ 750 nm) Spektralbereich ist die Absorption eines Photons hauptsächlich mit elektronischen Anregungen verbunden, im infraroten (IR, 750 nm $< \lambda <$ 1 cm) Spektralbereich dagegen mit Anregungen molekularer Schwingungen. Die Absorptionswahrscheinlichkeit hängt dabei jeweils von der Änderung des Dipolmoments des Moleküls bei dieser Anregung ab [67, 68].

Wie in Abb. 1.4b skizziert ist, kann das Molekül die Anregungsenergie ΔE auch dann von einem einfallenden Photon aufnehmen, wenn die Energie $\hbar\omega_i$ des Photons größer als ΔE ist, so dass die Resonanzbedingung $\Delta E = \hbar\omega$ nicht erfüllt ist [69]. Ein solcher inelastischer Stoß ist im Vergleich zu einer resonanten Absorption zwar viel unwahrscheinlicher, kann jedoch in Experimenten bei der sogenannten „Ramanspektroskopie“ beobachtet werden. Die nach dem Stoß verbleibende Energie $\hbar\omega_a = \hbar\omega_i - \Delta E$, die um die Anregungsenergie ΔE geringer ist als die Energie $\hbar\omega_i$ des einfallenden Photons, wird von dem gestreuten Photon wieder mitgenommen. Daher besitzt das gestreute Photon eine niedrigere Frequenz als das einfallende („Stokes Shift“). Abb. 1.4c zeigt ferner, dass das Photon in einem komplementären Prozess von einem thermisch angeregten Molekül Energie aufnehmen und als höherfrequentes Photon der Energie $\hbar\omega_a = \hbar\omega_i + \Delta E$ gestreut werden kann („Anti-Stokes Shift“) [69]. Im Gegensatz zu dem Absorptionsprozess hängt die Wahrscheinlichkeit für einen solchen Streuprozess quadratisch von der Änderung der Polarisierbarkeit des Moleküls bei dieser Anregung ab [67]. Abgesehen von diesem Unterschied kann man mit der Ramanspektroskopie aus dem Frequenzspektrum der gestreuten Photonen prinzipiell dasselbe Spektrum von Anregungsenergien ΔE ermitteln, wie mit der Absorptionsspektroskopie im IR Spektralbereich.

Bei der Anwendung dieser spektroskopischen Methoden ist entscheidend, dass das Spektrum der Anregungsenergien für jede Molekülsorte charakteristisch ist [70]. Zusätzlich ist das Spektrum eines Moleküls aber auch abhängig von seiner Polarisation, die es durch die elektrostatische Wechselwirkung mit seiner Umgebung erfährt (Solvatochromieeffekt). Als Beispiel habe ich die IR Absorption eines Peptidplättchens, die in einer α-Helix und in einem β-Faltblatt unterschiedlich ist [14–17], schon erwähnt. Allgemein folgt aus dieser Betrachtung, dass ein experimentell gemessenes Anregungsspektrum eines Moleküls strukturelle Informationen über das Molekül und über seine Umgebung in kodierter Form enthält. Zeitaufgelöste Spektren enthalten folglich kodierte Informationen über die Dynamik dieser Strukturen.

Das Problem besteht nun darin, diese strukturellen Informationen aus den Anregungsspektren zu dekodieren. Da sich die vorliegende Arbeit ausschließlich mit der Berechnung der Absorption von Molekülen im IR Spektralbereich befasst, werde ich einen möglichen Weg zur Lösung dieses Problems anhand von IR Spektren erläutern. Die jeweiligen Schritte sind jedoch auch auf andere Spektralbereiche und auf die Ramanspektroskopie übertragbar.

1.6 Dekodieren von Infrarotspektren von Proteinen

Ein nichtlineares Molekül mit N Atomen besitzt $3N-6$ Schwingungsmoden [67, 68], die durch Absorption eines IR Quants angeregt werden können. Bei Proteinen mit etwa 10^3-10^4 Atomen liegen die Absorptionsfrequenzen der einzelnen Moden so dicht beieinander, dass sie in einem experimentellen IR Spektrum ein quasi-kontinuierliches Absorptionsspektrum bilden.

Ein charakteristisches Merkmal der IR Spektren von Proteinen sind dabei die sogenannten Amidbanden, die dadurch entstehen, dass jedes einzelne Peptidplättchen des Proteins charakteristische Schwingungsmoden, die sogenannten Amidmoden, besitzt. Die Streckschwingung der C=O Bindung bezeichnet man z.B. als Amid-I Mode, die N–H Biegeschwingung in der Ebene des Peptidplättchens als Amid-II Mode, und so weiter. Durch die orientierungsabhängige Dipol-Wechselwirkung sind die Schwingungen der Amidmoden verschiedener Peptidplättchen eines Proteins stark untereinander gekoppelt. Zusätzlich wird jedes einzelne Peptidplättchen durch seine spezifische Umgebung individuell polarisiert. Die dadurch verursachten Verschiebungen der Absorptionsfrequenzen der Amidmoden spiegeln sich in den verbreiterten Absorptionsbanden wieder [14–16]. Die Amid-I Bande erstreckt sich zum Beispiel von 1600 bis 1700 cm^{-1}, die Amid-II Bande von 1530 bis 1580 cm^{-1} [16]. Die genaue Form dieser Banden gibt prinzipiell Aufschluss über die relative Orientierung der Peptidplättchen zueinander und damit über die Struktur des Proteins [16]. Methoden zur genauen Dekodierung dieser Information befinden sich jedoch erst in der Entwicklung [16, 71, 72].

Der experimentelle Zugang zu der IR Absorption des Zentrums der biologischen Aktivität eines Proteins, das sich wie erwähnt häufig nur über ein paar wenige 10 Atome erstreckt, ist schwierig, da die Absorption dieser Strukturen meist hinter den starken Amidbanden, der Absorption der Seitengruppen des Proteins und der Absorption des Lösungsmittels (insbesondere des Wassers) verborgen ist. Um diesen sogenannten Absorptionshintergrund zu beseitigen, bedient man sich der sogenannten Differenzspektroskopie. Dabei werden zwei IR Spektren des gleichen Proteins in verschiedenen Zuständen voneinander abgezogen. Der Unterschied zwischen den beiden Zuständen kann zum Beispiel durch eine Isotopenmarkierung, eine Punktmutation oder eine chemische Reaktion hervorgerufen werden. In der Differenz der beiden Spektren ist dann nur noch die Absorption jener Strukturen sichtbar, deren IR Absorption sich durch die Modifikation ändert. Je nach Art der Modifikation sind das z.B. in Proteine eingelagerte Kofaktoren [17, 73–78], Teile des Proteins [16, 17] oder Reaktanden der vom Protein katalysierten Reaktion [78, 79].

Die Abhängigkeit der IR Spektren dieser Fragmente von ihrem chemischen Zustand oder der Struktur ihrer Umgebung kann man nun ausnutzen, um strukturelle Informationen aus IR Spektren zu gewinnen [17, 76, 80, 81]. Um zu zeigen, wie man dabei üblicherweise vorgeht, verwende ich als Beispiel den Verlauf einer von einem Protein katalysierten chemischen Reaktion

$$\text{Edukt (E)} \rightarrow \text{Produkt (P)} \tag{1.1}$$

die zum Zeitpunkt t_0 starte und zum Zeitpunkt t_2 beendet sei.

Die schwarze Linie in Abbildung 1.5 zeigt schematisch die Änderung der IR Absorption des Reaktanden während der Reaktion. Man stellt fest, dass der Reaktand zu den drei verschiedenen Zeitpunkten t_0, t_1 und t_2 der Aufnahme jeweils verschiedene IR Spektren besitzt: Zum Zeitpunkt t_0 des Starts der Reaktion beobachtet man das IR Spektrum des Edukts, zum Zeitpunkt t_2 des Endes der Reaktion das IR

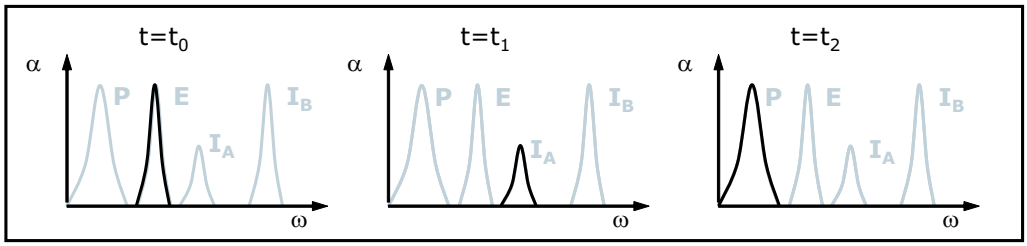

Abbildung 1.5: Die schwarze Linie zeigt schematisch das zeitaufgelöste IR Spektrum des Reaktanden der Reaktion (1.1). Die grauen Linien zeigen die berechneten und als fehlerfrei angenommenen IR Spektren der möglicherweise auftretenden Intermediate I_A und I_B sowie des Edukts (E) und des Produkts (P). Durch Vergleich der berechneten Spektren mit dem experimentellen IR Spektrum kann man schließen, dass der Reaktand zum Zeitpunkt t_1 der Reaktion als Intermediat I_A vorliegt.

Spektrum des Produkts. Aus der experimentellen Beobachtung, dass sich das Spektrum des Reaktanden zum dazwischenliegenden Zeitpunkt t_1 signifikant von dem des Produkts und dem des Edukts unterscheidet, kann man schließen, dass das Molekül zu diesem Zeitpunkt in einer dritten, intermediären Struktur vorliegt. Anhand des Spektrums kann jedoch nicht bestimmt werden, welche Struktur dieses Intermediat besitzt. Eine Möglichkeit, diese strukturelle Information aus dem Spektrum zu dekodieren besteht darin, die experimentellen Spektren folgendermaßen mit berechneten IR Spektren zu vergleichen:

Für die Reaktion (1.1) seien zwei verschiedene Wege postuliert worden. Der eine Weg laufe über das Intermediat I_A, der andere Weg über das Intermediat I_B. Die grauen Linien in Abbildung 1.5 zeigen die berechneten IR Spektren des Edukts (E), des Produkts (P) sowie jedes einzelnen Intermediats. Das schwarz gezeichnete, experimentelle Spektrum weist zum Zeitpunkt t_1 die beste Übereinstimmung mit dem berechneten Spektrum des Intermediats I_A auf. Nimmt man an, dass die Spektren allesamt fehlerfrei sind und die Reaktion nicht auf einem anderen, dritten Weg abläuft, kann man daraus schließen, dass die Reaktion über das Intermediat I_A abläuft.

Man erkennt an Abb. 1.5, dass eine zuverlässige Identifizierung von Intermediaten in zeitaufgelösten IR Spektren empfindlich von der Genauigkeit der berechneten und der experimentellen IR Spektren abhängt: Wäre der Fehler in den in Abb. 1.5 gezeigten Spektren größer als die Differenz zwischen ihnen, so wäre die eindeutige Identifikation eines Intermediats nicht möglich. Allgemeiner formuliert sind genaue Vorhersagen und genaue Messungen von IR Spektren zwingend notwendig, um experimentelle IR Spektren zuverlässig interpretieren zu können.

1.7 Berechnung von Schwingungsspektren – Stand der Technik

Durch die Entwicklung der Dichtefunktionaltheorie (DFT) [82, 83] erlaubt es der derzeitige Stand der Technik, die IR Spektren von mittelgroßen Molekülen (etwa einige 10 Atome) in der Gasphase mit hoher Genauigkeit zu berechnen [84]. Bei der dabei verwendeten Normalmodenanalyse (NMA) werden die Absorptionsfrequenzen des Moleküls in harmonischer Näherung aus der Hesse Matrix der zweiten Ableitungen der Grundzustandsenergie nach den Kernkoordinaten berechnet [67, 68]. Aus höheren Ableitungen können anharmonische Effekte störungstheoretisch berechnet werden [84, 85]. Die dafür notwendigen und in weit verbreiteten Softwarepaketen [86–89] implementierten DFT Verfahren sind allgemein als Standard anerkannt. Für die Berechnung von IR Spektren von Molekülen in kondensierter Phase, z.B. von Molekülen in Lösung oder in einer komplexen Proteinumgebung, sind diese Verfahren aufgrund ihres großen Rechenaufwands jedoch nicht geeignet.

Zur Untersuchung solch komplexer Systeme in kondensierter Phase stehen stattdessen zwei Möglichkeiten zur Verfügung: Eine Möglichkeit besteht darin, das Gesamtsystem durch kleine Modellsysteme zu repräsentieren, deren IR Spektren sich mit den oben erwähnten DFT Verfahren berechnen lassen. Hierbei kann jedoch der Fehler, der in den kleinen Modellsystemen z.B. durch die Vernachlässigung von langreichweitigen elektrostatischen Wechselwirkungen entsteht, unter Umständen eine eindeutige Interpretation von experimentellen IR Spektren verhindern [90, 91].

Eine genauere Modellierung eines Moleküls in komplexer Umgebung ist durch Hybridbeschreibungen möglich, welche die hochgenaue DFT mit der Effizienz von molekülmechanischen (MM) Kraftfeldern kombinieren [92, 93]. Durch deren Entwicklung lässt sich die Polarisation eines Moleküls durch seine Umgebung zumeist hinreichend genau erfassen [71, 76, 81, 91, 93, 94]. Zur expliziten Berechnung von IR Spektren aus entsprechenden DFT/MM-MD Simulationen sind die instantane Normalmodenanalyse (INMA), die Fourier Transformation der Zeitkorrelationsfunktion (FTTCF) des Dipolmoments und die sogenannte *principal mode analysis* (PMA) vorgeschlagen und angewendet worden.

Bei der INMA wird eine MD Trajektorie berechnet, um ein thermisches Ensemble von Umgebungsstrukturen zu generieren. Für jede Struktur dieses Ensembles wird die Umgebung des Moleküls festgehalten und die Energie des Moleküls in diesem „gefrorenen Käfig“ je nach Protokoll entweder minimiert [76, 81, 91, 93–95] oder nicht [96, 97]. Daraufhin werden die Schwingungsfrequenzen des Moleküls in dem Käfig analog zur Gasphase durch Normalmodenanalyse berechnet. Bei der FTTCF Methode wird die Korrelationsfunktion des Dipolmoments eines Moleküls aus einer MD Trajektorie gewonnen und zur Berechnung des IR Spektrums fouriertransformiert [91, 98, 99]. Bei der PMA wird die Kovarianzmatrix der Auslenkungen des Moleküls aus einer MD Trajektorie ermittelt. Die effektiven, mittleren Absorptionsfrequenzen des Moleküls werden dann aus den Eigenwerten der Kovarianzmatrix

berechnet [100, 101].

Die im Zusammenhang mit diesen Methoden zitierten Arbeiten haben sich in erster Linie zum Ziel gesetzt, die Genauigkeit der verwendeten Methode zusammen mit der verwendeten DFT/MM Beschreibung *für das jeweils betrachtete System* z.B. durch einen Vergleich mit experimentellen Daten nachzuweisen. Grundlegende methodische Fragen zur genauen Berechnung von IR Spektren von Molekülen in kondensierter Phase streifen sie nur am Rande. Eine rigorose theoretische Herleitung der jeweiligen Methoden bleiben sie schuldig.

Dieser Mangel ist die Ursache für die in den zitierten Studien dokumentierte Diskussion über die richtige Wahl der zu verwendenden Methode [76, 91, 93, 94, 100, 101]. Wie ich zeigen werde, stehen dabei die beiden folgenden offenen Fragen im Vordergrund: Welche Näherungen liegen den jeweiligen Methoden zugrunde und welche Fehler verursachen diese Näherungen in den berechneten Spektren?

1.8 Ziel und Gliederung dieser Arbeit

Das erste Ziel meiner Arbeit war die Klärung der Frage, anhand welcher spektraler Signaturen Intermediate der CPD Reparatur durch die Photolyase in zeitaufgelösten IR Spektren identifiziert werden können. Zu diesem Zweck habe ich die IR Spektren von Modellstrukturen dieses Systems mit Standard DFT Methoden berechnet. Die entsprechenden Resultate habe ich zum Teil zusammen mit Paul Tavan und Marco Nonella in dem Artikel [1] publiziert, der als Abschnitt 2.1 in die vorliegende Arbeit eingebunden ist. Im Abschnitt 2.2 werde ich diese Ergebnisse durch eine Präsentation und Diskussion weiterer, bisher unveröffentlichter Berechnungen von IR Spektren isotopenmarkierter Modellstrukturen ergänzen.

Aus den bisherigen Diskussionen in dieser Einleitung folgt, dass eine genauere Berechnung der IR Spektren der Intermediate der CPD Reparatur die Anwendung der im letzten Abschnitt erwähnten DFT/MM Beschreibungen erfordert. Es war jedoch bisher nicht klar, unter welchen Umständen die bisher vorgeschlagenen Methoden INMA, FTTCF und PMA das IR Spektrum eines Moleküls in kondensierter Phase genau berechnen können.

Um diese Frage zu klären, war es das zweite und eigentliche Hauptziel meiner Arbeit, die bisher vorgeschlagenen Vorschriften zur Berechnung von IR Spektren einzelner Moleküle in kondensierter Phase aus DFT/MM Simulationen theoretisch zu begründen, weiterzuentwickeln, auf Modellsysteme anzuwenden und miteinander zu vergleichen. Dabei beschränke ich mich auf das Mittel-IR (MIR, $0.25\ \text{mm} < \lambda < 2\ \text{mm}$, oder $500\ \text{cm}^{-1} < \lambda^{-1} < 4000\ \text{cm}^{-1}$) Spektrum eines kleinen Moleküls in einem polaren Lösungsmittel, weil diese Beschränkung, angesichts des großen Aufwands solcher Rechnungen, die Erhebung einer relativ guten statistischen Datenbasis ermöglicht. Ferner können durch diese Beschränkung komplizierte intermolekulare Kopplungen ausgeschlossen werden. Dennoch eignet sich dieses Modellsystem dazu, erstens, den Einfluss der Polarisation durch eine komplexe Umgebung auf das IR

Spektrum eines Moleküls zu untersuchen und, zweitens, zu prüfen, welche numerischen Vorschriften zur Berechnung von IR Spektren geeignet sind, diesen Einfluss genau zu analysieren. Die Ergebnisse dieses zweiten Teils meiner Arbeit habe ich zusammen mit Paul Tavan in den Artikeln [2] und [3] ausführlich beschrieben. Diese beiden Artikel werden in Kürze veröffentlicht und sind als Abschnitt 3.1 bzw. 3.2 in die vorliegende Arbeit eingebunden.

Im anschließenden Kapitel 4 fasse ich alle wichtigen Ergebnisse dieser Arbeit noch einmal ausführlich zusammen.

2 Schwingungsanalyse von Intermediaten der DNS Reparatur

Die berechneten IR Spektren neutraler und anionischer Modellstrukturen von Intermediaten der CPD Reparatur werde ich im nun folgenden Abschnitt 2.1 beschreiben. Dieser Abschnitt 2.1 ist ein Abdruck des Artikels[1]

> Matthias Schmitz, Paul Tavan, and Marco Nonella: „Vibrational Analysis of Carbonyl Modes in Different Stages of Light-Induced Cyclopyrimidine Dimer Repair Reactions“, Chem. Phys. Lett. **349**, 342-348 (2001),

den ich zusammen mit Paul Tavan und Marco Nonella in der Zeitschrift *Chemical Physics Letters* veröffentlicht habe. Die dort beschriebenen Ergebnisse werde ich im Abschnitt 2.2 durch weitere, nicht veröffentlichte Berechnungen von IR Spektren isotopenmarkierter Modellstrukturen ergänzen.

2.1 Schwingungsspektren neutraler und anionischer Modellstrukturen

[1] Mit freundlicher Genehmigung des Elsevier Verlags.

30 November 2001

Chemical Physics Letters 349 (2001) 342–348

CHEMICAL PHYSICS LETTERS

www.elsevier.com/locate/cplett

Vibrational analysis of carbonyl modes in different stages of light-induced cyclopyrimidine dimer repair reactions

Matthias Schmitz, Paul Tavan, Marco Nonella *

Lehrstuhl für BioMolekulare Optik, Ludwig-Maximilians-Universität München, Oettingenstr. 67, D-80538 München, Germany

Received 31 July 2001; in final form 27 September 2001

Abstract

The formation of cyclopyrimidine dimers is a DNA defect, which is repaired by the enzyme DNA photolyase in a light-induced reaction. Radical anions of the dimers have been suggested to occur as short-lived intermediates during repair. For their identification time-resolved Fourier-transform infrared (FTIR) spectroscopy will be a method of choice. To support and guide such spectroscopic studies we have calculated the vibrational spectra of various pyrimidine compounds using density functional methods. Our results suggest that the carbonyl vibrations of these molecules can serve as marker modes to identify and distinguish intermediates of the repair reaction. © 2001 Elsevier Science B.V. All rights reserved.

1. Introduction

The formation of cyclopyrimidine dimers is a frequent DNA-defect caused by UV-radiation. Cells protect themselves against such damages by excision repair or by photoreactivation. Here, the enzyme DNA photolyase binds specifically to such lesions and catalyzes monomerization by light-induced electron transfer from an FAD cofactor to the dimer [1–3]. For this process the reaction path depicted in Fig. 1 has been suggested: After electron transfer, the C_5—C'_5 bond of the dimer splits leading to an anionic radical dimer intermediate. The subsequent splitting of the C_6—C'_6 bond is assumed to occur in a thermal reaction catalyzed by the enzyme [2]. Electron transfer from the anionic pyrimidine back to the FAD cofactor finally restores the native state of the DNA and the initial state of the photolyase.

However, quantum mechanical calculations of model systems in vacuo have predicted a high-activation barrier for the thermal cleavage of the C_6—C'_6 bond in the anionic radical dimer intermediate. Therefore, a simultaneous splitting of the C_5—C'_5 and C_6—C'_6 bonds upon photoattachment of an electron to the cyclopyrimidine dimer has been proposed [4]. This alternative mechanism would lack the singly connected anionic dimer intermediate.

Time-resolved Fourier-transform infrared (FTIR) spectroscopy is a promising approach to resolve such a discussion on reaction paths by experimental means, if the corresponding spectra

* Corresponding author. Fax: +49-89-21-78-29-02.
E-mail address: nonella@physik.uni-muenchen.de (M. Nonella).

0009-2614/01/$ - see front matter © 2001 Elsevier Science B.V. All rights reserved.
PII: S0009-2614(01)01221-0

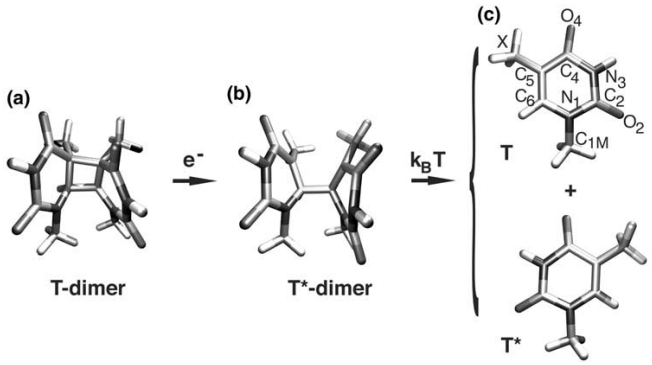

Fig. 1. Reaction path as proposed by Sancar [2]. In the drawing the deoxyribose linkages of the pyrimidine bases in DNA are replaced by methyl groups at the N_1 positions rendering *1-methyl-thymine* derivatives as *model compounds*. The resulting model of the dimeric defect is denoted by **T-dimer**, that of the suggested anionic intermediate by **T*-dimer** and those of the products by **T*** for the anionic and by **T** for the neutral monomer. Also indicated are the atom names used for the various pyrimidine derivatives studied here. A prime indicates an atom of the second monomer. X may stand for CH_3 (**T** as shown in the figure), H (1-methyl-uracil), or F (1-methyl-5-flouro-uracil).

allow an interpretation of observed bands in terms of structural intermediates. Interpretations are strongly facilitated whenever the reactants, intermediates and products exhibit intense and well-separated bands. Examples for successful applications of this approach are investigations of the electron transfer processes in photosynthetic reaction centers (see, for example [5,6]), of the photocycle of *Bacteriorhodopsin* (see, for example [7,8]), or of the GTP hydrolysis in Ras P21 [9,10].

The vibrational spectra of pyrimidine monomers are well characterized both by experimental and theoretical methods [11–15], whereas only a few experimental data and, to our knowledge, no theoretical descriptions are available for the vibrations of the pyrimidine dimers [16,17]. For the monomers infrared and Raman spectroscopy as well as ab initio calculations have shown that a series of strongly coupled ring modes are found in the congested frequency range between 900 and 1500 cm s^{-1}. The C=C stretch and the two C=O stretches, on the other hand, are well separated in the range 1600–1800 cm^{-1}.

The isolated spectral positions and the large intensities of the C=O stretching modes suggest that they might represent suitable candidates for the desired marker modes. It is the aim of this paper to check, whether this hope can be substantiated by theoretical arguments.

For this purpose we will apply a quantum mechanical vibrational analysis using density functional theory (DFT) to a set of models for the molecular states of the reaction scheme in Fig. 1. This analysis will indicate to what extent the C=O bands can change their spectral positions during light-induced DNA repair. It will provide estimates for the spectral positions of the C=O bands in the various stages of the reaction. Thus, it will show whether one can hope to distinguish and identify the various compounds by time-resolved FTIR observation of the C=O bands.

DFT is our method of choice for this purpose since it has proven to provide highly reliable molecular force fields [18–20]. Of course, one has to expect that the enzyme environment will affect these force fields particularly in the case of the highly polarizable anionic compounds through electrostatic interactions. Therefore, our in vacuo calculations are only a first step in the characterization of the vibrational spectra relevant for the analysis of the light-induced DNA repair. But they do provide the necessary reference for intended descriptions of pyrimidines within the enzyme environment by so-called hybrid methods [21]. Fur-

thermore, they can guide ongoing efforts to characterize the repair reaction by time-resolved FTIR spectroscopy [22].

2. Compounds and computational methods

The vibrational spectra of thymine and uracil derivatives, of associated cyclobutane-dimers and of corresponding radical anions have been calculated by DFT (cf. the caption to Fig. 1 for a characterization of these compounds). Hessian matrices in Cartesian coordinates were computed with the program GAUSSIAN 98 [23] using the gradient corrected BLYP [24,25] functional and a 6-31G* basis set [26–28]. The Hessians were then transferred to the program GAMESS [29] for the calculation of normal modes, of force constants for internal coordinates and of total energy distributions serving to characterize the normal modes in terms of internal coordinates.

3. Results and discussion

3.1. The monomers and the effects of hydrogen bonds

To test the applied method and to estimate the effects of hydrogen bonding we have first calculated the vibrational spectra of thymine and of 1-methyl-thymine **T**. The C=O and C=C frequencies are collected in Table 1 and are compared with experimental data as well as with results of previous calculations.

First, consider the case of thymine. Inspection of Table 1 immediately shows, that the calculated spectral positions of the two C=O modes (first column) are in excellent agreement with available experimental data [12,15]. Furthermore, our harmonic DFT force field underestimates the frequency of the C_5=C_6 mode only slightly by about 20 cm^{-1}. Note, that this nice agreement is found without scaling of force constants. In Table 1 we also quote the results of a recent DFT study applying the B3LYP functional and a 6-31G(d,p) basis set [15]. A comparison with the experimental data demonstrates that these predictions represent sizable overestimates. Here, a scaling procedure was required to achieve a satisfactory agreement. The excellent results for the thymine monomer let us expect that the BLYP functional combined with a standard 6-31G* basis set should yield accurate predictions also for the C=C and C=O modes of related compounds.

Consideration of DNA structures shows that the oxygen O_2 of thymine is in close contact with a proton of the deoxyribose moiety. To model this interaction we have added a methyl group to the nitrogen N_1 resulting in the model compound **T** depicted in Fig. 1. The frequencies calculated for **T** are also displayed in Table 1 and show, that the C_2=O_2 mode is red-shifted by about 30 cm^{-1}, whereas the other two modes are nearly unaffected. The red-shift is due to the formation of weak internal hydrogen bonds between the added methyl group and the C_2=O_2 group ([30] analyzes a related case). Correspondingly, in crystalline powder of thymine the two C=O bands are both red-shifted by about 40 cm^{-1} [11,14] which has

Table 1
Experimental and calculated vibrational frequencies of thymine in the 1600–2000 cm^{-1} region

Mode	BLYP[a]	Gas[b]	Ar[c]	B3LYP$_{un}$[d]	B3LYP$_{sc}$[e]	**T**, BLYP[f]
C_2=O_2	1765	1772	1767	1842	1774	1736
C_4=O_4	1707	1725	1711	1791	1722	1707
C_5=C_6	1645	1668	1668	1715	1673	1641

[a] Our calculation.
[b] IR spectrum of thymine in the gas phase [12].
[c] IR spectrum of thymine in an argon matrix [15].
[d] Unscaled DFT/B3LYP calculation [15].
[e] Scaled DFT/B3LYP calculation [15].
[f] Our calculation for 1-methyl-thymine (**T**).

been attributed to hydrogen bonding interactions in the crystal [15].

The same photolyase that repairs thymine dimerizations also cleaves uracil dimers (uracil lacks the methyl group at C_5, cf. Fig. 1). Therefore, we have also calculated the vibrational spectra of 1-methyl-uracil and, because of its spectroscopic properties, of 1-methyl-5-flouro-uracil. Omission of the methyl group at C_5 removes internal hydrogen bonds with O_4 and, as expected from the discussion above, causes a 19 cm^{-1} blue-shift of the $C_4{=}O_4$ stretch. The $C_2{=}O_2$ mode stays nearly unaffected but the $C_5{=}C_6$ mode is red-shifted to 1621 cm^{-1}. Replacing the proton at C_5 with a fluorine atom exclusively shifts that mode back up to 1648 cm^{-1} leaving the C=O modes unaffected.

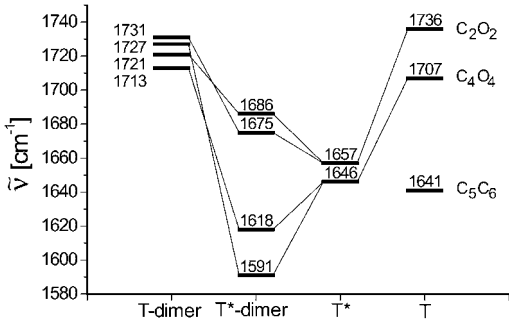

Fig. 2. Carbonyl stretch frequencies for 1-methyl-thymine models of proposed states in the cyclopyrimidine repair reaction: the neutral **T-dimer**, the anionic radical **T*-dimer**, the anionic and neutral monomers **T*** and **T**, respectively.

3.2. Dimer

The optimized geometry of **T-dimer** is shown in Fig. 1a. In excellent agreement with results of HF calculations [31,4] and with our own DFT results on thymine and uracil dimers we found for **T-dimer** a puckered, non-planar four-membered ring with a $C_5{-}C_5'{-}C_6'{-}C_6$ dihedral angle of 18°. In the crystal structure the uracil dimer exhibits a somewhat larger dihedral of 25° [32] which could be due to crystal packing. The puckered four-membered ring leads to a non-symmetric arrangement of the two pyrimidine rings. In **T-dimer** they are twisted against each other by a $C_4{-}C_5{-}C_5'{-}C_4'$ dihedral angle of −25°. Similar values were found for the other dimers investigated and closely match the value of −23° observed in the crystal structure of the uracil dimer [32]. In the dimers, the pyrimidine rings are non-planar. For instance, the $C_6{-}C_5{-}C_4{-}N_3$ dihedral angle was calculated at −18° for the **T-dimer**, the corresponding angle in the other ring was obtained at −14°. The crystal structure of the uracil dimer reveals similar values of −15° and of −16°, respectively, for these angles.

The non-symmetrical structures of the dimers cause differences among the carbonyl stretch frequencies in the two rings. Fig. 2 compares the level scheme obtained for **T-dimer** with that of **T**-monomer; also shown are the levels of the respective anions **T*-dimer** and **T*** discussed further below. For **T-dimer** the $C_2{=}O_2$ and $C_2'{=}O_2'$ bands are predicted at 1721 and 1731 cm^{-1}, respectively, and the $C_4{=}O_4$ and $C_4'{=}O_4'$ bands at 1727 and 1713 cm^{-1}, respectively. In the 1-methyl-uracil dimer, we find corresponding frequencies of 1720, 1730, 1741, and 1726 cm^{-1}. Compared to the thymine case, the $C_4{=}O_4$ and $C_4'{=}O_4'$ modes of the uracil dimer are blue-shifted by about 13 cm^{-1} due to the absence of internal hydrogen bonding.

In the 1-methyl-5f-uracil dimer the frequencies of all modes are blue-shifted ($C_2{=}O_2$ to 1725, $C_2'{=}O_2'$ to 1738, $C_4{=}O_4$ to 1753, and $C_4'{=}O_4'$ to 1732 cm^{-1}). This blue-shift is of interest, since in experiments in aqueous solution the C=O absorptions of thymine and uracil dimers are hidden below dominating water bands covering the 1600–1700 cm^{-1} range. In experiments with single strand DNA fluor uracil dodecamers, however, one infrared band at about 1740 cm^{-1} has been found [16] and has been assigned to the dimer. This high-frequency position agrees well with our results and, up to now, is the only observation of a pyrimidine dimer band in the high-frequency region.

Upon dimerization one generally finds a blue-shift of the $C_4{=}O_4$ stretching modes and a red-shift of the $C_2{=}O_2$ modes. This effect is readily understood by consideration of the resonance structures shown in Fig. 3. Dimerization is compatible with all resonance structures of the monomer except for structure III, in which the

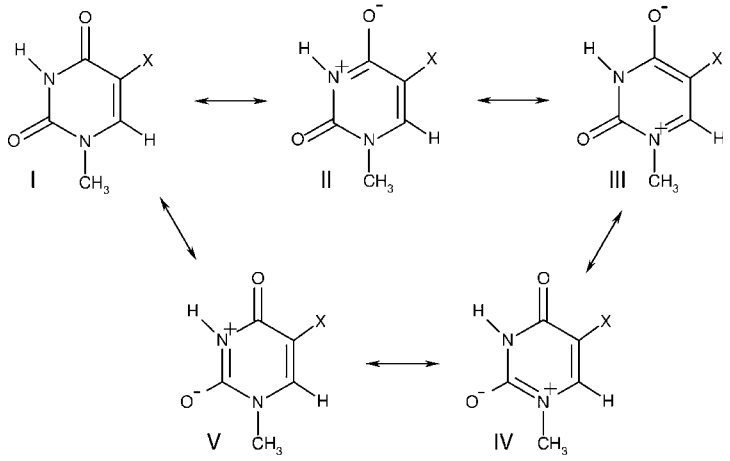

Fig. 3. The dominating resonance structures of neutral pyrimidines.

C_5═C_6 double bond required for dimerization is absent. Together with structure III a C_4—O_4 single bond and a C_2═O_2 double bond are lost upon dimerization which explains the strengthening of the former and the weakening of the latter bond.

3.3. Anionic structures

The radical anion **T*** of 1-methyl-thymine is found to be almost planar. A chair structure with C_6 and N_3 slightly out of the plane defined by the other ring atoms leads to a C_6—N_1—C_2—N_3 dihedral angle of 8.7°.

As exhibited by Fig. 2, reduction causes strong red-shifts of the C═O frequencies in the thymine monomers. Compared to **T**, the **T*** modes are red-shifted by about 70 cm^{-1} to 1657 (C_2═O_2) and 1646 cm^{-1} (C_4═O_4). The corresponding force constants are reduced from about 10.5 to 9.2 (C_2═O_2) and 8.5 mdyn/Å (C_4═O_4).

In contrast to previous quantum mechanical calculations on thymine dimers at the HF level [4,33], we did not find a stable structure for the 1-methyl-thymine dimer radical anion with both bonds, C_5—C_5' and C_6—C_6', conserved. Instead, starting from the structure of the neutral dimer, the C_5—C_5' bond was spontaneously broken after electron uptake leading to the intermediate structure **T*-dimer** shown in Fig. 1a.

After cleavage of the C_5—C_5' bond sterical constraints are lost and the two rings can change their relative orientation. Upon reduction the C_4—C_5—C_5'—C_4' dihedral angle grows from −24.4° in **T-dimer** to −38.7° in **T*-dimer**. Also, both rings are predicted to be non-planar with dihedral angles of −43.1° (C_2—N_1—C_6—C_5) and 22.9° (C_2'—N_1'—C_6'—C_5'). The previous HF calculations on reduced uracil dimers had predicted a much larger twist of about 67° between the rings and almost planar ring structures [4].

Fig. 2 demonstrates that the C═O modes are strongly red-shifted by reduction also in the dimers. As compared to **T-dimer**, in **T*-dimer** the C_2═O_2 and C_2'═O_2' modes are red-shifted by about 50 cm^{-1} to 1686 and 1675 cm^{-1}, respectively, and the coupled C_4═O_4 and C_4'═O_4' modes by nearly 100 cm^{-1} to 1618 and 1592 cm^{-1}, corresponding to symmetric and anti-symmetric combinations, respectively. In **T*-dimer** the force constants of the two C_2═O_2 bonds are approximately 9.7 mdyn/Å and those of the C_4═O_4 bonds are about 8.8 mdyn/Å. Thus, they are larger than the corresponding force constants in the **T***-monomer. Like in the comparison of neutral pyrimidines and their dimers presented further above, also these changes of force constants in the anionic case can be explained in terms of valence bond arguments. The increase of force constants upon dimerization

seems to contradict the red-shift of the $C_4{=}O_4$ mode displayed in Fig. 2. However, reduction strongly affects the coupling of the $C_4{=}O_4$ modes with other ring modes, i.e. the $C_4{=}O_4$ frequency does not solely depend on the value of the $C_4{=}O_4$ force constant.

4. Conclusion

We have calculated C=O and C=C frequencies of models for various states of the light-induced cyclopyrimidine repair reaction. First, our calculations on thymine have provided evidence that the applied BLYP/6-31G* method can accurately predict the frequencies of C=C and C=O modes of such compounds. We have compared the equilibrium structures of thymine and uracil dimers with experimental data and results of previous HF calculations obtaining good agreement. C=O stretching frequencies are found to show typical shifting patterns upon dimerization. Very high frequencies are predicted for the C=O modes of 1-methyl-5-flouro-uracil dimers matching a corresponding spectroscopic observation.

The carbonyl modes in radical anions of monomers and dimers show large red-shifts as compared to the neutral molecules. The resulting spectral positions strongly differ for the monomers and the dimers. Therefore, the intense C=O modes should be well suited for monitoring the repair reaction using time resolved FTIR techniques.

Our investigations predict that the pyrimidine C=O stretching modes are quite sensitive to hydrogen bonding. Accordingly, we expect frequencies observed in photolyase to be red-shifted by about 20–40 cm^{-1} as compared to our vacuum calculations.

Acknowledgements

The authors would like to thank A. Bacher, K. Gerwert, G. Richter, and E. Schleicher for communicating FTIR difference spectra of cyclopyrimidines prior to publication. This work was supported by the *Volkswagen Stiftung* (Project I/73 224) and by the Deutsche Forschungsgemeinschaft (SFB533/C3).

References

[1] S.-T. Kim, A. Sancar, Photochem. Photobiol. 57 (5) (1993) 895.
[2] A. Sancar, Biochemistry 33 (1994) 2.
[3] H.-W. Park, S.-T. Kim, A. Sancar, J. Deisenhofer, Science 268 (1995) 1866.
[4] A. Voityuk, N. Rösch, J. Phys. Chem. A 101 (1997) 8335.
[5] R. Brudler, H.J.M. de Groot, W.B.S. van Liemt, W.F. Steggerda, R. Esmeijer, P. Gast, A.J. Hoff, J. Lugtenburg, K. Gerwert, EMBO J. 13 (1994) 5523.
[6] J. Breton, J.-R. Burie, C. Boullais, G. Berger, E. Nabedryk, Biochemistry 33 (1994) 12405.
[7] K. Gerwert, F. Siebert, EMBO J. 5 (1986) 805.
[8] A. Maeda, Isr. J. Chem. 35 (1995) 387.
[9] K. Gerwert, Biol. Chem. 380 (1999) 931.
[10] C. Allin, K. Gerwert, Biochemistry 40 (2001) 3037.
[11] M. Aida, M. Kaneko, M. Dupuis, T. Ueda, K. Ushizawa, G. Ito, A. Kumakura, M. Tsuboi, Spectrochim. Acta A 53 (1997) 393.
[12] P. Colarusso, K. Zhang, B. Guo, P.F. Bernath, Chem. Phys. Lett. 269 (1997) 39.
[13] M. Tsuboi, M. Komatsu, J. Hoshi, E. Kawashima, T. Sekine, Y. Ishido, M.P. Russell, J.M. Benevides, G.J. Thomas Jr., J. Am. Chem. Soc. 119 (1997) 2025.
[14] S.L. Zhang, K.H. Michaelian, G.R. Loppnow, J. Phys. Chem. A 102 (1998) 461.
[15] K. Szczepaniak, M.M. Szczesniak, W.B. Person, J. Phys. Chem. A 104 (2000) 3852.
[16] E. Schleicher, G. Richter, unpublished data, 2000.
[17] E. Schleicher, B. Heßling, V. Illarionova, R. Feicht, A. Bacher, G. Richter, K. Gerwert, unpublished data, 2001.
[18] M. Nonella, P. Tavan, Chem. Phys. 199 (1995) 19.
[19] X. Zhou, S.J. Mole, R. Liu, Spectroscopy 12 (1996) 73.
[20] P. Tavan, Spectroscopy of biological molecules, in: J.C. Merlin, S. Turrell, J.P. Huvenne (Eds.), Sixth European Conference on the Spectroscopy of Biological Molecules 1995, Villeneuve d'Asc, France, Kluwer Academic Publishers, Dordrecht, 1995, p. 3.
[21] M. Eichinger, P. Tavan, J. Hutter, M. Parrinello, J. Chem. Phys. 110 (1999) 10452.
[22] A. Bacher, personal communication, 2001.
[23] M.J. Frisch et al., GAUSSIAN 98, Revision A.5, Gaussian, Inc., 1998, Pittsburgh, PA.
[24] C. Lee, W. Yang, R.G. Parr, Phys. Rev. B 37 (1988) 786.
[25] A.D. Becke, Phys. Rev. A 38 (1988) 3098.
[26] R. Ditchfield, W.J. Hehre, J.A. Pople, J. Chem. Phys. 54 (1971) 724.
[27] W.J. Hehre, R. Ditchfield, J.A. Pople, J. Chem. Phys. 56 (1972) 2257.
[28] M.M. Francl, W.J. Pietro, W.J. Hehre, J.S. Binkley, M.S. Gordon, D.J. DeFrees, J.A. Pople, J. Chem. Phys. 77 (1982) 3654.

[29] M.W. Schmidt, K.K. Baldridge, J.A. Boatz, S.T. Elbert, M.S. Gordon, J.H. Jensen, S. Koseki, N. Matsunage, K.A. Nguyen, S.J. Su, T.L. Windus, M. Dupuis, J.A. Montgomery, J. Comput. Chem. 14 (1993) 1347.
[30] M. Nonella, J. Phys. Chem. A 103 (35) (1999) 7069.
[31] M. Aida, M. Kaneko, M. Dupuis, Int. J. Quantum Chem. 57 (1996) 949.
[32] E. Adman, M.P. Gordon, E.H. Jensen, J. Am. Chem. Soc. Chem. Comm. (1968) 1019.
[33] A. Voityuk, M.-E. Michel-Beyerle, N. Rösch, J. Am. Chem. Soc. 57 (1996) 9750.

2.2 Schwingungsspektren isotopenmarkierter Modellstrukturen

Die im vorangegangenen Abschnitt 2.1 beschriebenen Rechnungen geben keinen Hinweis darüber, wie man aus experimentellen IR Spektren den Aufenthaltsort des Überschußelektrons, das von dem FADH Kofaktor der Photolyase nach der Lichtabsorption auf die DNS übertragen wird, bestimmen kann, sobald die CPD Spaltung abgelaufen ist. Um diesen Sachverhalt zu veranschaulichen, betrachten wir die Basensequenz

$$5'\ldots \mathrm{AT{=}TA}\ldots 3', \tag{2.1}$$

bei der die beiden Thyminbasen (T) ein CPD bilden, und bei der 5' bzw. 3' die verschiedenen Enden des DNS Strangs bezeichnen [5]. Nach dem Elektrontransfer von dem FADH Kofaktor der Photolyase auf das CPD spaltet sich das Dimer in zwei Monomere. Dabei ist nicht bekannt, ob das Überschußelektron nach der Spaltung von der 3'-seitigen oder der 5'-seitigen Pyrimidinbase übernommen wird, d.h. ob sich das Intermediat

$$5'\ldots \mathrm{AT(T^-)A}\ldots 3' \tag{2.2}$$

oder

$$5'\ldots \mathrm{A(T^-)TA}\ldots 3' \tag{2.3}$$

bildet. Denkbar ist ferner, dass das überschüssige Elektron nach der Spaltung zwischen den Basen springt, bevor es an den FADH Kofaktor der Photolyase zurückübertragen wird.[2] Bei Vernachlässigung von Polarisationseffekten sind die IR Spektren der Intermediate (2.2) und (2.3) durch die Superposition der Spektren der einzelnen DNS Basen gegeben und unterscheiden sich daher nicht. Um die beiden Intermediate (2.2) und (2.3) in IR Spektren deutlich voneinander unterscheidbar zu machen, stellen chemische Markierungen (z.B. der Austausch einer Thyminbase durch ein Cytosin, welches ein CPD der Form C=T bilden kann [39]) oder Isotopenmarkierungen ein geeignetes Mittel dar.

Um für zukünftige Isotopenmarkierungen einen Leitfaden zu erstellen, habe ich die IR Spektren der in Kap. 2.1 skizzierten Modellstrukturen berechnet, wobei die vier Ringkohlenstoffe, der Methylkohlenstoff an der C5 Position und die Stickstoffe durch ^{13}C bzw. ^{15}N Isotope ausgetauscht wurden. Da die Frequenzen der isotopenmarkierten Strukturen aus den Atommassen und derjenigen masseunabhängigen Hessematrix folgen [67, 68], die ich bereits für die entsprechenden unmarkierten Strukturen berechnet hatte, waren dafür keine weiteren DFT Rechnungen notwendig.

Abb. 2.1 zeigt die Schwingungsfrequenzen des isotopenmarkierten (T*) und des unmarkierten (T) 1-Methyl-Thymin Monomers. Ferner sind auch die Schwingungsfrequenzen beider Moleküle im anionischen ($^-$) Zustand angegeben. Man erkennt

[2] Zu Beobachtungen von Wanderungen von Überschußelektronen in der DNS siehe z.B. Lit. [102, 103].

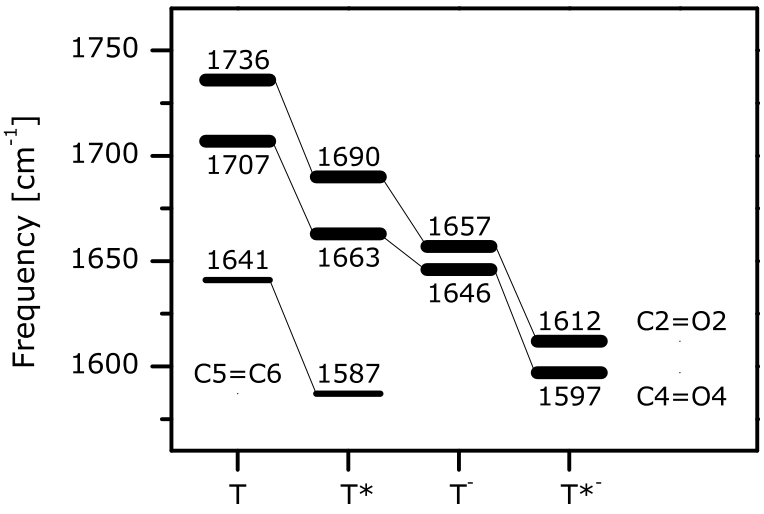

Abbildung 2.1: Berechneter Effekt der Isotopenmarkierung und des Radikalelektrons auf die IR Spektren des 1-Methyl-Thymin Monomers: a) neutral, unmarkiert (T), b) neutral, isotopenmarkiert (T*), c) anionisch, unmarkiert (T^-), d) anionisch, isotopenmarkiert (T^{*-}).

durch Vergleich der Schwingungsfrequenzen von T und T*, dass die Isotopenmarkierung die "reduzierte" Masse der Carbonylmoden erhöht und dadurch deren Schwingungsfrequenzen um etwa 45 cm^{-1} ins Rote verschiebt. Die Frequenzen des T^- Anions sind um etwa 70 cm^{-1} gegenüber denen des neutralen T rotverschoben. Ferner erkennt man an der Differenz von etwa 115 cm^{-1} zwischen den Frequenzen von T^{*-} und denen des T, dass die beiden rotverschiebenden Effekte der Isotopenmarkierung und des Radikalelektrons in guter Näherung additiv sind.

Auf der Basis der in Abb. 2.1 gezeigten Spektren habe ich mit Adelbert Bacher (TU München), Gerald Richter (inzwischen University Exeter, GB) und Paul Tavan in einer gemeinsamen Diskussion die Idee entwickelt, die Basensequenz

$$5'\ldots \mathrm{AT{=}T^*A}\ldots 3' \tag{2.4}$$

zu betrachten, die sich von der oben betrachteten Sequenz (2.1) nur durch die isotopenmarkierte Base T* unterscheidet. Die Spaltung dieses CPD läuft über die zu (2.2) und (2.3) analogen Intermediate

$$5'\ldots \mathrm{AT(T^{*-})A}\ldots 3' \tag{2.5}$$

oder

$$5'\ldots \mathrm{A(T^-)T^*A}\ldots 3'. \tag{2.6}$$

Wie man aus den in Abb. 2.1 abgebildeten Spektren ableiten kann, unterscheiden sie diese beiden Intermediate eindeutig. Daher sollte ein entsprechendes Experiment geeignet sein, den Ort des Überschußelektrons nach der CPD Spaltung zu bestimmen.

3 Berechnung von Infrarotspektren einzelner Moleküle in Lösungsmitteln

Auf dem Weg zu einer genauen Berechnung von MIR Spektren von Molekülen in kondensierter Phase aus DFT/MM-MD Simulationen beschreibt das folgende Kapitel verschiedene Methoden zur Berechnung von MIR Spektren einzelner Moleküle in polaren Lösungsmitteln.

3.1 Infrarotspektren aus Molekulardynamiksimulationen

Der nun folgende Abschnitt ist ein Abdruck des Artikels[1]

> Matthias Schmitz and Paul Tavan: „Vibrational spectra from atomic fluctuations in dynamics simulations: I. Theory, limitations, and a sample application“, J. Chem. Phys. **121**, 12233-12246 (2004).

Hier wird die MIR Absorption einzelner Moleküle in polaren Lösungsmitteln aus der linearen quantenmechanischen Störungstheorie [98] hergeleitet, die PMA Methode [100, 101] zur Berechnung von MIR Spektren ausführlich auf ihre Gültigkeit überprüft und ein neues Rechenverfahren hergeleitet, welches das MIR Spektrum eines Moleküls aus “generalisierten Virialfrequenzen” (GV) berechnet.

[1] Mit freundlicher Genehmigung des American Institute of Physics.

Vibrational spectra from atomic fluctuations in dynamics simulations. I. Theory, limitations, and a sample application

Matthias Schmitz and Paul Tavan
Theoretische Biophysik, Lehrstuhl für Biomolekulare Optik, Ludwig-Maximilians-Universität München, Oettingenstrasse 67, 80538 München, Germany

(Received 25 May 2004; accepted 29 September 2004)

Hybrid molecular dynamics (MD) simulations, which combine density functional theory (DFT) descriptions of a molecule with a molecular mechanics (MM) modeling of its solvent environment, have opened the way towards accurate computations of solvation effects in the vibrational spectra of molecules. Recently, Wheeler *et al.* [ChemPhysChem **4**, 382 (2002)] have suggested to compute these spectra from DFT/MM-MD trajectories by diagonalizing the covariance matrix of atomic fluctuations. This so-called principal mode analysis (PMA) allegedly can replace the well-established approaches, which are based on Fourier transform methods or on conventional normal mode analyses. By scrutinizing and revising the PMA approach we identify five conditions, which must be guaranteed if PMA is supposed to render exact vibrational frequencies. Besides specific choices of (a) coordinates and (b) coordinate systems, these conditions cover (c) a harmonic intramolecular potential, (d) a complete thermal equilibrium within the molecule, and (e) a molecular Hamiltonian independent of time. However, the PMA conditions [(c)–(d)] and [(c)–(e)] are generally violated in gas phase DFT-MD and liquid phase DFT/MM-MD trajectories, respectively. Based on a series of simple analytical model calculations and on the analysis of MD trajectories calculated for the formaldehyde molecule in the gas phase (DFT) and in liquid water (DFT/MM) we show that in both phases the violation of condition (d) can cause huge errors in PMA frequency computations, whereas the inevitable violations of conditions (c) and (e), the latter being generic to the liquid phase, imply systematic and sizable underestimates of the vibrational frequencies by PMA. We demonstrate that the huge errors, which are caused by an incomplete thermal equilibrium violating (d), can be avoided if one introduces mode-specific temperatures T_j and calculates the frequencies from a ''generalized virial'' (GV) expression instead from PMA. Concerning ways to additionally remove the remaining errors, which GV still shares with PMA, we refer to Paper II of this work [M. Schmitz and P. Tavan, J. Chem. Phys. **121**, 12247 (2004)].
© *2004 American Institute of Physics.* [DOI: 10.1063/1.1822914]

I. INTRODUCTION

Spectroscopy in the midinfrared (MIR) spectral region, which covers frequencies from about 600 to about 4000 cm^{-1}, is an important tool in chemical analysis and in biomolecular research because the MIR absorption of molecules is experimentally well accessible and contains information on the chemical state of the molecules and on the structure and dynamics of their condensed phase environment.[1–9]

For a decoding of the MIR spectra in terms of molecular structure, accurate descriptions of intramolecular force fields and of their perturbation by the environment are required.[1,4–7,10] In principle such descriptions have become accessible due to the development of hybrid methods, which combine highly precise density functional theory (DFT) for a solute molecule with a microscopic molecular mechanics (MM) modeling of its environment.[11,12] This paper and Paper II of this work[13] aim at comparing and developing computational methods by which one can accurately derive MIR spectra of solute molecules from DFT/MM simulations.

The infrared (IR) absorption spectrum of a thermal ensemble of molecules is given by[14]

$$\alpha(\omega)=\frac{c_0\omega}{\hbar}(1-e^{-\hbar\omega/k_BT})\sum_{F,I}\rho_I|\langle F|\vec{e}_E\cdot\vec{D}|I\rangle|^2\delta(\omega-\omega_{FI}), \tag{1}$$

where the velocity of light c and the refractive index n are collected into $c_0=4\pi^2/(cn)$, $|F\rangle$ and $|I\rangle$ are eigenstates of the Hamiltonian H, $\vec{D}$ is the dipole operator of the system, $\vec{e}_E$ the unit vector in the direction of the electric field $\vec{E}$, and ρ_I is the population of the state $|I\rangle$. The discrete absorption frequencies $\omega_{FI}=(E_F-E_I)/\hbar$ derive from the energy differences between the states after ($|F\rangle$) and before ($|I\rangle$) absorption. Due to the large number of mutually interacting molecules, many of which are excited in a common experimental setup, and because of finite spectral resolution, the discrete spectrum (1) appears to be composed of broad absorption bands. If these bands are well separated from each other, they can be characterized by central frequencies and linewidths.

By the equivalent expression

 © 2004 American Institute of Physics

$$\alpha(\omega)=\frac{c_0\omega}{\hbar}(1-e^{-\hbar\omega/k_BT})\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega t}\langle[\vec{e}_E\cdot\vec{D}(t')]\times[\vec{e}_E\cdot\vec{D}(t'+t)]\rangle dt, \quad (2)$$

the IR spectrum is the Fourier transform of the time correlation function (FTTCF) of the dipole operator of the system. Here, the dipole operator has been written as a time-dependent observable in the Heisenberg representation and the brackets $\langle\cdots\rangle$ denote the thermal ensemble of initial states $|I\rangle$ (see Ref. 14).

For a system of M molecules isolated from each other in the gas phase, the Hamiltonian can be decomposed into the sum

$$H=\sum_{m=1}^{M}H_m \quad (3)$$

of molecular Hamiltonians H_m each describing one of the M molecules. After averaging over all molecular orientations, whose distribution is assumed to be isotropic, the IR spectrum (1) is then made up of single molecule contributions,

$$\alpha_m(\omega)=\frac{c_0\omega}{3\hbar}(1-e^{-\hbar\omega/k_BT})\times\sum_{f_m,i_m}\rho_{i_m}|\langle f_m|\vec{d}_m|i_m\rangle|^2\delta(\omega-\omega_{f_mi_m}) \quad (4)$$

or, equivalently,

$$\alpha_m(\omega)=\frac{c_0\omega}{3\hbar}(1-e^{-\hbar\omega/k_BT})\times\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega t}\langle\vec{d}_m(t')\vec{d}_m(t'+t)\rangle dt. \quad (5)$$

Here $\vec{d}_m$ is the dipole operator of the molecule m, $|i_m\rangle$ and $|f_m\rangle$ are the eigenstates of the molecular Hamiltonian H_m, and $\omega_{f_mi_m}=(E_{m,f}-E_{m,i})/\hbar$ are the associated absorption frequencies. Note that the Hamiltonians H_m cover the coupled vibrational and rotational motions of the molecules.[15,16] By the choice of suitable coordinate systems these motions can be approximately decoupled, and perturbation theory can be applied to largely correct the corresponding errors.[17] Due to the accessibility of accurate DFT methods the gas phase spectra (4) can be predicted at a high precision even for relatively large molecules.[18]

The computation of the condensed phase spectra (1) is much more difficult, because the interactions between the molecules in the system generally prevent the separation (3) of the Hamiltonian into individual molecular contributions and, therefore, correspondingly decoupled descriptions (4) and (5). For molecules in polar solvents (e.g., in water), however, the molecular interactions exhibit a special frequency dependence, which can be exploited for deriving descriptions in a certain spectral region. In the far IR ($\omega<600$ cm^{-1}) a decoupling is impossible because here the vibrational modes and librational motions of the solute molecules couple to librations and conformational changes of their solvent cages. In contrast, in the MIR ($\omega>600$ cm^{-1}) couplings are induced either by dipole–dipole interactions among vibrational dipoles in close resonance or by dipole–induced-dipole interactions caused by the electronic polarization of molecules near a vibrational dipole. It is the simplicity of these coupling effects, which opens the way towards a description of MIR spectra.

Here we will address the specific question as to how one can calculate the MIR spectra of molecules, which are dilutely soluted in polar solvents. The focus on the MIR spectral region allows us to exclude the complicated low-frequency couplings from our analysis. Furthermore, by the use of hybrid simulations, in which a single solute molecule is described by a highly accurate DFT Hamiltonian and the solvent by a rigid, nonpolarizable MM model, we avoid the complicating resonant couplings between the vibrational modes of the molecules: First, the restriction to a single solute molecule in the simulation system excludes couplings between molecules of the same species. Second, by the choice of a rigid MM model the vibrations of the solvent molecules are suppressed and, thus, couplings to the vibrations of the solute cannot arise. Third, by the choice of a nonpolarizable solvent model one avoids (and, thus, neglects) correlations between the vibrational dipole of the solute molecule and the dipole moments induced in the solvent by electronic polarization.

For strongly polarizing ionic solutes, the neglect of the solvent polarizability can cause an underestimation of the solvent-induced frequency shifts.[10] For neutral dipolar solutes, however, this neglect has been shown to scarcely affect calculated IR intensities and frequencies.[1,19] Apart from the neglect of the solvent polarizability the virtual reality generated by such a DFT/MM hybrid simulation of the molecular dynamics (MD) corresponds to an experimental situation, in which the dilution is large enough to prevent dipole–dipole couplings among the solute molecules and in which the absorption frequencies of the solvent molecules exhibit sufficiently large spectral distances from those of the solute as to prevent resonant couplings.

All methods suggested so far for the computation of MIR spectra by DFT/MM-MD simulations rely on the approximations and assumptions sketched above.[1,2,19–21] However, the lively discussion concerning the validity of the various computational approaches, which is documented in the quoted papers, suffers from the fact that these assumptions were never explicitly stated. To remove some of the resulting confusion and to enable a deeper insight into the merits and deficiencies of the various suggestions, we first want to look more closely at these methods.

The so-called instantaneous normal mode analysis (INMA), which is described by Nonella *et al.* in Ref. 2, is implicitly based on a separation of the Hamiltonian

$$H=H_m+H_s+H_{ms} \quad (6)$$

into the Hamiltonians H_m and H_s, which describe the solute molecule and the solvent, respectively, and the Hamiltonian H_{ms}, which contains the interactions between the two. Because the thermally excited motions of the rigid solvent molecules are much slower than the MIR vibrations of the solute, one can assume that the vibrational eigenstates $|i_m\rangle$ of

the solute *instantaneously* adjust to the solvent configurations. In this approximation the eigenstates associated with vibrational excitations of the solute molecule are determined by the Schrödinger equation

$$(H_m + H_{ms})|i_m\rangle = E|i_m\rangle \tag{7}$$

for the solute molecule at a fixed solvent environment configuration. Note that this approximation is analogous to the Born–Oppenheimer approximation, according to which the electronic eigenstates are considered to instantaneously adjust to the nuclear configurations.

Furthermore, for sufficiently large solute molecules in viscous solvents one can approximately separate the vibrational from the translational and rotational contributions to H_m, because the solvent cages strongly suppress the latter motions. Then the initial ($|i_m\rangle$) and final ($|f_m\rangle$) states of an MIR absorption process reduce to the eigenstates of the vibrational contributions to $H_m + H_{ms}$. Finally, because of the fact that at room temperature all MIR modes of a solute molecule are in the ground state and because of the small transition moments of multiple or combined excitations, one can restrict the discussion to the fundamental absorption frequencies $\omega_{f_m i_m}$ associated with excitations from the ground state $|i_m\rangle$ to the first vibrational level $|f_m\rangle$ of the MIR modes.

Based on these ideas, the INMA protocol[2] proceeds as follows: First, an MD simulation serves to generate an ensemble of solvent cages. These cages are subsequently kept fixed and the vibrational spectra of the embedded solute molecule are calculated from the DFT/MM force field by conventional techniques of normal mode analysis (NMA), in which the geometry optimizations and the computations of the Hessians are restricted to the coordinates of the solute. Note that the thus applied harmonic approximation is solely a matter of computational efficiency. If one can afford an increased computational effort, one can also include anharmonic quantum corrections.[17,18] The calculated ensemble of frequencies then allows us to reconstruct the MIR spectrum, which, with the noted approximations and the assumption of an isotropic distribution of solute orientations, has the form (4).

Note that the calculated absorption frequencies $\omega_{f_m i_m}$ of the solute are different in each frozen solvent cage, because the Schrödinger equation (7) contains the interactions H_{ms} of the solute with its respective solvation shell, by which the solute becomes specifically and individually polarized. Therefore, using INMA one can study in great detail, how the vibrational spectrum of a solute molecule is steered by the structure of its environment (cf. Ref. 1).

INMA requires a sizable computational effort for the energy minimization and for the subsequent frequency calculation of the solute in each solvent cage implying that one may have to resort to the harmonic approximation and that, nevertheless, the statistics may be limited.[2] Furthermore, due to the neglect of the coupling of the kinetic energy operator in H_s to the vibrational eigenstates $|i_m\rangle$ in the Schrödinger equation (7), the solutes vibrations are effectively decoupled from the dynamics of its environment. Therefore, INMA does not capture certain dynamical effects affecting the linewidths (such as, e.g., motional narrowing[22–24]). We will address this issue in Paper II of this work.[13]

However, concerning calculated linewidths we note already here that the INMA prescription[2] deviates from other suggestions[25–27] for the computation of ''instantaneous normal modes'' (INM). According to these INM suggestions and in contrast to the INMA protocol, the vibrational frequencies of a molecule are calculated from its Hessian without a preceding energy minimization. Apart from increasing the calculated linewidths,[13] the omitted energy minimization can lead to imaginary values for some of the vibrational frequencies. Although the imaginary frequencies entail unphysical divergences in the computation of time correlation functions,[25] the resulting INM density of states (DOS) is assumed to provide a suitable approximation to the true DOS.[25,26,28] As an *ad hoc* remedy, the imaginary frequencies are either projected into the real space[29] by multiplication with $-i$ or are simply neglected.[25,30]

Considering the FTTCF approach (2) for the calculation of MIR spectra from DFT/MM-MD simulations, recall that our construction of the DFT/MM setup, in which a single solute molecule is described by DFT and the solvent by a rigid, unpolarizable MM model, excludes resonant couplings between the MIR vibrational dipole of the solute and those of the solvent while neglecting the weak effects of dipole–induced-dipole interactions. For a sufficiently extended DFT/MM trajectory the correlations between the slowly varying permanent dipoles of the solvent molecules and the MIR vibrational dipole of the solute will average to zero. Following the arguments in Ref. 14 the MIR absorption of an isotropically distributed solute therefore decomposes into the form (5). Here, the dipole operators are replaced by the classical dipole moments, if one applies the classical description to the nuclear motion as is common in MD simulations.

In contrast to INMA, FTTCF is not based on the decoupled description (7) of the solute and, therefore, accounts for the dynamical effects of motional narrowing. Furthermore, FTTCF covers those parts of the anharmonic corrections, which are sampled by a classical MD trajectory. On the other hand, the explicit time dependence of the frequencies is not accessible in the FTTCF representation (5), because the time correlation function represents an average over the complete DFT/MM-MD trajectory. Furthermore, the classical ensemble generated by a DFT/MM-MD trajectory, from which the dipole correlation function in Eq. (5) is calculated, does not coincide with the ensemble of an experimental probe in which nuclear motion proceeds quantum mechanically. For a correction, various frequency dependent factors have been suggested,[31,32] whose validity can be checked by theoretical considerations[33] or by comparison of FTTCF with INMA results[13,29] or experimental data.[29,34]

In a recent paper, Wheeler *et al.* have proposed to calculate the vibrational spectra from classical MD trajectories by a so-called ''principle mode analysis'' (PMA),[20] which allegedly represents a simple and efficient alternative to the established methods sketched above. Similar to the quasiharmonic analysis proposed by Brooks *et al.*[35] and to the more recent covariance analysis of atomic velocities proposed by Strachan,[36] these authors suggested to calculate the covari-

ance matrix σ of the displacements $\vec{x}$ of the atoms.[20] Assuming a harmonic potential $U(\vec{x})=\vec{x}F_{\text{eff}}\vec{x}$, where F_{eff} is the Hessian matrix, and a Boltzmann-like distribution $p(\vec{x})\sim\exp(-\vec{x}F_{\text{eff}}\vec{x}/2k_BT)$ [cf. Eq. (3) in Ref. 20], they claimed that the Hessian can be identified by $F_{\text{eff}}=k_BT\sigma^{-1}$ with the inverse of σ multiplied by the Boltzmann factor k_B and the temperature T [cf. Eq. (4) in Ref. 20]. Correspondingly, they proposed to calculate the vibrational normal modes and frequencies by diagonalization of σ^{-1}. Furthermore, they claimed that this PMA can account for nonharmonicities in the potential energy.[20,21]

Due to its simplicity the PMA approach[20] appears to be promising for the construction of vibrational spectra from MD trajectories. Therefore we want to check in this paper under which circumstances the PMA method can be used to compute MIR band positions at an accuracy, which is comparable to INMA or to FTTCF. By carefully revising the PMA scenario we first identify its basic assumptions (cf. Sec. II). In Sec. III we subsequently explain the differences of the revised PMA approach to the original suggestion by Wheeler *et al.* Furthermore, we prove by simple analytical model calculations, as to why PMA must fail in scenarios, in which one or more of its basic assumptions are violated. All these scenarios are relevant for the computation of vibrational spectra from DFT/MM trajectories by PMA. In this context we derive a related method that avoids one of the key PMA defects. To exemplify our conclusions, we apply this method and the original PMA prescription in Sec. IV to the DFT and DFT/MM trajectories of a formaldehyde molecule in the gas phase and in aqueous solution, respectively.

This test case is of interest because formaldehyde contains a carbonyl group. As is well known, the vibrations of carbonyl groups are strongly modified by external electric fields. Taking advantage of this effect, carbonyl vibrations were spectroscopically studied in many biologically relevant molecules to characterize, e.g., biochemical reactions catalyzed by enzymes and structural properties of peptides and proteins.[3–6,37]

II. BASIC THEORY

We start by deriving the tacit assumptions, on which any attempt to calculate the vibrational spectrum of a molecule by PMA from a MD trajectory is based. First we introduce our notation by briefly sketching the standard NMA within the framework of classical mechanics.

A. NMA of an isolated molecule at rest

Consider an isolated molecule consisting of N atoms in a nonlinear arrangement, which does not undergo any translational or rotational motion. Collecting the Cartesian coordinates $\vec{r}_i$ of the atoms into a vector $\vec{r}=(\vec{r}_1,\ldots,\vec{r}_N)^T$, the equilibrium configuration of the molecule is given by $\vec{r}^0$. Let us choose a frame of reference, whose origin is located at the center of mass $\vec{X}^0=\Sigma_{i=1}^N m_i\vec{r}_i^0$ of this equilibrium configuration. Small vibrations around $\vec{r}^0$ are conveniently described by the mass-weighted displacements

$$\vec{y}=M^{1/2}\vec{x}, \tag{8}$$

where $\vec{x}:=\vec{r}-\vec{r}^0$ describes the set of Cartesian displacements and M is the mass matrix, which is diagonal and in which the atomic masses m_i occupy the diagonal.

Applying the harmonic approximation to the potential energy $U_0(\vec{y})$ of the molecule, the molecular Hamiltonian is given by

$$H_0(\vec{y},\dot{\vec{y}})=\tfrac{1}{2}\dot{\vec{y}}^2+\tfrac{1}{2}\vec{y}^TF_0^{(M)}\vec{y}, \tag{9}$$

where $F_0^{(M)}=M^{-1/2}F_0M^{-1/2}$ is the mass-weighted Hessian F_0, which is the matrix of second derivatives of $U_0(\vec{y})$ at $\vec{y}=0$. The eigenvalues ω_j^2 of $F_0^{(M)}$ are the vibrational frequencies and the associated eigenvectors $\vec{c}_j$ yield the normal modes $\vec{n}_j$ by $\vec{n}_j=M^{-1/2}\vec{c}_j$. We label the six normal modes, which correspond to the translational or rotational motions, by indices $j\in\{1,\ldots,6\}$. They are characterized by the frequency zero. Collecting the eigenvectors into the matrix $C:=(\vec{c}_1,\ldots,\vec{c}_{3N})$, the transformation of a mass-weighted displacement vector $\vec{y}$ into normal coordinates $\vec{a}$ is given by

$$\vec{a}=C^T\vec{y}. \tag{10}$$

In normal coordinates, the Hamiltonian is then

$$H_0(\vec{a},\dot{\vec{a}})=\frac{1}{2}\sum_{j=7}^{3N}(\dot{a}_j^2+\omega_j^2a_j^2). \tag{11}$$

As is well known, the absorption (4) of a corresponding quantum mechanical harmonic oscillator is δ peaked at the classical vibrational frequencies ω_j. The associated absorption strengths

$$I_j:=\frac{2\omega_j}{\hbar}|\langle f|\vec{d}|i\rangle|^2 \tag{12}$$

are given by the derivatives of the molecular dipole moment $\vec{d}$ with respect to the normal coordinates a_j through

$$I_j=\left\langle\frac{\partial\vec{d}}{\partial a_j}\right\rangle^2. \tag{13}$$

These dipole derivatives can be calculated from a classical trajectory as follows: Given an MD time series of normal coordinates $\vec{a}(t)$ and of corresponding dipole moments $\vec{d}(t)$ sampled at an integration time step size Δt, the increment $\Delta\vec{d}(t)=\vec{d}(t+\Delta t)-\vec{d}(t)$ can be written as

$$\Delta\vec{d}(t)=\sum_j\frac{\partial\vec{d}}{\partial a_j}\Delta a_j(t)+O([\Delta a_j(t)]^2), \tag{14}$$

where $\Delta a_j(t)=a_j(t+\Delta t)-a_j(t)$. Neglecting terms of the order $O([\Delta a_j(t)]^2)$ and assuming that the expectation values

$$\left\langle\frac{\Delta a_i(t)}{\Delta a_j(t)}\right\rangle=\delta_{ij} \tag{15}$$

vanish for $i\neq j$ in a sufficiently long trajectory, Eq. (13) can be calculated by

$$I_j=\left\langle\left\langle\frac{\Delta\vec{d}(t)}{\Delta a_j(t)}\right\rangle\right\rangle^2. \tag{16}$$

B. Frame of reference

A molecule in solution will exhibit slow translations and rotations in addition to its vibrational motions. Therefore, the above assumption of a molecule at rest is not met. The motion of the molecule entails the following complications: First, the normal modes $\vec{n}_j$ are moving with the rotating molecule and, already for this reason, become time dependent in the Cartesian coordinates of the laboratory frame of reference. Second, the vibrations are coupled to the rotations through the kinetic energy term in the Hamiltonian.[15]

However, it is well known that the kinetic energy of the molecule can be separated into a rigid body term, a purely vibrational term, and a term coupling the vibrational to the rotational motion, if the coordinates are transformed into a particular molecular frame of reference.[15] The origin of this frame is defined by the condition that the center of mass

$$\vec{X}(\vec{y}) := \frac{1}{m}\sum_{i=1}^{N} \sqrt{m_i}\vec{y}_i = 0 \tag{17}$$

vanishes, where $m = \Sigma_{i=1}^{N} m_i$ denotes the total mass. Furthermore, the orientations of the axes have to obey the condition

$$\vec{\Phi}(\vec{y}) := \sum_{i=1}^{N} \vec{y}_i^{\,0} \times \vec{y}_i = 0, \tag{18}$$

where $\vec{y}^0 = M^{1/2}\vec{r}^0$ is the mass-weighted equilibrium configuration $\vec{r}^0$.[15] Taking the time derivatives of Eqs. (17) and (18) one can observe in the limit of small displacements $\vec{y}_i$ that the thus defined molecular frame is characterized by vanishing total linear and angular momenta $\vec{P} = m\dot{\vec{X}}$ and $\vec{L} = \dot{\vec{\Phi}}$, respectively.[15,16,38]

Because in viscous solvents the rotational motion of a sufficiently large solute molecule is strongly suppressed by friction, the vibration–rotation coupling can be neglected. The rigid body rotation can be taken into account by employing the molecular frame of reference specified above. Therefore, by using this frame one can describe the vibrational motion in the same way as if the molecule were at rest.[15] If not stated otherwise, we will use this molecular frame throughout this paper for the atomic displacement coordinates $\vec{y}_i$ of the molecule, whose vibrations are to be analyzed.

C. Principal mode analysis (PMA)

Having fixed the frame of reference we now want to explore, how one can obtain a vibrational spectrum from an MD trajectory of a molecule in a polar solvent using PMA. For simplification, we will first assume a purely harmonic potential and will neglect any influence of the environment on the potential energy of the molecule, i.e., we will use the Hamiltonian H_0 given by Eq. (9). Nevertheless, we will assume that there is some energy transfer between the environment and the molecule such that the molecule is in thermal equilibrium with its environment at a given temperature T.

Under these conditions we can assume a Boltzmann-like distribution given by

$$p(\vec{y},\dot{\vec{y}}) = \frac{1}{Z'}\exp[-H_0(\vec{y},\dot{\vec{y}})/k_BT] \times \delta[\vec{X}(\vec{y})]\delta[\vec{\Phi}(\vec{y})]\delta[\dot{\vec{P}}(\dot{\vec{y}})]\delta[\dot{\vec{L}}(\dot{\vec{y}})]. \tag{19}$$

Here, Z' is the partition function normalizing the distribution and the δ functions account for the restrictions imposed to the coordinates by the conditions (17) and (18) defining the molecular frame of reference. These δ functions render $p(\vec{y},\dot{\vec{y}})$ infinitely narrow in the directions of the translational and rotational motions. Integration over the mass-weighted velocities $\dot{\vec{y}}$ yields

$$p(\vec{y}) = \frac{1}{Z}\exp\left(-\frac{\vec{y}^T F_0^{(M)}\vec{y}}{2k_BT}\right)\delta[\vec{X}(\vec{y})]\delta[\vec{\Phi}(\vec{y})], \tag{20}$$

which can be expressed in terms of the normal coordinates (10) by

$$p(\vec{a}) = \frac{1}{Z}\exp\left(-\sum_{j=7}^{3N}\frac{\omega_j^2 a_j^2}{2k_BT}\right)\prod_{j=1}^{6}\delta(a_j). \tag{21}$$

The covariance matrix $\sigma^{(a)} = \langle \vec{a}\cdot\vec{a}^T\rangle - \langle\vec{a}\rangle\cdot\langle\vec{a}^T\rangle$ of this distribution is obviously diagonal. Here, the dot "$\cdot$" denotes the dyadic product of two vectors. The eigenvectors of $\sigma^{(a)}$ are the unit vectors $\vec{e}_j$ and its eigenvalues $\lambda_j^{(a)}$ are

$$\lambda_j^{(a)} = \sigma_{jj}^{(a)} = \begin{cases} 0 & \text{for} \quad j \in \{1,\ldots,6\}, \\ k_BT/\omega_j^2 & \text{for} \quad j \in \{7,\ldots,3N\}. \end{cases} \tag{22}$$

By transforming these results back to the mass-weighted displacement coordinates $\vec{y}$, one immediately finds that the covariance matrix

$$\sigma^{(M)} = \langle \vec{y}\cdot\vec{y}^T\rangle - \langle\vec{y}\rangle\cdot\langle\vec{y}^T\rangle = C\sigma^{(a)}C^T \tag{23}$$

of the distribution (20) has the same eigenvalues $\lambda_j = \lambda_j^{(a)}$ as $\sigma^{(a)}$ and that the eigenvectors $\vec{d}_j = C\vec{e}_j$ are identical to the eigenvectors $\vec{c}_j$ of the mass-weighted Hessian matrix. Hence, a principal component analysis of the distribution $p(\vec{y})$ yields the mass-weighted normal modes $\vec{c}_j$ as the principal axes.

According to Eq. (22) the vibrational frequency of a vibrational mode j follows by $\omega_j^2 = k_BT/\lambda_j$ from the temperature T and from the eigenvalue λ_j of the covariance matrix $\sigma^{(M)}$. Since λ_j measures the variance $\mathcal{V}(a_j)$ of the normal coordinate a_j at the given temperature, we can write

$$\omega_j^2 = k_BT/\mathcal{V}(a_j), \quad j \in \{7,\ldots,3N\}. \tag{24}$$

Further below this relation will be important in our test on the applicability of PMA.

As a result we may state that upon using (a) mass-weighted displacement coordinates, (b) the specified molecular frame of reference, (c) the harmonic approximation, assuming (d) thermal equilibrium for all modes, and (e) a time-independent Hamiltonian PMA appears to be a valid approach to the vibrational analysis of molecules from MD trajectories. In Sec. III we will check whether PMA retains its validity in scenarios, in which one or more of the conditions (a)–(e) are violated.

III. VALIDITY AND LIMITATIONS OF PMA

By comparing the above derivation of the PMA with the original proposal by Wheeler *et al.*,[20] we will first discuss the relevance of (a) mass-weighted coordinates and (b) the particular molecular frame of reference. Subsequently, we consider scenarios in which conditions (c)–(e) are violated.

A. Comparison with Ref. 20

In contrast to Eq. (20), which describes the distribution of mass-weighted displacements $\vec{y}$, Wheeler *et al.* apparently discuss in Eq. (3) of Ref. 20 the distribution of unweighted displacements $\vec{x}$. As a consequence, they obtain the eigenvalues of the Hessian matrix F instead of the vibrational frequencies, which are the square roots of the eigenvalues of the mass-weighted Hessian $F^{(M)}$. Nevertheless, we will assume from now on that $\vec{x}$ and F were actually supposed to denote mass-weighted entities, and that this has not been mentioned by the authors for reasons of brevity.

Likewise, Wheeler *et al.* do not mention the necessity to describe the atoms in the molecular frame of reference specified in Sec. II B. Therefore, we have to check the consequences, which arise for PMA if the atomic displacements are described in the laboratory frame L.

The transformation of displacements $\vec{y}_L$ from the laboratory frame into the molecular frame can be expressed by $\vec{y} = R[\vec{\varphi}(t)]\vec{y}_L + \vec{t}(t)$. Here, $\vec{t}(t)$ is the vector, which accounts for the translation of the center of mass and $R[\vec{\varphi}(t)]$ is the matrix describing the orientation of the molecule. It consists of N identical 3×3 rotation matrices and depends on the three Eulerian angles collected in the vector $\vec{\varphi}(t) = [\varphi_1(t), \varphi_2(t), \varphi_3(t)]$. In the laboratory frame of reference, the potential energy of the molecule is then

$$U_0(\vec{y}_L, t) = \tfrac{1}{2}\vec{y}_L^T F_{0,L}^{(M)}(t)\vec{y}_L, \tag{25}$$

where the Hessian matrix

$$F_{0,L}(t) = R^T[\vec{\varphi}(t)]F_0^{(M)}R[\vec{\varphi}(t)] \tag{26}$$

is explicitly time dependent due to the rigid-body rotation of the molecule. If the molecule assumes all orientations at equal probabilities during a MD trajectory, the distribution

$$p(\vec{y}_L) = \frac{1}{Z_L}\exp\left(-\frac{\vec{y}_L^T\langle F_{0,L}^{(M)}(t)\rangle_t \vec{y}_L}{2k_BT}\right) \tag{27}$$

of displacements $\vec{y}_L$ is characterized by the rotationally averaged mass-weighted Hessian

$$\langle F_{0,L}^{(M)}(t)\rangle_t = \langle R^T[\vec{\varphi}(t)]F_0^{(M)}R[\vec{\varphi}(t)]\rangle_{\vec{\varphi}} \tag{28}$$

from which one cannot derive vibrational frequencies by any means. Hence, one is forced to assume that the authors have tacitly assumed the use of the molecular frame of reference specified above.

However, if this is the case, their derivation contains an error, which becomes apparent in their Eq. (3). In this equation, which specifies the distribution $p(\vec{y})$, the δ distributions showing up in Eq. (20) and marking the constraints imposed by the molecular frame of reference are missing. This erroneously suggests that $p(\vec{y})$ is flat instead of sharply peaked in the translational and rotational directions. By identifying this distribution as a multivariate Gaussian, they erroneously continued to conclude that the associated covariance matrix $\sigma^{(M)}$ is the inverse of the mass-weighted Hessian matrix $F^{(M)}$ [see Eq. (4) in Ref. 20].

This unjustified conclusion has practical consequences for the calculation of the normal modes: In the molecular frame of reference, the covariance matrix $\sigma^{(M)}$ determined from a MD trajectory is singular (or numerically nearly singular). In contrast to our proposal to diagonalize $\sigma^{(M)}$ and to invert the nonzero eigenvalues afterwards according to Eq. (24), Wheeler *et al.* suggested to invert $\sigma^{(M)}$ before diagonalization,[20] despite its singularity. In applications, in which the determinant of $\sigma^{(M)}$ may not vanish completely due to numerical inaccuracies, this may be possible. However, the inversion of an almost singular matrix is numerically unstable when using standard inversion algorithms,[39] and will therefore lead to erroneous results for the nonsingular modes. Hence, we have to state that results obtained following the original proposal[20] will not be reliable. This statement represents a critique of least possible severity, because we have done our best to assume that the noted omissions concerning conditions (a) and (b) were mere consequences of notational brevity.

B. PMA for molecules with a time-independent Hamiltonian

Having shown above that the conditions (a) and (b) are crucial for a correct calculation of vibrational frequencies by PMA, we will now turn to the significance of conditions (c) and (d), which require a harmonic Hamiltonian and thermal equilibrium, respectively. Here, we still assume the Hamiltonian to be constant (e).

1. Isolated molecules with a purely harmonic potential

Dropping condition (d) but retaining (c), we first consider the case of an isolated molecule, freely oscillating in a purely harmonic potential. The solution of the equations of motion associated with the Hamiltonian (11) is given by the normal coordinates $a_j(t) = a_{j,0}\cos(\omega_j t + \varphi_j)$. The associated covariance matrix

$$\sigma_{ij}^{(a)} = \langle a_i(t)a_j(t)\rangle = \frac{a_{j,0}^2}{2}\delta_{ij} \tag{29}$$

is diagonal for nondegenerated frequencies ω_j. Considering Eq. (23), one can easily show that the covariance matrix $\sigma^{(M)}$ of Cartesian displacements $\vec{y}$ is diagonalized by the matrix C of mass-weighted normal modes. Thus, in the given case, PMA determines the mass-weighted normal modes correctly through the eigenvectors of $\sigma^{(M)}$.

According to the virial theorem, the total energy (11) is on the average equally distributed among its kinetic and potential parts for every single mode j. Because $\langle a_j\rangle$ vanishes for the given harmonic oscillation and because, therefore, $\langle a_j^2\rangle$ is the variance $\mathcal{V}(a_j)$ of the normal coordinate a_j, one obtains

$$\omega_j^2 = \langle \dot{a}_j^2\rangle / \mathcal{V}(a_j), \quad j \in \{7,\ldots,3N\}. \tag{30}$$

If the average kinetic energy $\langle \dot{a}_j^2 \rangle$ has the same value $k_B T$ for all modes j, then the virial expression (30) is equivalent to the PMA expression (24). Hence, PMA yields exact frequencies for an isolated polyatomic molecule with a purely harmonic Hamiltonian *if the kinetic energy is equally distributed among all modes*. However, if one simulates the oscillations of such a molecule by MD starting from an arbitrary initial deformation, this condition will be generally violated and the PMA frequencies (24) will deviate from the correct virial frequencies (30).

2. Molecules with a weakly anharmonic potential

Relaxing the PMA condition (c) we now introduce weak anharmonicities into the Hamiltonian while retaining the classical treatment of nuclear motion. Note that the resulting classical anharmonic frequency corrections will differ from corresponding quantum mechanical results (see Paper II of this work[13] for further details).

Dropping once again also condition (d), we first consider an isolated one-dimensional anharmonic oscillator, whose potential is given by

$$U(a) = \tfrac{1}{2}\gamma a^2 + \tfrac{1}{3}\beta a^3, \tag{31}$$

where a is a mass-weighted normal coordinate and, consequently, $\gamma \equiv \omega_H^2$ is given by the harmonic frequency ω_H [cf. Eq. (11)]. The frequency of a classical oscillator (31) is

$$\omega^2 = \omega_H^2 \left(1 - \frac{5\varepsilon^2}{6} + O(\varepsilon^3) \right), \tag{32}$$

where $\varepsilon^2 = 2\beta^2 k_B T/\gamma^3 \ll 1$ is assumed to be small and $k_B T$ is the total energy of the oscillator.[40] In Appendix A we show that the PMA prescription (24) yields the frequency

$$\omega_{\mathrm{iso}}^2 = \omega_H^2 \left(1 - \frac{27\varepsilon^2}{36} + O(\varepsilon^3) \right) \tag{33}$$

for a free trajectory of the anharmonic oscillator (31), which is almost correct to the order of ε^2. Hence, in the given case PMA gives a better approximation to the classical vibrational frequency than NMA, which completely neglects anharmonic effects.

Recovering condition (d) again, we next assume that the one-dimensional anharmonic oscillator (31) is in thermal equilibrium with an external heat bath. In Appendix B we show that in this case the PMA prescription (24) gives a frequency of

$$\omega_{\mathrm{eq}}^2 = \omega_H^2 [1 - 2\varepsilon^2 + O(\varepsilon^3)] \tag{34}$$

which *systematically underestimates* the frequency given in Eq. (32) by about as much as NMA overestimates it, which yields ω_H^2. In higher-dimensional cases, anharmonic couplings between the normal modes can additionally affect the results of PMA. These effects will be qualitatively discussed in Paper II of this work.[13]

3. Molecules with a purely harmonic potential, weakly coupled to a heat bath

According to condition (d) PMA requires a complete equilibrium with an external heat bath. Because in solution the thermal relaxation times of solute molecules are of the order of picoseconds, one does not necessarily obtain a thermally equilibrated system in a DFT/MM-MD simulation within manageable computation time. Therefore, in a given DFT/MM-MD trajectory the vibrational modes of the solute generally will be excited at different energies thus violating condition (d).

However, while retaining the PMA conditions of a (c) purely harmonic and (e) time-independent Hamiltonian, but dropping the equilibrium condition (d), one can still compute vibrational modes and frequencies exactly if one assumes that every single normal mode j has a Boltzmann-like distribution of potential energy with an associated mode specific temperature

$$T_j := \langle \dot{a}_j^2 \rangle / k_B . \tag{35}$$

Replacing in Eq. (21) the temperature T by the mode specific temperatures T_j yields a correspondingly modified distribution $p(\vec{a})$ of normal coordinates. Transforming $p(\vec{a})$ back into the coordinate system of the mass-weighted displacements $\vec{y}$ renders a distribution $p(\vec{y})$, whose principal axes are the mass-weighted normal modes $\vec{c}_j$ as assumed by PMA (cf. Sec. II C). Furthermore, the PMA frequencies (24) reduce to the virial expression (30), if T is replaced by the T_j and if Eq. (35) is used. Thus, the virial frequencies (30), originally derived for MD trajectories of an isolated multidimensional harmonic oscillator, remain exact upon introducing a weak coupling to a heat bath, whereas the PMA frequencies (24) contain errors determined by $\sqrt{T/T_j}$. Therefore, we call the correct values derived from Eq. (30) "generalized virial" (GV) frequencies and will focus on this expression instead of considering the error-prone PMA equation (24).

In the given case of unequally distributed excitation energies, special care has to be taken when the vibrational modes are calculated by PMA: Two eigenvalues λ_i and λ_j of the covariance matrix $\sigma^{(M)}$ are degenerated, if

$$\lambda_i = \frac{k_B T_i}{\omega_i^2} = \frac{k_B T_j}{\omega_j^2} = \lambda_j . \tag{36}$$

Such degeneracies can accidentally occur, even if the corresponding modes have different frequencies $\omega_i \neq \omega_j$. If they occur, the respective eigenvectors $\vec{d}_i$ and $\vec{d}_j$ of the covariance matrix may be mixed and hence cannot be associated to vibrational modes. In order to identify such cases, one has to check the eigenvalues λ_j of the covariance matrix $\sigma^{(M)}$ for degeneracies. Note that, on the other hand, degenerated normal modes may become separated in PMA if they are excited at different energies.

C. GV frequencies for molecules with a time-dependent Hamiltonian

Up to this point we have considered molecules which are either isolated or coupled to a heat bath. Correspondingly, it was justified to assume a time-independent molecular Hamiltonian H_0 as required by condition (e). However, the molecular Hamiltonians become time dependent in condensed phase due to interactions with the fluctuating environment. Time-dependent fluctuations of a molecular Hamiltonian

$H(t)$ cause time-dependent fluctuations of the equilibrium geometry $\vec{r}^0(t)$, of the normal modes $\vec{n}_j(t)$, and of the vibrational frequencies $\omega_j(t)$. This raises the question as to how these fluctuations affect the GV frequencies (30) calculated from a DFT/MM trajectory.

If the fluctuations of the molecular Hamiltonian $H(t)$ are fast and uncorrelated as compared to the time scales $\tau_j = 2\pi/\omega_j$ characterizing the mean field Hamiltonian $\langle H(t)\rangle$, then one can apply the theory of Brownian motion.[14,41] Correspondingly, Wheeler *et al.* have shown that PMA accurately yields the vibrational frequency of a Brownian particle in a one-dimensional harmonic potential.[21] However, for a molecule in solution the above conditions are violated, because the fluctuations of the molecular Hamiltonian $H(t)$, which, e.g., may be due to rearrangements of the solvent shell, are generally as fast or even *slower* than the oscillation periods τ_j of the molecule (see Paper II of this work[13] for an example proving this claim). Consequently, the vibrations of a molecule in solution cannot be modeled by a Brownian particle in a potential specified by $\langle H(t)\rangle$.

In order to elucidate the effects of a time-dependent Hamiltonian on the GV frequencies (30), we will now separately suspend the assumptions of constant normal modes $\vec{n}_j(t)$ of a constant equilibrium geometry $\vec{r}^0$ and of constant force constants assuming slow fluctuations.

Let us start with the effects of time-dependent normal modes $\vec{n}_j(t)$ on the GV frequencies (30). In the worst case, the composition of the normal modes $\vec{n}_j(t)$ may be very sensitive to the environment such that the mass-weighted normal modes $\vec{c}_j(t)$ rotate in the $(3N-6)$-dimensional configuration space occasionally yielding $\vec{c}_j(t)\cdot\vec{c}_j(t')\approx 0$ at different times t and t'. In this case, the assumption of constant normal modes is certainly wrong and the calculation of vibrational frequencies (30) from the moments of the associated normal coordinates $a_j(t)$ will lead to erroneous results. In the best case, the normal modes $\vec{n}_j(t)$ are only slightly fluctuating, such that $\vec{c}_i(t)\cdot\vec{c}_j(t')\approx\delta_{ij}$ for any t and t', and the inaccuracies in the associated normal coordinates $a_j(t)$ are also small. However, this latter case is expected to apply only to small molecules with well separated bands.

But even in the optimal case of constant normal modes, the assumptions of a constant equilibrium geometry and of constant force constants may not be valid. To estimate the effects of corresponding time dependencies, we have calculated the GV frequencies (30) of a one-dimensional harmonic oscillator for a slowly fluctuating equilibrium position or force constant in Appendix C.

For a harmonic oscillator of constant frequency ω_H, whose equilibrium position $\overline{a}(t)$ fluctuates with a standard deviation of $\sigma_{\overline{a}}$ around its average value, we find

$$\omega_{\mathrm{GV}} \approx \omega_H \left(1 - \frac{1}{2}\frac{\sigma_{\overline{a}}^2}{\mathcal{V}(a)}\right). \tag{37}$$

Thus, the GV frequency systematically underestimates ω_H. This result follows from Eq. (C4), which shows that the fluctuations $\sigma_{\overline{a}}^2$ of the equilibrium position $\overline{a}(t)$ add to the variance $\mathcal{V}(a)$ of the normal coordinate $a(t)$ measured in the calculation of the GV frequency (30). This enlargement of the measured variance mimics a weaker force constant and leads to the frequency underestimate exhibited by Eq. (37).

Similarly, for a harmonic "constant" $\gamma(t)\equiv\omega_H^2(t)$ fluctuating with a standard deviation σ_γ around the average value $\gamma_0\equiv\langle\omega_H^2\rangle$ we obtain

$$\omega_{\mathrm{GV}} \approx \langle\omega_H\rangle \left(1 - \frac{\sigma_\gamma^2}{4\gamma_0^2}\right), \tag{38}$$

where $\langle\omega_H\rangle$ is the average harmonic frequency. Therefore, also fluctuations of the force constant will entail a systematic underestimate of the GV frequency as compared to the average frequency $\langle\omega_H\rangle$ of our ensemble of harmonic oscillators.

When taken together the results (37) and (38) suggest that the GV approach will systematically underestimate the average vibrational frequencies $\langle\omega_j(t)\rangle$ of molecules in polar solvents. In contrast, the average frequency $\langle\omega_j(t)\rangle$ is exactly the one, which is sampled by the INMA procedure. Within the applied approximation scheme (harmonic vs anharmonic, quality of DFT/MM description) INMA can therefore yield exact average vibrational frequencies whereas GV cannot.

This observation raises two questions: (a) How large are the systematic frequency underestimates and the other GV artifacts outlined further above? (b) Are there means by which one can remedy the GV failures? We will give a preliminary answer to the first question by applying the GV prescription (30) to DFT/MM-MD trajectories of a sample molecule in Sec. IV. A detailed analysis and an answer to the second question will be presented in the subsequent paper.[13]

IV. A TEST APPLICATION OF PMA AND GV TO FORMALDEHYDE

In the above development and analysis of the PMA approach, we have identified a number of effects, which can cause differences between vibrational frequencies calculated by PMA and INMA, respectively. To avoid one of the key problems hampering the PMA prescription (24), we have derived the GV expression (30). Using a DFT-MD trajectory of formaldehyde (FA) in the gas phase and a DFT/MM-MD trajectory of FA in aqueous solution, we will now provide first estimates on the sizes of the analytically predicted deviations.

A. Methods

DFT/MM calculations of FA (H_2CO) in aqueous solution have been carried out with the MD program package EGO-MMII (Ref. 42) into which the DFT package CPMD (Ref. 43) has been integrated[11] for the treatment of DFT fragments. For FA representing the DFT fragment, we have used the functionals of Becke and Perdew[44,45] and pseudopotentials of Troullier and Martins.[46] We have applied an 80 Ry cutoff to the plane wave basis, since test calculations have shown that

the calculated vibrational frequencies of FA in the gas phase do not change significantly when the cutoff is raised above the given value (data not shown).

For the oxygen, carbon, and hydrogen atoms of FA, the van der Waals parameters $\varepsilon_{vdW}=0.20/0.12/0.01$ and $\sigma_{vdW}=2.85/3.30/2.74$ have been used, respectively.[47] The MM fragment consisted of 1970 rigid MM water molecules as described by Jorgensen's four-point transferable intermolecular potential (TIP4P).[48] As our simulation system, we have chosen a rhombic dodecahedron with periodic boundary conditions. The electrostatic interactions have been calculated up to a cutoff radius of 20 Å by means of the fast structure-adapted multipole method (SAMM, Refs. 49–51) and by a reaction field approach[52] beyond this cutoff.

In order to control the thermodynamic parameters of the simulation system, we have applied a barostat[53] (compressibility 0.49 GPa^{-1}, coupling time constant of 10 ps) and a thermostat[53] (coupling time constant 1 ps) to the solvent molecules, if required in the respective simulated ensembles. We have chosen a time step as large as 0.5 fs, instead of the usual 0.25 fs (which may have been slightly overoptimistic; see Paper II of this work[13]).

The system has been equilibrated for 200 ps in an N-P-T ensemble at a pressure of 1 atm and a temperature of 300 K. Here, a MM force field has also been applied to the FA molecule. After this preliminary equilibration, the force field has been switched to the above described DFT/MM force field. After a further equilibration of 1 ps, the system has been simulated in the N-V-T ensemble for 35 ps. During this DFT/MM simulation, the FA molecule has been excluded from the coupling to the thermostat in order to avoid artifacts affecting the vibrational spectra.

In order to compare the PMA and GV results with INMA, we have selected 70 snapshots of the 35 ps trajectory at temporal distances of 0.5 ps, although these short distances do not guarantee complete statistical independence. However, as will become apparent in Paper II of this work[13] a substantial decorrelation takes place at the chosen time scale. The structure of minimum potential energy of the FA molecule has been calculated at each of these snapshots by steepest gradient descent, while the water environment was kept immobile (frozen cage). The Hessian matrix has been determined by displacing each atom by 0.01 Å from its minimum position.

The 10 ps trajectory at 339 K and the vibrational frequencies of FA in the gas phase have been calculated using the same DFT force field in the N-V-E ensemble. This trajectory started from an arbitrarily displaced geometry.

B. Results and discussion

The diagonalization of the covariance matrix (23) obtained from the trajectory of FA in the gas phase yielded six vibrational modes, which are the symmetric (CH^{sym}, $A1$) and the asymmetric (CH^{asy}, $B2$) hydrogen stretches, the carbonyl stretch (CO, $A1$), the in-plane bending (HCH, $A1$) and rocking (HCO, $B2$) mode, and the out-of-plane bending mode (oop, $B1$). In parentheses, we have given abbreviations and symmetries for the various modes. Essentially the same normal modes have been obtained by diagonalization of the covariance matrix (23) obtained from the trajectory of FA in solution, by NMA in the gas phase, and by INMA for all snapshots in solution.

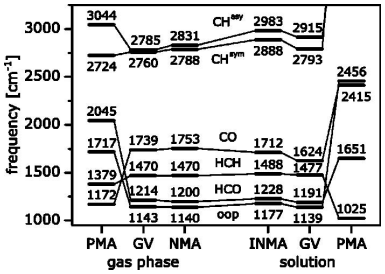

FIG. 1. Vibrational frequencies of FA in the gas phase (left) and in solution (right) for each phase calculated with PMA (24), GV (30), and INMA.

1. Vibrational frequencies of FA in the gas phase

The left part of Fig. 1 displays the vibrational frequencies of FA obtained by the PMA approach (24), our GV prescription (30), and by NMA in the gas phase.

The harmonic frequencies determined by NMA agree within an 8 cm^{-1} root mean square deviation (rmsd) with those determined by a recent DFT study, in which the same density functional but an extended basis set were used.[18] In contrast, the PMA frequencies (24) strongly deviate by a rmsd of 457 cm^{-1} from the NMA results. The main reason for this huge deviation is identified by the following consideration.

According to Sec. III B 1 the assumption (d) of an equal distribution of kinetic energy among the normal modes is essential for the validity of the PMA prescription (24). However, this assumption is violated in the given 10 ps DFT trajectory. Here, each of the six vibrational modes has been initially excited at a different energy leading to a 226 K standard deviation of the mode-specific temperatures T_j from the average temperature $T=339$ K. In contrast, Fig. 1 shows that the GV frequencies (30) are much closer to the NMA results. Here, the rmsd is only 10 cm^{-1} for the modes with frequencies below 2000 cm^{-1}, is 38 cm^{-1} for the hydrogen stretching modes, and is 23 cm^{-1} for all modes. We conclude that the PMA approach cannot be applied to gas phase trajectories of molecules, if the initial displacement is chosen arbitrarily.

As explained in Sec. III B 2, NMA neglects anharmonic terms in the potential energy, whereas GV can reproduce anharmonic shifts in the fundamental frequency to the extent, to which the anharmonic contributions to the potential of an isolated molecule are sampled by the classical trajectory. On the other hand, as also explained in the quoted section, the GV frequencies may be subject to inaccuracies due to anharmonic mode coupling. We attribute the remaining 23 cm^{-1} rmsd between the GV and NMA frequencies observed in the left part of Fig. 1 to these two effects. By quoting experimental data[54] and computational results[18] on the true sizes of anharmonic frequency shifts as well as by applying various

other methods to the DFT trajectory, Paper II of this work[13] explains in detail the relative sizes, by which the two effects contribute to the noted deviations.

2. Vibrational frequencies of FA in solution

The right part of Fig. 1 displays the vibrational frequencies obtained by the PMA expression (24) and by the GV prescription (30) from the DFT/MM solution trajectory. These results are compared with the average frequencies obtained by INMA. We immediately note that INMA predicts a 41 cm^{-1} redshift of the CO mode upon solvation of FA in the highly polar aqueous solvent, which approximately matches the well-known behavior of carbonyl modes upon hydrogen bonding.

For the solution case, Fig. 1 exhibits even larger deviations between the PMA frequencies (24) and those obtained by INMA than for the isolated FA molecule as witnessed by the rmsd of 1216 cm^{-1}. The main source of this huge deviation is identical to the one identified above in connection with the gas phase spectra: Thermal equilibrium was absent within the FA molecule as demonstrated by a 173 K standard deviation of the normal mode energy distribution around an average temperature of 258 K.

For the given DFT/MM solution case the large 173 K standard deviation of the kinetic energy distribution is surprising, because the FA molecule had the chance to exchange energy with its 300 K environment. As explained in the supplementary information[55] the observed lack of equilibration is caused by the slow thermal relaxation of FA and by rare and artificial force jumps acting on the atoms of the DFT fragment. These jumps, which artificially inject heat into specific normal modes, are due to an unsufficiently tight convergence criterion imposed by us on the DFT wave function.[55] Partially due to this computational artifact (that cannot affect other methods such as GV and its variants, FTTCF, or INMA) but mainly because of the slow thermal relaxation, the PMA assumption (d) of thermal equilibrium has been strongly violated in our DFT/MM-trajectory and therefore, the PMA frequencies obtained by Eq. (24) are completely wrong.

In contrast, the GV frequencies (30) displayed in Fig. 1 are redshifted with respect to the INMA frequencies by a rmsd of only 64 cm^{-1} and, due to the use of mode-specific temperatures, apparently unaffected by the artificial force jumps.

For a more detailed analysis of these differences, we concentrate on the important CO mode. In the gas phase, anharmonicities have shifted the corresponding GV frequency by 14 cm^{-1} to the red of the corresponding harmonic NMA frequency, whereas in solution the GV frequency is shifted by 88 cm^{-1} to the red of the INMA result. According to our analysis in Sec. III B 2 the anharmonic redshifts monitored by GV are expected to be about two times larger in solution than in the gas phase [compare Eqs. (33) and (34)]. This argument applies, if the mode specific temperatures T_{CO} are equal in both cases. However, in the gas phase DFT simulation the CO mode was excited at $T_{CO}=747$ K whereas in the DFT/MM simulation T_{CO} was smaller by a factor of ~ 6. This difference leads to the estimate that the GV anharmonic redshift is by a factor of 3 smaller in solution than in the gas phase [cf. the temperature dependence of the parameter ε appearing in Eqs. (32)–(34)]. Accordingly, the GV frequency of the CO mode in solution should be redshifted by only 5 cm^{-1} with respect to the harmonic INMA frequency. As a result, the large 88 cm^{-1} redshift cannot be attributed to anharmonicities. There must be a much more important second effect, which causes the redshifts of the GV frequency with respect to the INMA result.

TABLE I. IR intensities of FA in the gas phase and in solution as obtained by Eq. (16) from the MD trajectories and by NMA/INMA using Eq. (13). The experimental values (Ref. 56) have been uniformly scaled to yield the same average value as NMA.

	IR intensities (a.u.)				
	Gas phase			Solution	
	Eq. (16)	NMA	Expt.[a]	Eq. (16)	INMA
CH^{asy}	3.48	3.35	2.80	0.93	0.81
CH^{sym}	1.99	1.92	2.33	1.00	1.11
CO	2.42	2.48	2.31	3.32	3.09
HCH	0.24	0.22	0.40	0.63	0.66
HCO	0.20	0.23	0.31	0.20	0.18
Imp	0.15	0.14	0.25	0.29	0.20

[a]Reference 56.

To identify this second effect, recall the results of Sec. III C, which show that the violation of the assumption (e) of a time-independent molecular Hamiltonian entails systematic frequency underestimates by GV in the condensed phase. Therefore, we preliminarily attribute the 64 cm^{-1} rmsd redshift of the complete spectrum and, in particular, the 88 cm^{-1} redshift of the CO mode mainly to the violation of the assumption (e).

By comparing the vibrational frequencies obtained by PMA and by Fourier methods from a hybrid MD trajectory of a quantum mechanical water molecule (as described by Hartree–Fock) in MM bulk water, Wheeler *et al.* have found PMA frequencies redshifted by a rmsd of about 550 cm^{-1} with respect to the results of the Fourier methods.[21] Apparently unaware of the systematic frequency underestimates characteristic to PMA they have assigned these deviations to an alleged inability of Fourier methods to properly account for anharmonic motions.[21] We will thoroughly discuss these issues and the preliminary explanation voiced further above in Paper II of this work.[13]

3. IR intensities

Table I compares the IR intensities, which are calculated by Eq. (16) from the MD trajectories using normal coordinates $a_j(t)$ derived from NMA/INMA normal modes $\vec{n}_j$, with the IR intensities obtained by NMA/INMA in the usual way. Also shown are gas phase experimental IR intensities.[56] Apparently, both in the gas phase and in solution the calculated IR intensities agree quite well with each other and, for the gas phase case, also with the experimental data. For the solution case experimental data are not available because unfortunately the FA molecule reacts with water. Thus, we refer to Ref. 10 for quantitative comparisons of DFT/MM INMA intensities with IR data. Nevertheless, the sizable increase of

the IR intensity of the CO mode, which is predicted by our calculations upon hydration, generally agrees with experimental data and other calculations on the behavior of various carbonyl compounds upon solvation in polar solvents.[57]

V. SUMMARY

In this paper, we have thoroughly reinvestigated a proposal aiming at the computation of vibrational spectra from classical MD trajectories by means of a principal component analysis.[20] By rederiving this so-called PMA approach we have identified the five conditions under which PMA can yield exact frequencies: (a) One has to use mass-weighted displacements, (b) the total linear and angular momenta of the molecule have to vanish in the frame of reference, (c) the intramolecular potential has to be harmonic, (d) the kinetic energy has to be equally distributed among the normal modes during the MD simulation, and (e) the molecular Hamiltonian has to be time independent. In addition, we have given a prescription as to how one can calculate the IR intensities of the various normal modes from a MD trajectory.

However, in gas phase DFT-MD and in solution DFT/MM-MD simulations of molecules the conditions [(c),(d)] and [(c)–(e)], respectively, are generally violated: The potential energy surface of the molecule is anharmonic, thermal equilibrium is not guaranteed, and in solution the molecular Hamiltonian is time dependent due to the interaction of the molecule with its fluctuating environment.

According to our analysis the violation of condition (d) entails PMA errors determined by $\sqrt{T_j/T}$, where T_j is the mode specific temperature (35) and T is the one of the simulation system. To avoid these errors we have suggested to apply T_j in the frequency computations leading to the expression (30), which has been identified as a generalized virial frequency (GV). As shown by simple analytical arguments for the solution case, the violations of conditions (c) and (e) both entail systematic underestimates of the GV and PMA frequencies.

Sample DFT and DFT/MM simulations of a FA molecule in the gas phase and in aqueous solution, respectively, have demonstrated that the GV approach cures the most severe artifacts of PMA frequency computation, which are induced by violations of condition (d). However, as also shown by our solution simulation, large artifacts presumably caused by violation of condition (e) persist. The IR intensities calculated from the MD trajectory by our prescription closely agree with experimental data and NMA/INMA results.

The question remains, to what extent the systematic frequency underestimates still hampering also the GV approach can be avoided by further modifications of the computational procedures. These issues will be addressed in Paper II of this work.[13]

ACKNOWLEDGMENTS

Support by the Volkswagen Stiftung (Project No. I/73224) and by the Deutsche Forschungsgemeinschaft (Grant No. SFB 533/C3) is gratefully acknowledged.

APPENDIX A: ISOLATED, WEAKLY ANHARMONIC OSCILLATOR

Here we discuss PMA applied to an isolated one-dimensional anharmonic oscillator with a potential energy given by Eq. (31). The classical equation of motion in this potential is given by

$$-\ddot{a}=\gamma a+\beta a^2. \tag{A1}$$

Let us assume a motion with a small amplitude $a_{\max}>0$ in which the relative contribution $\beta a_{\max}^2/\gamma a_{\max}\ll 1$ of the anharmonic term to the potential energy is weak. Then the motion is confined around $a=0$, such that we can make the ansatz of an oscillatory motion,

$$a(t)=\sum_{n=0}^{\infty} a^{(n)}\cos n\omega t, \tag{A2}$$

where $\dot{a}(0)=0$ has been chosen without loss of generality as the initial condition. After applying the ansatz (A2) to the equation of motion (A1), a comparison of coefficients yields the oscillation frequency

$$\omega^2=\gamma\left(1-\frac{5\eta^2}{6}+O(\eta^3)\right) \tag{A3}$$

and the coefficients

$$\begin{aligned}
a^{(0)}&=\frac{\gamma}{\beta}[-\eta^2/2+O(\eta^3)],\\
a^{(1)}&=\frac{\gamma\eta}{\beta},\\
a^{(2)}&=\frac{\gamma}{\beta}[\eta^2/6+O(\eta^3)],\\
a^{(n)}&=\frac{\gamma}{\beta}O(\eta^n),\quad n\geqslant 3,
\end{aligned} \tag{A4}$$

where the small parameter $\eta=\beta a^{(1)}/\gamma\ll 1$ accounts for the assumption of small displacements (for a similar derivation see Ref. 40).

With these coefficients, one can easily derive the temperature

$$k_BT=\langle\dot{a}^2\rangle=\frac{\eta^2\gamma^3}{2\beta^2}\left(1-\frac{26\eta^2}{36}+O(\eta^3)\right) \tag{A5}$$

of the oscillator and the variance

$$\mathcal{V}(a)=\frac{\eta^2\gamma^2}{2\beta^2}\left(1+\frac{\eta^2}{36}+O(\eta^3)\right) \tag{A6}$$

of the normal coordinate a. Using Eq. (24) for the calculation of the frequency yields

$$\omega_{\mathrm{iso}}^2=\gamma\left(1-\frac{27\eta^2}{36}+O(\eta^3)\right), \tag{A7}$$

which represents a good approximation to the oscillation frequency (A3).

By multiplying Eq. (A5) with $2\beta^2/\gamma^3$, one can introduce a new parameter ε by

$$\varepsilon^2=\frac{2\beta^2 k_B T}{\gamma^3}=\eta^2+O(\eta^4), \tag{A8}$$

which only depends on the temperature T of the oscillator. With this parameter, we find for the temperature dependence of the classical oscillation frequency

$$\omega^2=\gamma\left(1-\frac{5\varepsilon^2}{6}+O(\varepsilon^3)\right) \tag{A9}$$

and of the PMA frequency

$$\omega_{\mathrm{iso}}^2=\gamma\left(1-\frac{27\varepsilon^2}{36}+O(\varepsilon^3)\right), \tag{A10}$$

respectively.

APPENDIX B: WEAKLY ANHARMONIC OSCILLATOR, COUPLED TO A HEAT BATH

Here we discuss PMA applied to a classical one-dimensional anharmonic oscillator, which is coupled to a heat bath of the temperature T. Figure 2 shows the one-dimensional potential $U(a)$ as given by Eq. (31) and the associated normalized Boltzmann distribution

$$p(a)=\exp\left(-\frac{U(a)}{k_B T}\right)=\exp\left(-\frac{a^2}{2b^2}-\frac{a^3}{3c^3}\right), \tag{B1}$$

where

$$b=\sqrt{\frac{k_B T}{\gamma}},\quad c=\sqrt[3]{\frac{k_B T}{\beta}}>0. \tag{B2}$$

The distribution (B1) diverges for $a\to-\infty$ and, therefore, cannot be integrated as required by Eq. (24) for PMA. In a dynamical picture, any thermal distribution at $T>0$, which is initially confined at $a=0$ would leak across the finite potential barrier, whose maximum is located at $a=-d$, where $d=\gamma/\beta>0$. In order to prevent such a leaking and to ensure the integrability, we modify the potential by introducing a potential wall at the position of the maximum. Figure 2 shows this wall in the modified potential

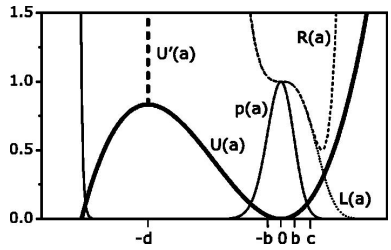

FIG. 2. Explanation of the geometry applied for the normalization of the Boltzmann distribution (B1). Thick solid line: anharmonic potential $U(a)$ given by Eq. (31), scaled to fit into the graph. Thin solid: unnormalized Boltzmann distribution $p(a)$ in this potential given by Eq. (B1). Thick dashed: modified potential $U'(a)$ given by Eq. (B3), scaled with the same factor as $U(a)$. Dot: left-hand side $L(a)$ of Eq. (B7). Dash: right-hand side $R(a)$ of Eq. (B7). For all functions, the parameters are $\gamma=1$, $\beta=0.1$, and $k_B T=1$.

$$U'(a)=\begin{cases}U(a) & \text{for } a\geqslant -d,\\ \infty & \text{for } a<-d.\end{cases} \tag{B3}$$

In the associated unnormalized density

$$p'(a)=\exp\left(-\frac{a^2}{2b^2}-\frac{a^3}{3c^3}\right)\Theta(d+a), \tag{B4}$$

this wall is expressed by the Heaviside function Θ. In order to keep the effects of this barrier small, we require that the relative density at the cutoff $a=-d$ is low, i.e., that

$$\frac{p'(-d)}{p'(0)}=\exp\left(-\frac{d^2}{2b^2}+\frac{d^3}{3c^3}\right)=\exp\left(-\frac{1}{3\varepsilon^2}\right)\ll 1 \tag{B5}$$

with ε^2 as defined in Eq. (A8). Thus, the cutoff effect vanishes exponentially with $\varepsilon\to 0$. Accordingly, we will assume from now on that

$$\varepsilon^2\ll 1. \tag{B6}$$

In order to obtain analytical expressions for the moments of the distribution (B4), we have to approximate the cubic part of the exponential by the first terms of its Taylor expansion

$$\exp\left(-\frac{a^3}{3c^3}\right)\approx 1-\frac{a^3}{3c^3}+\frac{a^6}{18c^6} \tag{B7}$$

around $a^3/c^3=0$. Figure 2 compares the left-hand side $L(a)$ with the right-hand side $R(a)$ of Eq. (B7). Note that the approximation is exact in the limit $a/c\to 0$ but deviates if a is of the order of c.

We now have to check, to what extent these deviations affect the total density $p'(a)$ if the approximation (B7) is applied to (B4), yielding

$$p'(a)\approx\exp\left(-\frac{a^2}{2b^2}\right)\left(1-\frac{a^3}{3c^3}+\frac{a^6}{18c^6}\right)\Theta(d+a). \tag{B8}$$

According to Fig. 2, the density $p'(a)$ is small at distances $a>c$. Analytically, the relative density $p'(a)$ is always smaller than

$$p'(c)=\exp\left(-\frac{c^2}{2b^2}-\frac{1}{3}\right)=\exp\left(-\frac{1}{(2\varepsilon)^{2/3}}-\frac{1}{3}\right)$$

in that region. Similar results are obtained for $a<-c$ and for the approximated density (B8). This implies that the error caused by the approximation (B8) is of the order $\exp[-1/(2\varepsilon)^{2/3}]$ and vanishes exponentially for $\varepsilon\to 0$.

For the analytical calculation of the moments $\langle a^m\rangle$ of this distribution, the approximation

$$\int_{-d}^{\infty}a^n\exp\left(-\frac{a^2}{2b^2}\right)da\approx\int_{-\infty}^{\infty}a^n\exp\left(-\frac{a^2}{2b^2}\right)da \tag{B9}$$

has to be applied to the occurring integrals. The error introduced by this approximation is of the order $\exp(-d^2/2b^2)=\exp(-1/\varepsilon^2)$ and also vanishes exponentially for $\varepsilon\to 0$.

With these approximations, the moments of $p'(a)$ can be easily calculated. The mean displacement

$$\langle a\rangle=-d\left(\frac{\varepsilon^2}{2}+O(\varepsilon^3)\right) \tag{B10}$$

is at the leading order equal to the mean displacement (A4) of the free oscillator. Now Eq. (24) allows us to calculate the frequency from the variance

$$\mathcal{V}(a)=\frac{d^2\varepsilon^2}{2}[1+2\varepsilon^2+O(\varepsilon^3)], \tag{B11}$$

which yields

$$\omega_{\mathrm{eq}}^2=\gamma[1-2\varepsilon^2+O(\varepsilon^3)]. \tag{B12}$$

Hence, in the limits considered here, the classical oscillation frequency (A9) of an anharmonic oscillator in thermal equilibrium with an external heat bath is underestimated by PMA.

APPENDIX C: EFFECT OF TIME-DEPENDENT PARAMETERS OF THE OSCILLATOR

We now want to discuss the GV prescription (30) applied to a molecule with a time-dependent Hamiltonian $H(t)$. For this purpose we consider a one-dimensional oscillator with a potential energy of the form

$$U(a,t)=\tfrac{1}{2}\gamma(t)[a-\overline{a}(t)]^2. \tag{C1}$$

First we will discuss the case, in which the force constant $\gamma(t)\equiv\omega_H^2$ is time independent and the equilibrium position $\overline{a}(t)$ is time dependent. Subsequently, we will discuss the complementary case of a time-dependent force constant $\gamma(t)$ and of a constant equilibrium position $\overline{a}(t)\equiv 0$.

1. Time-dependent equilibrium position

We start with assuming the harmonic constant $\gamma(t)$ to be time independent, i.e., $\gamma\equiv\omega_H^2$. Let us assume that the time-dependent equilibrium position $\overline{a}(t)$ is a random variable with a probability distribution $p(\overline{a})$ centered around the mean value $\langle\overline{a}\rangle=0$ at a variance $\mathcal{V}(\overline{a})=\sigma_{\overline{a}}^2$. If $\overline{a}(t)$ is slowly varying, we can assume a Boltzmann distribution

$$p(a|\overline{a})=\sqrt{\frac{\omega_H^2}{2\pi k_BT}}\exp\left(-\frac{\omega_H^2(a-\overline{a})^2}{2k_BT}\right) \tag{C2}$$

of the normal coordinate a for each given $\overline{a}$. The total variance of a is then

$$\mathcal{V}(a)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}a^2p(a|\overline{a})p(\overline{a})d\overline{a}\,da. \tag{C3}$$

Subsequent integration over a and $\overline{a}$ gives

$$\mathcal{V}(a)=\frac{k_BT}{\omega_H^2}+\sigma_{\overline{a}}^2. \tag{C4}$$

Inserting Eq. (C4) into the GV expression (30) and neglecting terms of the order $O([\sigma_{\overline{a}}^2/\mathcal{V}(a)]^2)$ yields Eq. (37).

2. Time-dependent force constant

We continue with the complementary case of a time-dependent force constant $\gamma(t)$ by considering a potential

$$U(a,t)=\tfrac{1}{2}\gamma(t)a^2. \tag{C5}$$

We assume that $\gamma(t)$ is a random variable with probability distribution $p(\gamma)$ centered around its mean value $\langle\gamma\rangle=\gamma_0\gg\sigma_\gamma$ at a variance $\mathcal{V}(\gamma)=\sigma_\gamma^2$. The average oscillation frequency is

$$\langle\omega_H\rangle=\int\sqrt{\gamma}p(\gamma)d\gamma. \tag{C6}$$

Expanding $\sqrt{\gamma}$ around γ_0 and neglecting terms of the order $(\gamma-\gamma_0)^3/\gamma_0^3$ or higher leads to the average frequency

$$\langle\omega_H\rangle=\sqrt{\gamma_0}\left(1-\frac{\sigma_\gamma^2}{4\gamma_0^2}\right). \tag{C7}$$

Assuming that $\gamma(t)$ is slowly varying leads to the Boltzmann distribution

$$p(a|\gamma)=\sqrt{\frac{\gamma}{2\pi k_BT}}\exp\left(-\frac{\gamma a^2}{2k_BT}\right) \tag{C8}$$

for each given force constant γ. Hence, the variance of a is

$$\mathcal{V}(a)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}a^2p(a|\gamma)p(\gamma)d\gamma\,da. \tag{C9}$$

Integration over a gives

$$\mathcal{V}(a)=\int_{-\infty}^{\infty}\frac{k_BT}{\gamma}p(\gamma)d\gamma. \tag{C10}$$

Expanding $1/\gamma$ around γ_0 and neglecting terms of the order $O[(\gamma-\gamma_0)^3/\gamma_0^3]$ yields after integration

$$\mathcal{V}(a)\approx\frac{k_BT}{\gamma_0}\left(1+\frac{\sigma_\gamma^2}{\gamma_0^2}\right). \tag{C11}$$

Inserting Eq. (C11) into the GV expression (30) yields Eq. (38).

[1] M. Nonella, G. Mathias, M. Eichinger, and P. Tavan, J. Phys. Chem. B **107**, 316 (2003).
[2] M. Nonella, G. Mathias, and P. Tavan, J. Phys. Chem. A **107**, 8638 (2003).
[3] M. Schmitz, P. Tavan, and M. Nonella, Chem. Phys. Lett. **349**, 342 (2001).
[4] A. Barth and C. Zscherp, Q. Rev. Biophys. **35**, 369 (2002).
[5] R. Vogel, F. Siebert, G. Mathias, P. Tavan, G. Fan, and M. Sheves, Biochemistry **42**, 9863 (2003).
[6] K. Hauser, M. Engelhard, N. Friedman, M. Sheves, and F. Siebert, J. Phys. Chem. A **106**, 3553 (2002).
[7] J. Breton, J.-R. Burie, C. Berthomieu *et al.*, Biochemistry **33**, 4953 (1994).
[8] C. Allin, M. R. Ahmadian, A. Wittinghofer, and K. Gerwert, Proc. Natl. Acad. Sci. U.S.A. **98**, 7754 (2001).
[9] K. G. R. Pachler, F. Matlok, and H.-U. Gremlich, *Merck FTIR Atlas* (VCH, Weinheim, 1988).
[10] M. Klähn, G. Mathias, C. Kötting, M. Nonella, J. Schlitter, K. Gerwert, and P. Tavan, J. Phys. Chem. A **108**, 6186 (2004).
[11] M. Eichinger, P. Tavan, J. Hutter, and M. Parrinello, J. Chem. Phys. **110**, 10452 (1999).
[12] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[13] M. Schmitz and P. Tavan, J. Chem. Phys. **121**, 12247 (2004) (following paper).
[14] D. A. McQuarrie, *Statistical Mechanics* (Harper and Row, New York, 1976).
[15] E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular Vibrations* (McGraw-Hill, London, 1955).
[16] L. A. Woodward, *Introduction to the Theory of Molecular Vibrations and Vibrational Spectroscopy* (Clarendon, Oxford, 1972).
[17] H. H. Nielsen, Rev. Mod. Phys. **23**, 90 (1951).
[18] J. Neugebauer and B. A. Hess, J. Chem. Phys. **118**, 7215 (2003).
[19] Q. Cui and M. Karplus, J. Chem. Phys. **112**, 1133 (2000).

[20] R. A. Wheeler, H. Dong, and S. E. Boesch, ChemPhysChem **4**, 382 (2003).
[21] R. A. Wheeler and H. Dong, ChemPhysChem **4**, 1227 (2003).
[22] S. Mukamel, *Principles of Nonlinear Optical Spectroscopy* (Oxford University Press, Oxford, 1995).
[23] R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II—Nonequilibrium Statistical Mechanics* (Springer, Heidelberg, 1985).
[24] R. Kubo, Adv. Chem. Phys. **15**, 101 (1969).
[25] R. M. Stratt, Acc. Chem. Res. **28**, 201 (1995).
[26] T. Keyes, J. Phys. Chem. **101**, 2921 (1997).
[27] T. Kalbfleisch and T. Keyes, J. Chem. Phys. **108**, 7375 (1998).
[28] H. Ahlborn, X. Ji, B. Space, and P. B. Moore, J. Chem. Phys. **111**, 10622 (1999).
[29] H. Ahlborn, B. Space, and P. B. Moore, J. Chem. Phys. **112**, 8083 (2000).
[30] B. M. Ladanyi and R. M. Stratt, J. Phys. Chem. **99**, 2502 (1995).
[31] J. Borysow, M. Moraldi, and L. Frommhold, Mol. Phys. **56**, 913 (1985).
[32] P. A. Egelstaff, Adv. Phys. **11**, 203 (1962).
[33] R. Ramirez, T. Lopez-Ciudad, P. Kumar, and D. Marx, J. Chem. Phys. **121**, 3973 (2004).
[34] M. P. Gaigeot and M. Sprik, J. Phys. Chem. B **107**, 10344 (2003).
[35] B. R. Brooks, D. Janezic, and M. Karplus, J. Comput. Chem. **16**, 1522 (1995).
[36] A. Strachan, J. Chem. Phys. **120**, 1 (2004).
[37] J. Bredenbeck, J. Helbing, A. Sieg *et al.*, Proc. Natl. Acad. Sci. U.S.A. **100**, 6452 (2003).
[38] C. Eckart, Phys. Rev. **47**, 552 (1935).
[39] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
[40] N. W. McLachlan, *Ordinary Non-Linear Differential Equations in Engineering and Physical Sciences* (Clarendon, Oxford, 1950).
[41] S. Chandrasekhar, Rev. Mod. Phys. **15**, 1 (1943).
[42] G. Mathias, M. Eichinger, H. Carstens *et al.*, EGO-MMII *Users Guide* (Ludwig Maximilians Universität München, München, 2004).
[43] J. Hutter, A. Alavi, T. Deutsch, M. Bernasconi, S. Goedecker, D. Marx, T. Tuckermann, and M. Parinello, CPMD *Version 3.7.2* (MPI für Festkörperforschung Stuttgart, Suttgart, 2003).
[44] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).
[45] J. P. Perdew, Phys. Rev. B **33**, 8822 (1986).
[46] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
[47] J. S. Bader and B. J. Berne, J. Chem. Phys. **104**, 1293 (1996).
[48] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[49] C. Niedermeier and P. Tavan, J. Chem. Phys. **101**, 734 (1994).
[50] C. Niedermeier and P. Tavan, Mol. Simul. **17**, 57 (1996).
[51] M. Eichinger, H. Grubmüller, H. Heller, and P. Tavan, J. Comput. Chem. **18**, 1729 (1997).
[52] G. Mathias, B. Egwolf, M. Nonella, and P. Tavan, J. Chem. Phys. **118**, 10847 (2003).
[53] H. J. Berendsen, J. P. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[54] D. A. Clabo, W. D. Allen, R. B. Remington, Y. Yamaguchi, and H. F. Schaefer III, Chem. Phys. **123**, 187 (1988).
[55] See EPAPS Document No. E-JCPSA6-121-511447 for detailed information on the rare artificial force jumps affecting the DFT fragment during the DFT/MM simulation. A direct link to this document may be found in the online article's HTML reference section. The document may also be reached via the EPAPS homepage (http://www.aip.org/pubservs/epaps.html) or from ftp.aip.org in the directory /epaps/. See the EPAPS homepage for more information.
[56] T. Nakanaga, S. Kondo, and S. Saeki, J. Chem. Phys. **76**, 3860 (1982).
[57] M. W. Wong, K. B. Wiberg, and M. Frisch, J. Chem. Phys. **95**, 8991 (1991).

3.2 Zeitaufgelöste Infrarotspektren und Vergleich verschiedener Methoden

Im folgenden Abschnitt, der ein Abdruck des Artikels[2]

> Matthias Schmitz and Paul Tavan: „Vibrational spectra from atomic fluctuations in dynamics simulations: II. Solvent-induced frequency fluctuations at femtosecond time-resolution“, J. Chem. Phys. **121**, 12247-12258 (2004).

ist, wird das in Abschnitt 3.1 entwickelte GV Verfahren zur Berechnung von MIR Spektren einzelner Moleküle in polaren Lösungsmitteln weiter verbessert und derart modifiziert, dass es eine subpikosekunden-zeitaufgelöste Berechnung von IR Spektren ermöglicht. Schließlich werden die verschiedenen Methoden zur Berechnung von MIR Spektren anhand der Ergebnisse einer exemplarischen Anwendung miteinander verglichen.

[2]Mit freundlicher Genehmigung des American Institute of Physics.

Vibrational spectra from atomic fluctuations in dynamics simulations. II. Solvent-induced frequency fluctuations at femtosecond time resolution

Matthias Schmitz and Paul Tavan
Theoretische Biophysik, Lehrstuhl für Biomolekulare Optik, Ludwig-Maximilians-Universität München, Oettingenstrasse 67, 80538 München, Germany

(Received 25 May 2004; accepted 29 September 2004)

The midinfrared (MIR) spectra of molecules in polar solvents exhibit inhomogeneously broadened bands whose spectral positions are shifted as compared to the gas phase. The shifts are caused by interactions with structured solvation shells and the broadenings by fluctuations of these interactions. The MIR spectra can be calculated from hybrid molecular dynamics (MD) simulations, which treat the solute molecule by density functional theory and the solvent by molecular mechanics by the so-called instantaneous normal mode analysis (INMA) or by Fourier transforming the time correlation function (FTTCF) of the molecular dipole moment. In Paper I of this work [M. Schmitz and P. Tavan, J. Chem. Phys. **121**, 12233 (2004)] we explored an alternative method based on generalized virial (GV) frequencies noting, however, that GV systematically underestimates frequencies. As shown by us these artifacts are caused by solvent-induced fluctuations of the (i) equilibrium geometry, (ii) force constants, and (iii) normal mode directions as well as by (iv) diagonal and (v) off-diagonal anharmonicities. Here we now show, by analyzing the time scales of fluctuations and sample MD trajectories of formaldehyde in the gas phase and in water, that all these sources of computational artifacts can be made visible by a Fourier analysis of the normal coordinates. Correspondingly, the error sources (i) and (iii)–(v) can be removed by bandpass filtering, as long as the spectral signatures of the respective effects are well separated from the fundamental band. Furthermore, the artifacts arising from effect (ii) can be strongly diminished by a time-resolved version of the GV approach (TF-GV). The TF-GV method then yields for each mode j a trajectory of the vibrational frequency $\omega_j(t|\tau)$ at a time resolution $\tau > \tau_j$, which is only limited by the corresponding oscillation time $\tau_j = 2\pi/\omega_j$ and, thus, is in the femtosecond range. A correlation analysis of these trajectories clearly separates the librational motions from the conformational dynamics of the solvation shells and yields the inhomogeneously broadened MIR spectra, if the theory of motional narrowing is properly included. The MIR spectrum of formaldehyde in solution obtained by TF-GV agrees very well with the FTTCF result, if one applies the so-called ''harmonic approximation'' quantum correction factor and a temperature scaling to the FTTCF intensities. Also for INMA an excellent agreement is achieved if one disregards a slight INMA overestimate of linewidths. © *2004 American Institute of Physics.*
[DOI: 10.1063/1.1822915]

I. INTRODUCTION

Spectroscopy in the midinfrared (MIR) spectral region, which covers frequencies from about 600 to about 4000 cm^{-1}, is an important tool in chemical analysis and biomolecular research because the MIR absorption of molecules is experimentally well accessible.[1–7] Whereas the gas phase spectra are solely determined by the chemical structures and intramolecular force fields of the molecules, in condensed phase these spectra also code the structures of their environments.[3,7–10] For a decoding accurate descriptions of the intramolecular force fields and, in condensed phase, of the interactions between the molecules are required.

Due to the development of density functional theory (DFT) the gas phase IR spectra of molecules can be predicted at a high accuracy.[11] In principle the development of hybrid methods, which combine a DFT description of a molecule with a molecular mechanics (MM) model of the polarizing environment, has opened the way towards comparably accurate descriptions of condensed phase MIR spectra.[12,13] However, the computational protocols for such calculations are still the subject of a current debate.[9,12,14–18] Here, the case of least complexity is the one of molecules dilutely soluted in polar solvents. We will focus the following discussion on this case.

In Paper I of this work[18] we have already discussed the three methods, which have been proposed for the computation of the MIR absorption spectra of molecules in polar solvents from hybrid molecular dynamics (MD) simulations: the instantaneous normal mode analysis (INMA), the Fourier transform of the time correlation function (FTTCF) of the dipole moment, and the so-called ''principal mode analysis'' (PMA). To avoid redundancy we will now rehash only the basic concepts and will concentrate on the specific issues to be addressed in this paper.

When restricting the computation of the MIR absorption to the solute including its polarization by the solvent environment, which is the case considered by us, the three meth-

 © 2004 American Institute of Physics

ods listed above are based on the assumptions that there is no resonance between MIR vibrational modes of the solute and of other molecules and that the polarization induced by the solute in the solvent on the time scale of a MIR vibration is negligible.[18] Note that these assumptions are fulfilled in the generic case of a hybrid MD simulation, in which a single solute molecule is described by DFT and a rigid, nonpolarizable MM model is used for the solvent molecules.[18]

The INMA approach additionally assumes that the solutes vibrations can be decoupled from its rotational and translational motions in its solvent cage and from the dynamics of the solvent for the computation of the vibrational frequencies (cf. Ref. 18 for discussion). Under these assumptions, the MIR spectrum of an isotropically distributed solute is given by[18]

$$\alpha(\omega)=\frac{c_0\omega}{3\hbar}(1-e^{-\hbar\omega/k_BT})\sum_{f,i}\rho_i|\langle f|\vec{d}|i\rangle|^2\delta(\omega-\omega_{fi}), \quad (1)$$

where the velocity of light c and the refractive index n are collected into $c_0=4\pi^2/(cn)$, $\vec{d}$ is the dipole operator of the solute, $|i\rangle$ and $|f\rangle$ are the eigenstates of the molecular vibrational Hamiltonian H_{vib} of the solute in its solvent cage, $\omega_{fi}=(E_f-E_i)/\hbar$ are the associated absorption frequencies, and ρ_i is the population of the state $|i\rangle$.

To calculate the values required in Eq. (1) one proceeds as follows: First, a MD simulation serves to generate an ensemble of solvent cages. Each of these cages is subsequently kept fixed and the vibrational spectrum of the embedded solute molecule is calculated from the DFT/MM force field by the following procedure: According to Refs. 12 and 9 and in contrast to other proposals,[19,20] the potential energy of the solute is first minimized in each cage. Subsequently, the vibrational frequencies ω_j of the solute are calculated from the Hessian. Finally, the derivatives $\partial\vec{d}/\partial a_j$ of the solutes molecular dipole moment $\vec{d}$ with respect to the normal coordinates a_j are calculated. These frequencies and dipole derivatives are different in each solvent cage because the vibrational Hamiltonian H_{vib} contains the interactions of the solute with its respective solvation shell.

Restricting to the fundamentals one can then identify the resulting ensemble of vibrational frequencies ω_j with the ensemble of absorption frequencies ω_{fi}, and can calculate the transition matrix elements $\langle f|\vec{d}|i\rangle$ in linear approximation from the dipole derivatives.[18] In order to account also for anharmonic effects, corrections can be applied to the calculated absorption frequencies and to the transition matrix elements.[11,21,22]

Because INMA completely decouples the solute vibrations from the solvent dynamics for the computation of the vibrational frequencies, it does not capture dynamical effects such as motional narrowing[23–25] or vibrational dephasing,[26] which modify the distribution of the absorption frequencies and, thus, the linewidths.

In the FTTCF approach defined by

$$\alpha(\omega)=\frac{c_0\omega}{3\hbar}(1-e^{-\hbar\omega/k_BT})\frac{1}{2\pi}\int_{-\infty}^{\infty}e^{-i\omega t}\langle\vec{d}(t')\vec{d}(t'+t)\rangle dt \quad (2)$$

the time correlation function of the dipole moment $\vec{d}$ of an isotropically distributed solute is calculated from a DFT/MM-MD trajectory and subsequently Fourier transformed.[18,27] By considering an extended trajectory one has to make sure that the statistics suffices for meaningful results. Because in contrast to the INMA approach, FTTCF does not decouple the solutes eigenstates from the dynamics of its environment, FTTCF includes dynamical effects such as motional narrowing and, therefore, can render exact linewidths. On the other hand, because Eq. (2) has been derived for quantum mechanical ensembles[18,27] and because the time correlation function is obtained from a classical trajectory of nuclear motion, suitable corrections have to be applied to the resulting MIR intensities.[28–32]

As an alternative to the INMA and FTTCF methods Wheeler *et al.* have suggested the so-called PMA.[15] Here, the average mass-weighted normal modes $\langle\vec{c}_j\rangle$ and associated frequencies ω_j of the solute molecule are calculated from a classical MD trajectory by evaluating the eigenvectors and eigenvalues of the covariance matrix associated with the mass-weighted displacements of the molecular coordinates from their equilibrium positions. We have analyzed this suggestion in the preceding paper and have corrected its most severe deficiencies.[18] One of these corrections is based on introducing for each vibrational mode j the so-called ''generalized virial'' (GV) frequency

$$\omega_j^2=\langle\dot{a}_j^2\rangle/\mathcal{V}(a_j), \quad j\in\{7,\ldots,3N\}, \quad (3)$$

which can be calculated from the variance of the trajectory $a_j(t)$ of the associated normal coordinate and from the second moment of its temporal derivative $\dot{a}_j(t)$.[18] The GV expression for the frequency differs from its PMA precursor in that it replaces the global temperature T of the molecule by a mode specific temperature $T_j:=\langle\dot{a}_j^2\rangle/k_B$. It has been derived by considering the virial theorem for an isolated harmonic oscillator. Equation (3) is exact, if (a1) the vibrational molecular Hamiltonian H_{vib} is constant and if (a2) the potential of nuclear motion is harmonic.[18]

In DFT/MM trajectories these assumptions are generally violated. As we have shown,[18] such violations entail serious consequences for the applicability of the GV expression (3) to the computation of frequencies: First, one notes that Eq. (3) neither provides information about the time dependence of the frequencies $\omega_j(t)$ nor about line shapes, because average frequencies are calculated from global moments. Therefore, the GV frequencies cannot account for the fluctuations of the $\omega_j(t)$. But even worse, fluctuations of H_{vib}, which entail fluctuations of (i) the equilibrium coordinates and (ii) the intramolecular force constants lead to strong artificial redshifts of the GV frequencies.[18] Furthermore, if fluctuations of (iii) the normal modes lead to a significant mode mixing in a given molecule, an assignment of average frequencies to average normal modes by Eq. (3) may become

arguable. Moreover, (iv) diagonal anharmonicities and (v) anharmonic mode couplings furnish additional but smaller inaccuracies of the GV frequencies.[18]

Below, we will repeatedly address and more deeply analyze these five error sources with the aim to eliminate the shortcomings of the GV prescription (3) and, thus, to derive an essentially accurate procedure for the computation of average frequencies. The insights gained during this process of error correction will allow us to suggest a modified prescription, by which one can compute vibrational frequencies $\omega_j(t)$ temporally resolved at femtosecond time scales. To verify our arguments and estimate the sizes of the various effects we will apply the corresponding computational procedures to an extended 35 ps DFT/MM trajectory of formaldehyde in water.

II. THEORY

Here we develop a set of methods serving to identify and remove the sources (i)–(v) of errors in GV frequency calculations. As our guideline we use the hypothesis that the dominant sources (i)–(iii) of inaccuracies are due to slow fluctuations of the solvent cages, whose time scales τ_{fluct} are much slower than those given by the MIR frequencies ($\tau_j = 2\pi/\omega_j$) of the solute. In the frequency domain, processes occurring at different time scales occupy well-separated regions. Therefore, one should be able to eliminate the slow fluctuations from the GV frequency computation (3) by Fourier transformations and frequency space filtering. Because FTTCF is based on these transformations, we use a short discussion of its properties for the most simple case of (a1) a constant Hamiltonian and (a2) a harmonic potential to introduce basic concepts and our notation.

A. FTTCF for an isolated molecule with a harmonic potential

Consider a DFT-MD trajectory of an isolated molecule with a constant and harmonic Hamiltonian H_{vib}. In linear approximation the Taylor expansion of the molecular dipole moment $\vec{d}$ with respect to the normal coordinates $\vec{a}$ is

$$\vec{d}(\vec{a}) = \vec{d}(0) + \sum_{j=7}^{3N} \frac{\partial \vec{d}}{\partial a_j} a_j . \tag{4}$$

For harmonic oscillations $a_j(t) = a_{j,0}\sin(\omega_j t + \varphi_j)$ the FTTCF expression (2) becomes

$$\alpha(\omega) = \frac{c_0}{3\hbar\omega}(1 - e^{-\hbar\omega/k_B T}) \sum_{j=7}^{3N} k_B T_j \left(\frac{\partial \vec{d}}{\partial a_j}\right)^2 \delta(\omega - \omega_j), \tag{5}$$

which determines the vibrational frequencies ω_j exactly. In the linear approximation considered here, the derivatives of the dipole moment represent the quantum mechanical transition moments for transitions from the vibrational ground states $|i\rangle$ to the singly excited states $|f\rangle$ of the modes j, that is,

$$(\partial \vec{d}/\partial a_j)^2 = \frac{2\omega_j}{\hbar} |\langle f|\vec{d}|i\rangle|^2 . \tag{6}$$

Since a DFT-MD trajectory is of finite duration L, by Nyquists theorem[33] the accuracy of frequency computation is

$$\delta\omega_j = 1/L \quad \text{for} \quad \omega_j < 1/(2\Delta t), \tag{7}$$

where Δt is size of the integration time step. For example, the duration $L = 10$ ps and the integration time step $\Delta t = 0.5$ fs of the gas phase trajectory analyzed in Sec. III implies a spectral resolution of $\delta\omega_j \approx 3\ \text{cm}^{-1}$ for $\omega_j < 30\,000\ \text{cm}^{-1}$.

B. General discussion of FTTCF

The FTTCF approach (2) exemplified above has two drawbacks, which become particularly relevant in the costly DFT/MM setting. First, vibrational frequencies of modes with vanishing transition moments (6) cannot be detected. Second, the spectrum (2) does not provide any information about the normal modes. However, these drawbacks can be removed by applying Fourier transformations

$$\hat{a}_j(\omega) = \int_{-\infty}^{\infty} a_j(t) e^{-i\omega t} dt \in \mathcal{C} \tag{8}$$

to the trajectories $a_j(t)$ of approximate normal coordinates derived from projecting the MD trajectory to average mass-weighted normal modes $\langle \vec{c}_j \rangle$.[9] The required $\langle \vec{c}_j \rangle$ may be estimated by symmetry considerations[9] or from an ensemble of INMA calculations. The main peaks in the spectra $\hat{a}_j(\omega)$ then determine the frequencies ω_j independently of the transition dipole (6).

The expressions (1) and (2) for the IR spectra of solution systems both rest on the assumption of thermal equilibrium,[27] which is generally not guaranteed in a necessarily short DFT/MM trajectory. By Eq. (5), which we have derived for an isolated molecule ignoring this restriction, the IR absorptions $\alpha(\omega_j)$ of the modes j are proportional to the mode specific temperatures T_j mentioned further above. However, the T_j equals the global temperature T only if the average kinetic energy of the molecule is distributed equally among all modes. Generalization of this observation to the DFT/MM setting, which describes molecules in thermal contact with a heat reservoir, provides now the clue as to how one can correct the errors caused by incomplete equilibration.

If one calculates an MIR spectrum according to Eq. (2) by FTTCF from an insufficiently equilibrated ensemble, within which each mode is excited at a different temperature T_j, then one can extrapolate to the spectrum of an ensemble equilibrated at temperature T by scaling the absorption $\alpha(\omega)$ by $T/T(\omega)$, where $T(\omega)$ is the step function which equals T_j within the interval covering ω_j and ranging halfway to the frequencies ω_{j-1} and ω_{j+1} of the neighboring modes.

C. FTTCF for molecules in polar solvents

In DFT/MM-MD simulations of molecules in polar solvents the assumptions of harmonic (a1) and constant (a2) Hamiltonians H_{vib} are generally violated. For these cases, Fig. 1 schematically sketches a typical spectrum $|\hat{a}_j(\omega)|$ as obtained by Eq. (8) from the trajectory $a_j(t)$ of a given approximate normal coordinate. The spectrum exhibits, besides

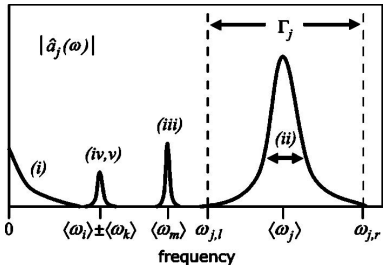

FIG. 1. General scheme as to how the frequency spectrum $\hat{a}_j(\omega)$ of a typical molecular normal coordinate $a_j(t)$ can be affected by slow fluctuations of the (i) equilibrium geometry $\vec{r}^0(t)$, (ii) frequencies $\omega_j(t)$, and (iii) mass-weighted normal modes $\vec{c}_j(t)$, and by (iv,v) an anharmonic potential.

a strong peak at the average fundamental frequency $\langle\omega_j(t)\rangle$, several additional features, whose origins can be related to the effects (i)–(v) introducing errors into the GV frequency computation (3).

The slowly fluctuating shifts (i) of the equilibrium geometry $\vec{r}^0(t)$ along the average mass-weighted normal mode $\langle\vec{c}_j\rangle$ cause the low-frequency ($\omega\to 0$) signals shown in Fig. 1 and marked by (i). The slow fluctuations (ii) of the force constants entail corresponding fluctuations of the vibrational frequencies and, therefore, a broadening of the strong peak centered around the average frequency $\langle\omega_j(t)\rangle$. Thus, the shape of this peak represents the distribution of vibrational frequencies $\omega_j(t)$ sampled by the DFT/MM trajectory.

The slow fluctuations (iii) of the mass-weighted normal modes $\vec{c}_j(t)$ around their average values may have the following effects: Other modes $\vec{c}_m(t)$ may acquire components into the direction of the given and suitably estimated average mode $\langle\vec{c}_j\rangle$ implying that these vibrations are also projected onto the corresponding approximate normal coordinate a_j. Figure 1 shows a smaller peak at $\langle\omega_m\rangle$ belonging to such a mode m. If the fluctuations of the mass-weighted normal modes $\vec{c}_m(t)$ are very strong, the spectrum $|\hat{a}_j(\omega)|$ will contain many sizable signals of other modes and the association of a frequency to mode j becomes impossible. Moreover, in such cases of a strong solvent-induced mode mixing, reasonable estimates for average mass-weighted normal modes $\langle\vec{c}_j\rangle$ cannot be given.

If the potential of classical nuclear motion contains (iv) diagonal anharmonicities, the spectrum $|\hat{a}_j(\omega)|$ will exhibit peaks at the higher harmonics, that is, at $2\langle\omega_j\rangle, 3\langle\omega_j\rangle, \ldots$. Finally, in the case (v) of nonlinear couplings between modes i, j, and k the spectrum $|\hat{a}_j(\omega)|$ will show peaks at the sum or difference frequencies $\langle\omega_i\rangle\pm\langle\omega_k\rangle$ as indicated in Fig. 1.

Summarizing we can state that each of the effects (i)–(v) leaves a characteristic finger print in the spectra $|\hat{a}_j(\omega)|$. Therefore, an inspection of such spectra enables a precise judgement about the extent to which these effects contribute to the various normal coordinates. Correspondingly one can estimate the size, by which a given DFT/MM trajectory violates the assumptions of a constant (a1) and harmonic (a2) Hamiltonian that are basic for the computation (3) of the GV frequencies. But more importantly, the existence of the fingerprints immediately suggests as to how one can eliminate the sources (i)–(v) of errors hampering that computation.

D. GV frequencies from bandpass filtering

Figure 1 additionally visualizes the key idea for the error elimination announced above. Assuming that the spectral features, which are associated with the sources (i) and (iii)–(v) of inaccuracies in the GV frequency computation, are well separated from the fundamental peak, the application of a bandpass filter centered around $\langle\omega_j\rangle$ to the spectrum (8) promises to eliminate these sources. Analytically, one constructs a truncated spectrum

$$\hat{a}_j^*(\omega)=\begin{cases}\hat{a}_j(\omega) & \text{for} \quad \omega\in\Gamma_j,\\ 0 & \text{else,}\end{cases} \tag{9}$$

where the frequency interval Γ_j should cover all fluctuations of the fundamental $\omega_j(t)$. Fourier back transformation of $\hat{a}_j^*(\omega)$ then yields filtered normal coordinates $a_j^*(t)$. One can expect that the bandpass filtered GV frequencies (F-GV)

$$\omega_j^{*2}=\langle\dot{a}_j^{*2}\rangle/\mathcal{V}(a_j^*), \quad j\in\{7,\ldots,3N\}, \tag{10}$$

are much more accurate than the GV frequencies (3) calculated from the original normal coordinate trajectories $a_j(t)$, because the artificial redshifts[18] caused by (i) slow equilibrium geometry fluctuations and the inaccuracies[18] generated by (iv) diagonal anharmonicities are always filtered, whereas the effects (iii) and (v) are removed to the extent that the frequencies $\langle\omega_m\rangle$ and $\langle\omega_i\rangle\pm\langle\omega_k\rangle$, respectively, do not fall into the frequency interval Γ_j. Because the vibrational spectra of small molecules generally exhibit well separated bands, the case $\langle\omega_m\rangle\in\Gamma_j$ will be a rare exception. For larger molecules with relatively dense spectra, however, the bandpass filtering can remove only those mode coupling artifacts that are due to spectrally distant modes. Further below, using DFT/MM sample calculations, we will check these expectations and the size of the improvements introduced by bandpass filtering.

The slow fluctuations (ii) of the force constants, however, are retained by the filtering (9). Therefore, the associated artificial redshifts of the GV frequencies[18] remain in the F-GV frequencies (10) and we need an additional method, which can remove also these errors.

E. Time-local GVF

According to our basic hypothesis the fluctuations of the force constants proceed on time scales τ_{fluct}, which are larger than the oscillation periods $\tau_j=2\pi/\omega_j$ of typical MIR vibrations. Therefore, if one can calculate the F-GV or GV frequencies by suitable moving averages on time scales $\tau\approx\tau_j$ the errors caused by the force constant fluctuations should become negligible. Moreover, such a procedure will immediately render time-resolved frequencies $\omega_j(t|\tau)$ at a temporal resolution defined by τ.

To proceed along these lines we introduce the convolution

$$\langle A\rangle_\tau(t):=A^*g_\tau=\int_{-\infty}^{\infty}A(t')g_\tau(t'-t)dt' \tag{11}$$

of a time series $A(t)$ with convolution kernels $g_\tau(t)$, which are localized around $t=0$, have a width τ, and are normalized according to

$$\int_{-\infty}^{\infty} g_\tau(t)dt=1. \tag{12}$$

By means of the convolution (11) one can now calculate the moments of the normal coordinates required in the F-GV prescription (10) in a time-resolved fashion yielding, e.g., the time-resolved variance

$$\mathcal{V}_\tau[a_j^*](t) \coloneqq \langle a_j^{*2}\rangle_\tau(t)-\langle a_j^*\rangle_\tau^2(t) \tag{13}$$

of the normal coordinate $a_j^*(t)$. As a result we obtain the time-resolved F-GV frequencies (TF-GV)

$$\omega_j^2(t|\tau) \coloneqq \frac{\langle \dot{a}_j^{*2}\rangle_\tau(t)}{\mathcal{V}_\tau[a_j^*](t)}, \tag{14}$$

which reduce to the F-GV frequencies (10) in the limit $\tau\to\infty$.

For $\tau<\tau_j$, the moments appearing in Eq. (14) strongly depend on the phase of the oscillating normal coordinate $a_j^*(t)$. Correspondingly, $\omega_j^2(t|\tau<\tau_j)$ does not represent a reasonable frequency estimate and, therefore, τ_j sets a lower limit to the time resolution τ. To achieve a time resolution close to this lower limit, Gaussian convolution kernels

$$g_\tau(t)=\frac{1}{\sqrt{2\pi\tau^2}}\exp\left(-\frac{t^2}{2\tau^2}\right) \tag{15}$$

should be used, because they render very smooth averages and are scarcely sensitive to the phase of the oscillation for $\tau\geqslant\tau_j$. For such a kernel, Eq. (14) determines the frequency of a harmonic oscillation $a^*(t)=a^*(0)\cos\omega_j t$ at an accuracy given by $\exp(-\omega_j^2\tau^2)$, which is quite accurate even for an averaging time $\tau=\tau_j$ at the lower limit. This averaging time is, for instance, about 20 fs for a typical MIR vibration at about 1700 cm^{-1}.

F. Motional narrowing of TF-GV frequencies

Now the question arises, how one can derive the linewidths $\Delta\omega_j$ from these time series. This question can be approximately answered in the framework of simple stochastic models for random frequency modulations outlined, e.g., in Ref. 24. Adopting these concepts we assume that the TF-GV frequency of a mode j changes in time as

$$\omega_j(t|\tau)=\omega_j+\delta\omega_j(t|\tau), \tag{16}$$

where the fluctuation $\delta\omega_j(t|\tau)$ is a stationary stochastic process. We furthermore assume that this process is Gaussian with an average $\langle\delta\omega_j\rangle=0$ and with a correlation function defined by

$$\langle\delta\omega_j(t'|\tau)\,\delta\omega_j(t'+t|\tau)\rangle=\sigma_j^2(\tau)C_{\omega,j}(t|\tau), \tag{17}$$

where

$$\sigma_j^2(\tau)=\langle\delta\omega_j^2\rangle \tag{18}$$

measures the magnitude of the random frequency modulation. Then the line shape $I_j(\omega)$ of mode j is given by[24]

$$I_j(\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}dt\exp\left[-i\omega t-\sigma_j^2(\tau)\times\int_0^t(t-t')C_{\omega,j}(t'|\tau)dt'\right], \tag{19}$$

where the frequency is measured from the center ω_j. If the correlation function (17) decays like a simple exponential

$$C_{\omega,j}(t|\tau)=e^{-t/\tau_c}, \tag{20}$$

where the correlation time τ_c measures the rate of the frequency modulation, the line shape (19) can be further evaluated and two limiting cases can be distinguished. Depending on the value of the parameter

$$\alpha_j\equiv\tau_c\sigma_j(\tau), \tag{21}$$

Eq. (19) either reduces to a Gaussian of standard deviation $\sigma_j(\tau)$ for $\alpha_j>1$ or to a Lorentzian of width $\sigma_j(\tau)\alpha_j$ for $\alpha_j<1$. As is apparent from the definition (21), a Gaussian line shape is obtained in the limit of strong and slow frequency modulations, whereas a Lorentzian corresponds to weak and fast perturbations.[24]

Although one cannot expect that the frequency modulations of a molecule in a complex polar solvent are caused by Gaussian noise with a single exponential decay, the above considerations can provide guidelines for approximately deriving line shapes from the frequency autocorrelation functions $C_{\omega,j}(t|\tau)$ of the calculated TF-GV frequencies $\omega_j(t|\tau)$.

III. RESULTS AND DISCUSSION

To test and evaluate the F-GV and TF-GV methods of frequency computation introduced above we will use the DFT-MD and DFT/MM-MD trajectories of a FA molecule in the gas phase and in aqueous solution, respectively, which we have presented in Paper I of this work.[18] In the quoted paper we have already applied the (I)NMA and GV procedures (but not yet FTTCF, F-GV, and TF-GV) to describe the six vibrational modes j of FA, that is the asymmetric ($j=12$, $\mathrm{CH}^{\mathrm{asy}}$) and the symmetric ($j=11$, $\mathrm{CH}^{\mathrm{sym}}$) hydrogen stretches, the carbonyl stretch ($j=10$, CO), the symmetric ($j=9$, HCH), and asymmetric ($j=8$, HCO) in-plane bending modes and the out-of-plane mode ($j=7$, oop). In that study, the analysis of the gas phase DFT trajectory mainly served us to estimate the effect, which the anharmonicities (iv) and (v) have on the GV frequencies. Concerning the DFT/MM trajectory of FA in water, the observed artificial redshifts of the GV frequencies were attributed by us to the effects (i)–(iii).[18]

A. Details of methods

FTTCF. For a detailed description of the MD simulations we refer to Paper I of this work.[18] By Eq. (7) the durations of the trajectories of 10 ps in the gas phase and 35 ps in solution imply resolutions of 3 and 0.9 cm^{-1}, respectively, for the FTTCF spectra. Applying the temperature scaling (cf. Sec. II B) to the FTTCF expression (2) yields

$$\alpha_{\mathrm{FTTCF}}(\omega)=\frac{c_0\omega}{6\pi\hbar}\frac{T}{T(\omega)}\int_{-\infty}^{\infty}e^{-i\omega t}\langle\vec{d}(t')\vec{d}(t'+t)\rangle dt. \quad (22)$$

Here the factor $(1-e^{-\hbar\omega/k_BT})$ occurring in Eq. (2) has been neglected, because we have $\hbar\omega\gg k_BT$ in the considered MIR frequency range $\omega>1000\ \mathrm{cm}^{-1}$.

INMA. In order to model the absorption spectrum (1) we have calculated the average ω_j and the standard deviation σ_j of the vibrational frequency of each mode from the INMA snapshots described in Ref. 18. Using Eq. (6), second order statistics of the distribution of vibrational frequencies, and neglecting terms of the order $e^{-\hbar\omega/k_BT}$, Eq. (1) becomes

$$\alpha_{\mathrm{INMA}}(\omega)=\frac{c_0}{6}\sum_j\left\langle\left(\frac{\partial\vec{d}}{\partial a_j}\right)^2\right\rangle\frac{\exp\left[-\frac{(\omega-\omega_j)^2}{2\sigma_j^2}\right]}{\sqrt{2\pi\sigma_j^2}}. \quad (23)$$

The average dipole derivatives required for the INMA-MIR spectrum of FA are listed in Table I of Ref. 18.

INM. Following a different proposal,[19,20] we have also calculated the vibrational frequencies of the FA molecule from its Hessian in the frozen solvent cages without prior energy minimization. We will denote the corresponding results by INM.

GV, F-GV, and TF-GV. The mass-weighted normal modes $\langle\vec{c}_j\rangle$ needed for the computation of the normal mode trajectories $a_j(t)$ were obtained from the NMA and from averages over the INMA results in the gas phase and solution cases, respectively. The frequency intervals Γ_j defining the bandpass filters for the F-GV frequency calculation were determined by visual inspection of the normal mode spectra $|\hat{a}_j(\omega)|$.

For the analysis of the trajectory of the vibrational frequencies, the time-resolved TF-GV frequencies were calculated by applying Gaussian kernels (15) of widths $\tau=2\pi/\omega_j$, which cover one vibrational period for each mode j, to the bandpass filtered normal mode trajectories $a_j^*(t)$. Here, the ω_j resulting from the F-GV calculation were used. To determine and interpret the time scales characteristic for the fluctuations of the TF-GV frequencies $\omega_j(t|\tau)$ we have calculated the average

$$C_\omega(t)=\frac{1}{6}\sum_{j=7}^{12}C_{\omega,j}(t|\tau) \quad (24)$$

of the autocorrelation functions (17).

Analogously to Eq. (23), that is, using Eq. (6) and neglecting terms of the order $e^{-\hbar\omega/k_BT}$, we have constructed the absorption spectrum (1) according to

$$\alpha_{\mathrm{TF\text{-}GV}}(\omega)=\frac{c_0}{6}\sum_j\left(\frac{\partial\vec{d}}{\partial a_j}\right)^2 p_j(\omega). \quad (25)$$

Here, the dipole derivatives are determined from the trajectories $\vec{d}(t)$ and $\vec{a}(t)$ by[18]

$$\left(\frac{\partial\vec{d}}{\partial a_j}\right)^2\approx\left\langle\left(\frac{\Delta\vec{d}(t)}{\Delta a_j(t)}\right)\right\rangle^2 \quad (26)$$

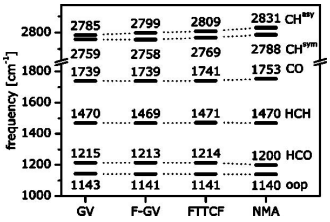

FIG. 2. Vibrational frequencies of FA in the gas phase, calculated from our 10 ps DFT trajectory with GV (3), F-GV (10), FTTCF (8), and NMA [cf. Paper I of this work (Ref. 18)].

and the distribution $p_j(\omega)$ of absorption frequencies is modeled by a histogram $p_j(\omega_j^*)$ of the TF-GV frequencies (14) calculated at an averaging time scale $\tau=240$ fs.

B. Band positions calculated by the various methods

In the gas phase case, which we will consider first, quite accurate MIR spectra can be efficiently calculated by combining DFT and NMA. Anharmonic perturbation theory can then enhance the accuracy of results.[21] Corresponding reference calculations on the FA molecule[11,18] will now allow us to check in detail, to what extent FTTCF of a DFT gas phase trajectory can account for the anharmonicities neglected by NMA and to what extent the anharmonicity artifacts (iv) and (v) can be removed from the GV frequencies by bandpass filtering (F-GV).

1. Gas phase spectra of FA

The last two columns of Fig. 2 compare the NMA results from Ref. 18 with the FTTCF frequencies obtained from the 10 ps DFT trajectory. In the low-frequency part of the spectrum, which is below $1800\ \mathrm{cm}^{-1}$, the NMA and FTTCF frequencies differ only by a root mean square deviation (rmsd) of $9\ \mathrm{cm}^{-1}$, whereas in the high-frequency part covering the CH stretches the anharmonicities derived from our classical description of nuclear motion shift the FTTCF frequencies on the average by $21\ \mathrm{cm}^{-1}$ to the red of the harmonic NMA results.

In contrast to these findings, both experiments[34] and perturbation theoretical calculations[11] accounting for the quantum nature of nuclear motion have found large anharmonic redshifts of $170\ \mathrm{cm}^{-1}$ and $30\ \mathrm{cm}^{-1}$ for the high-frequency modes and the low-frequency modes, respectively. The thus determined FTTCF underestimate of the anharmonic shifts is most probably due to the fact that the excitation of a molecule by an IR quantum $\hbar\omega_j$ leads to much larger oscillation amplitudes than the ones sampled at $T\approx340$ K by our DFT trajectory ($k_BT\ll\hbar\omega_j$), within which the nuclear motions were described by classical mechanics. Correspondingly, a much smaller range of the anharmonic potential surface has been explored as compared to the experimental situation.

This explanation leads to the general prediction that anharmonic quantum effects will be sizably underestimated by all methods (such as, e.g., GV or FTTCF), which derive vibrational frequencies from classical room temperature DFT(/MM) trajectories. Our data generally support this pre-

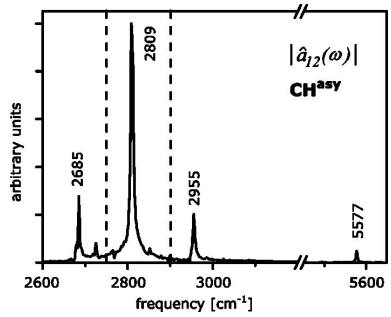

FIG. 3. Fourier transform $|\hat{a}_{12}(\omega)|$ of the normal coordinate $a_{12}(t)$ of the asymmetric CH stretching mode as obtained from the gas phase DFT trajectory of formaldehyde.

diction as one can see from the rough agreement of the GV and FTTCF frequencies displayed in the first and third columns of Fig. 2.

One may, however, try to remedy this underestimate of the anharmonic shift by the following consideration: If one assumes that the anharmonic shifts depend linearly upon the mode specific temperature T_j, as suggested for the classical one-dimensional anharmonic oscillator analyzed in Paper I of this work,[18] one can try extrapolations. For instance, the CO mode had been excited at a temperature $T_{CO}=747$ K during the gas phase trajectory. The 14 cm^{-1} (GV) and 12 cm^{-1} (FTTCF) anharmonic shifts observed for this mode extrapolate to 47 and 40 cm^{-1}, respectively, at the temperature of $T_{CO}=2522$ K corresponding to the energy of an IR quantum at the NMA frequency $\omega_{CO}=1753$ cm^{-1}. However, these extrapolated shifts sizably overestimate the anharmonic shift experimentally observed[34] at 32 cm^{-1} and calculated[11] at 26 cm^{-1}. This indicates that such a linear extrapolation of anharmonic shifts in classical oscillators cannot capture the complex problem of high-dimensional anharmonic quantum mechanical vibrations too well.

Upon closer comparison of columns one and three in Fig. 2, one observes an average 17 cm^{-1} underestimate of the GV frequencies with respect to the FTTCF frequencies in the high-frequency region. If the arguments outlined in Sec. II are correct, these differences should be caused by the anharmonic artifacts (iv) and (v) hampering the GV approach. These artifacts should be visible in the Fourier transforms of the normal coordinate trajectories (cf. Sec. II C) and should vanish by applying the F-GV extension.

To provide an example for the spectral fingerprints of the anharmonic artifacts (iv) and (v), Fig. 3 shows the spectrum $|\hat{a}_{12}(\omega)|$ obtained from the projection of the DFT trajectory onto the NMA mass-weighted normal mode $\vec{c}_{12}$ of the asymmetric CH stretch. At 2809 cm^{-1} the spectrum exhibits the expected large peak representing the fundamental frequency. The two additional peaks at 2685 and 2955 cm^{-1}, respectively, are due to nonlinear mode mixing (v) and correspond to the sums $\omega_8+\omega_9$ and $\omega_8+\omega_{10}$, respectively. The small peak at 5577 cm^{-1} represents the second harmonic of ω_{12} and signifies the diagonal anharmonicities (iv). These additional peaks are located outside the frequency interval Γ_{12} indicated by the dashed lines in Fig. 3. The associated anharmonicity effects are removed in the inverse Fourier transform $a_{12}^*(t)$ of the truncated spectrum $\hat{a}_{12}^*(\omega)$ located within Γ_{12}. If they represent the cause for the differences noted above, the F-GV frequency ω_{12} should be much closer to the FTTCF result than the error-prone GV frequency.

The inspection of the CH^{asy} frequencies in the level-diagram depicted in Fig. 2 confirms this expectation. Whereas the GV procedure underestimates the FTTCF value by 24 cm^{-1}, the bandpass filtering diminishes this underestimate by 14 cm^{-1}. In the spectrum $|\hat{a}_7(\omega)|$ of the HCO bending mode (not shown here) one finds similar signatures of a nonlinear mode mixing (v), although in this case GV agrees with FTTCF and the result does not change significantly upon removing the anharmonicities (F-GV). This invariance implies that in the HCO case the various GV artifacts of type (v) mutually cancel. In contrast, all anharmonicity fingerprints are much weaker in the spectra $|\hat{a}_j(\omega)|$ of the remaining modes explaining why the corresponding F-GV and GV frequencies are very close.

In the high-frequency CH stretching part of the spectrum, 10 cm^{-1} differences between the F-GV and FTTCF frequencies remain. As shown in the Appendix, the Verlet algorithm,[35] which has been applied to the numerical integration of the Newtonian equations of motion, generates artificially high vibrational frequencies for harmonic oscillators. As exhibited by Eq. (31) the relative frequency overestimate scales with $(\omega\Delta t)^2$. At the given time step $\Delta t=0.5$ fs one therefore expects a 12 cm^{-1} Verlet overestimate for the FTTCF frequencies of the CH stretches and overestimates smaller than 3 cm^{-1} in the low-frequency region. Interestingly, this Verlet overestimate does not apply to GV frequencies due to an accidental cancellation of errors also proven in the Appendix. Therefore and in contrast to the FTTCF results, the F-GV frequencies in Fig. 2 are unaffected by the Verlet artifact. In view of the 3 cm^{-1} spectral resolution of the Fourier transformation, this artifact hampering FTTCF quantitatively explains the remaining deviations between F-GV and FTTCF observed in Fig. 2.

As a result we may state that our study of the gas phase spectra has demonstrated that the bandpass filtering approach can identify and remove the anharmonic artifacts (iv) and (v) impeding the GV frequency computation. Furthermore, it has provided a hint towards an unexpected Verlet artifact affecting FTTCF.

2. Vibrational frequencies of FA in solution

Turning to the solution spectrum of FA we now want to check, to what extent FTTCF accounts for the anharmonicities neglected by INMA in solution and to what extent also the artifacts (i)–(iii) induced by the fluctuating molecular vibrational Hamiltonian H_{vib} can be removed from the GV frequencies by bandpass filtering (F-GV) and time resolution (TF-GV).

The last two columns of Fig. 4 compare the INMA results[18] with the FTTCF frequencies. From the rmsd of only 9 cm^{-1} between the FTTCF and the harmonic INMA frequencies, one concludes that the classical anharmonic effects are even weaker in the solution trajectory than in the gas

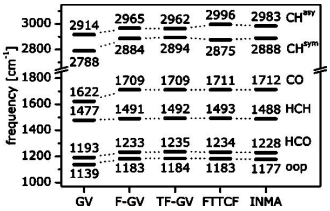

FIG. 4. Vibrational frequencies of FA in aqueous solution, calculated from our 35 ps DFT/MM trajectory with GV (3), F-GV (10), TF-GV (14), FTTCF (8), and INMA [cf. Paper I of this work (Ref. 18)].

phase trajectory analyzed above. Hence, the strong 66 cm^{-1} rmsd between the GV and FTTCF frequencies listed in the first and fourth columns of Fig. 4, respectively, must be mainly due to the artifacts (i)–(iii) induced by the fluctuating Hamiltonian H_{vib}.

As an example as to how one can identify and remove these artifacts by the bandpass filtering approach (cf. Secs. II C and II D), Fig. 5 shows the spectrum $|\hat{a}_{10}(\omega)|$ obtained from the projection of the DFT/MM trajectory onto the average INMA mass-weighted normal mode $\langle \vec{c}_{10} \rangle$ of the CO stretch.

At 1711 cm^{-1} the spectrum $|\hat{a}_{10}(\omega)|$ exhibits the expected large peak representing the CO fundamental frequency, which is significantly broadened as compared, e.g., with the gas phase peak in Fig. 3 (note the different frequency scales in the two figures). According to Sec. II C, this broadening is due to fluctuations (ii) of the CO force constant. In addition, one sees a broad, sizable, and noisy signal towards low frequencies, which by Sec. II C indicates slow fluctuations (i) of the equilibrium geometry $\vec{r}^{\,0}(t)$ along the approximate normal coordinate a_{10}. Furthermore, the spectrum $|\hat{a}_{10}(\omega)|$ exhibits a small maximum at the fundamental frequency $\omega_7 = 1180$ cm^{-1} of the oop mode, which is caused by the solvent-induced mode mixing and fluctuation effects (iii) discussed in Sec. II C. Conversely, in the spectrum $|\hat{a}_7(\omega)|$ of the oop mode (not shown) a corresponding small maximum is found at the fundamental frequency $\omega_{10} = 1711$ cm^{-1} of the CO mode. Very weak additional peaks occur at 3412, 3210, and 2983 cm^{-1}, respectively. They likely correspond to the CO second harmonic $2\omega_{10}$, to the sum frequency $\omega_{10}+\omega_9$, and to the HCH second harmonic $2\omega_9$, respectively, and, therefore, indicate the weak presence of the anharmonic effects (iv) and (v).

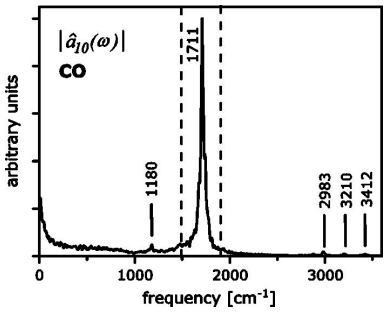

FIG. 5. Fourier transform $|\hat{a}_{10}(\omega)|$ of the normal coordinate $a_{10}(t)$ of the CO stretching mode as obtained from the DFT/MM trajectory of FA in solution. For noise reduction the spectrum has been smoothened with a Gaussian convolution kernel of width 5 cm^{-1}.

The additional peaks in Fig. 5 are all located well outside the interval Γ_{10} indicated by the dashed lines. Also the main part of the low-frequency noise is found outside Γ_{10}. Equivalent observations can be made by considering the spectra $|\hat{a}_j(\omega)|$ of all other low-frequency modes (data not shown). Hence, one expects that filtering significantly diminishes the large difference between the corresponding GV and FTTCF frequencies.

By comparing columns one, two, and four in Fig. 4, one indeed observes that the rmsd of 54 cm^{-1} between GV and FTTCF frequencies in the low-frequency part shrinks to a rmsd of only 2 cm^{-1} between the F-GV and FTTCF frequencies. This shows that the large deviations between the GV frequencies and the FTTCF frequencies are mainly due to the GV artifacts (i), (iii)–(v), which are removed by bandpass filtering.

For the two high-frequency CH stretches the bandpass filtering is more difficult, because in the corresponding spectra $|\hat{a}_{11}(\omega)|$ and $|\hat{a}_{12}(\omega)|$ (not shown), several signals from nonlinear mode mixing overlap with the strongly broadened fundamental peaks. Nevertheless, by our choice of the intervals Γ_j limited by the minima surrounding the fundamental peaks a significant improvement is achieved. According to Fig. 4 the rmsd of 84 cm^{-1} between GV and FTTCF shrinks to 23 cm^{-1} for F-GV and FTTCF.

The excellent agreement between the F-GV and FTTCF frequencies in the low-frequency part of the spectrum suggests that the remaining GV artifacts (ii) induced by the fluctuating force constants are negligible. This conclusion is confirmed by the average TF-GV frequencies, which were designed to remove these artifacts and are shown in the third column of Fig. 4. The TF-GV and F-GV frequencies are almost identical in the whole spectral range. Thus, also in the given case the differences concerning the FTTCF and (T)F-GV descriptions of the CH frequencies must be attributed to a too coarse time step Δt of integration and to anharmonicity artifacts (v), which are not completely removed by bandpass filtering.

C. Trajectories of vibrational frequencies

Figure 6 shows a typical trajectory of a time-resolved TF-GV frequency $\omega_j(t|\tau)$ using the CO mode as an example. Here, the time resolution is $\tau = 20$ fs. To make slower fluctuations underlying the strong and fast fluctuations in that trajectory more clearly visible, the figure also shows the TF-GV frequency $\omega_j(t|\tau)$ of the CO mode calculated at an averaging time scale $\tau = 240$ fs. This width has been estimated as follows.

To obtain an estimate on the time scales of the frequency fluctuations consider the solid line in Fig. 7, which represents the average frequency autocorrelation function $C_\omega(t)$ defined by Eq. (24). In the depicted range this function can be reasonably approximated by a bi-exponential fit, which yields decay times $\tau_1 \approx 79$ fs and $\tau_2 \approx 240$ fs. More slowly

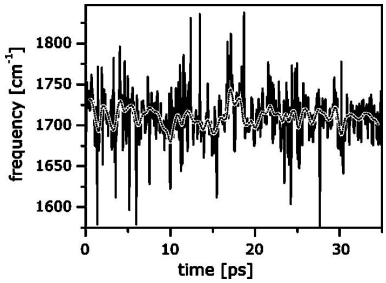

FIG. 6. Trajectory of the time-resolved TF-GV frequency $\omega_{10}(t|\tau)$ (14) of the CO mode of FA in solution calculated at averaging time scales $\tau=20$ fs (black solid line) and $\tau=240$ fs (black dotted line).

TABLE I. Half-widths of the FTTCF bands and standard deviations of the INMA, INM, and TF-GV frequencies. The TF-GV frequencies were calculated at averaging time scales of $\tau=2\pi/\omega_j$, $\tau=80$ fs, and $\tau=240$ fs, respectively.

	Linewidths (cm^{-1})					
				TF-GV		
	FTTCF	INM	INMA	$\sigma_{2\pi}$	σ_{80}	σ_{240}
CH^{asy}	28	118	36	60	38	29
CH^{sym}	23	125	27	66	33	25
CO	10	51	16	28	17	10
HCH	7	19	11	15	9	5
HCO	11	27	13	24	16	9
oop	10	31	15	25	16	9

decaying correlations cannot be detected due to the insufficient statistics provided by our short (35 ps) trajectory.

Figure 7 additionally displays the autocorrelation functions $C_{\vec{v}}(t)$ of the velocity $\vec{v}(t)$ of the center of mass and $C_{\vec{n}}(t)$ of the orientation $\vec{n}(t)$ of the molecule. Simple exponential fits yield the decay times $\tau_{\vec{v}}\approx 63$ fs and $\tau_{\vec{n}}\approx 700$ fs, respectively. These values are by a factor of 2 shorter than those determined by previous MM-MD simulations, which, however, have also shown that these values are extremely sensitive to the applied force field.[36,37]

Thus, $\tau_1\approx\tau_{\vec{v}}$ showing that the decay of the velocity autocorrelation is reflected in the fast component of the frequency autocorrelation decay. Similarly, in view of the more limited statistics concerning slower processes also τ_2 may be considered to be of the same order of magnitude as $\tau_{\vec{n}}$ implying that also the decay of orientational correlations shows up in the frequencies. Since the decays of $C_{\vec{v}}(t)$ and $C_{\vec{n}}(t)$ are caused by collisions of FA with the water molecules in the hydration shell, its frequencies thus represent sensors for these fast processes.

By the choice of the 240 fs averaging time scale for the TF-GV frequencies, the fluctuations associated with these fast processes are effectively filtered. As demonstrated by the corresponding trajectory in Fig. 6, slower fluctuations remain which occur at the time scale of a few picoseconds. This time scale is characteristic for conformational transitions within the hydration shell of C═O groups.[9] Therefore, we assign the slower fluctuations of the CO vibrational frequencies to rearrangements in the hydration shell.

By averaging over the fast fluctuations the standard deviation of the frequency decreases from 28 to 10 cm^{-1}. Therefore, the largest part of the CO frequency modulations is caused by environmental processes relaxing faster than 0.25 ps. Further below this observation will be of importance in our comparative discussion of linewidths derived by the INMA and INM methods.

Here we finally want to estimate the GV errors of type (ii), which are caused by fluctuations of the force constant $\gamma(t)$ and remain in the F-GV frequencies after bandpass filtering. For the CO mode the standard deviation σ_γ of the force constant can be estimated from the standard deviation $\sigma=28$ cm^{-1} of the TF-GV frequency $\omega(t|\tau=20$ fs) by $\sigma_\gamma/\langle\gamma\rangle\approx 2\sigma/\langle\omega\rangle\approx 3.3\%$. The GV frequency error of type (ii) then is[18] $|\delta\omega|=\omega\sigma_\gamma^2/4\gamma^2$, which is below 1 cm^{-1} in the given case. This estimate corroborates the conclusion drawn in connection with Fig. 4 that the GV errors of type (ii) remaining in the F-GV frequencies are very small. Therefore, the F-GV approach yields quite accurate frequencies.

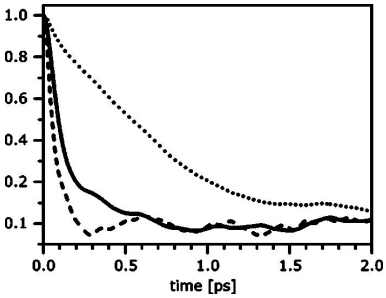

FIG. 7. Various autocorrelation functions calculated from the MD trajectory of FA in solution. Solid line: average (24) for the frequencies (14); dotted line: $C_{\vec{v}}(t)$ for the FA velocity; dashed line: $C_{\vec{n}}(t)$ for the FA orientation.

D. Linewidths

Table I lists the linewidths of the various modes as obtained by FTTCF, INMA, INM, and TF-GV, respectively. The data labeled by FTTCF are the half-widths at half-maximum extracted from the spectrum (22), whereas the INMA, INM, and TF-GV values are the standard deviations of the corresponding frequencies. The TF-GV results were calculated at averaging time scales of $\tau=2\pi/\omega_j$, $\tau=80$ fs, and $\tau=240$ fs, respectively. We will take the FTTCF results as our reference, since this method automatically accounts for dynamical effects such as, e.g., motional narrowing.[25]

According to Table I the FTTCF linewidths are seriously overestimated by INM. This error is due to the neglect of motional narrowing characteristic for INM. In contrast, the INMA linewidths agree much better with FTTCF. For an explanation recall that INMA, as opposed to INM, minimizes the energy of the molecule in the solvent cages before calculating Hessians and thus allows the molecule to escape from strong collisional interactions. According to our above analysis of the TF-GV frequency fluctuations, such an escape can significantly reduce the size of the fluctuations. As

shown by Table I, the INMA reduction of fluctuations partially repairs the INM overestimate of the FTTCF linewidths. This correction is responsible for the good agreement between INMA and FTTCF or experimental linewidths noted in previous DFT/MM studies.[9,14]

Let us now turn to the calculation of linewidths within TF-GV. Here, in principle Eq. (19) represents a prescription as to how one can calculate the line shape $I_j(\omega)$ for each mode j individually from the autocorrelation function $C_{\omega,j}$ (17) of the corresponding TF-GV frequency. Unfortunately our DFT/MM-MD trajectory does not provide a sufficient statistical basis for these sensitive calculations. Therefore, we have chosen a more approximate approach to calculate TF-GV linewidths, which relies on the time scales $\tau_1 \approx 80$ fs, $\tau_2 \approx 240$ fs, and $\tau_3 \approx 2$ ps of frequency fluctuations identified by our analysis of the mode-averaged autocorrelation function C_ω (24). Recall that we have attributed these three time scales to translational collisions (TC), to rotational collisions (RC), and to the solvents conformational dynamics (CD), respectively,

By choosing the times scales $\tau = 2\pi/\omega_j$, $\tau = 80$ fs, $\tau = 240$ fs, and $\tau = \infty$ in the TF-GV convolution kernel (15), one expects that frequency fluctuations faster than τ are effectively filtered from the respective TF-GV time series $\omega_j(t|\tau)$. Therefore, the associated TF-GV standard deviations $\sigma_{2\pi}$, σ_{80}, and σ_{240}, which are listed in Table I, must be due to environmental processes occurring at larger time scales (by definition $\sigma_\infty = 0$). If the three stochastic processes TC, RC, and CD are statistically independent, then the differences $\sigma_{TC} = \sigma_{2\pi} - \sigma_{80}$, $\sigma_{RC} = \sigma_{80} - \sigma_{240}$, and $\sigma_{CD} = \sigma_{240} - \sigma_\infty$ represent estimates for their respective contributions to the short-time standard deviations $\sigma_{2\pi}$.

By Eq. (21) these estimates for the magnitudes $\sigma_j(\tau)$ of random frequency modulations [cf. Eq. (18)], which follow from Table I for each mode, and the associated time scales τ_1, τ_2, and τ_3 allow us to estimate the parameter α for each mode and for each process. One obtains $\alpha_{TC} \approx 0.42$ for the high-frequency modes and $\alpha_{TC} \approx 0.13$ for the low-frequency modes. According to these values the TC fluctuations are in the motional narrowing limit and, therefore, scarcely contribute to the total linewidths. Because α_{RC} is about 0.33 for all modes, the RC fluctuations behave similarly. In contrast, for the CD fluctuations one obtains $\alpha_{CD} \approx 10$ at high frequencies and $\alpha_{CD} \approx 3$ at low frequencies. These values indicate that the slow and strong CD fluctuations are well outside the motional narrowing regime. Therefore, one can consider the contributions $\sigma_{CD} = \sigma_{240}$ of the CD fluctuations as reasonable estimates for the linewidths.

Indeed, Table I displays a good agreement of the TF-GV standard deviations σ_{240} with the FTTCF linewidths. Note that Fig. 6 shows for the CO mode the corresponding time series of TF-GV frequencies $\omega_j(t|240\text{ fs})$. Because the standard deviations $\sigma_{j,80}$ agree quite well with INMA linewidths we additionally suspect that the rotational relaxation of the molecule remained incomplete during our INMA energy minimizations. As a result, even an approximate analysis, which is based on the average autocorrelation function (24) of the TF-GV frequencies $\omega_j(t|\tau)$, can render insights how the various dynamical processes shape the MIR bands.

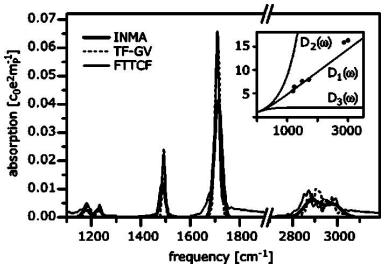

FIG. 8. MIR spectra of FA in aqueous solution, calculated by INMA (thick solid line), TF-GV (dotted), and the quantum corrected FTTCF expression (thin solid), as given by Eqs. (23), (25), and (28), respectively. For noise reduction, the latter spectrum has been smoothened by a Gaussian kernel (width of 5 cm^{-1}). The scale of the y axis is given by the factor c_0 introduced in Eq. (1), the electron charge e, and the proton mass m_p. Inset: The circles display the ratio $\alpha_{TF\text{-}GV}/\alpha_{FTTCF}$ between the maximum absorption of each mode j, as calculated by TF-GV and by the uncorrected FTTCF expression (22), respectively. The solid lines display the proposed quantum correction factors $D_1(\omega)$, $D_2(\omega)$, and $D_3(\omega)$, respectively (see the text).

E. MIR spectra of FA in solution calculated by the various methods

Due to the good agreement between INMA and TF-GV concerning band positions and IR intensities (see Ref. 18) of FA in aqueous solution, the differences between the corresponding MIR spectra (23) and (25) depicted in Fig. 8 are mainly caused by the different linewidths documented in Table I. In contrast to this satisfying agreement, the maximum TF-GV absorption of each mode is underestimated by the FTTCF expression (22) (spectrum not shown) by an average factor of about 10. We attribute this underestimate mainly to the difference between the classical dynamics used in our MD trajectory and the quantum mechanics assumed for the derivation of the FTTCF expressions (2) and (22).

In order to account for this difference various quantum correction factors have been proposed, which are multiplied to the FTTCF spectrum (22). By construction, such corrections are neither needed for the INMA spectrum (23) nor for the TF-GV spectrum (25). Because the inset in Fig. 8 shows that the FTTCF underestimate scales about linearly with the frequency and because the so-called harmonic approximation quantum correction factor[29]

$$D_1(\omega) = \frac{\hbar\omega}{k_B T} \frac{1}{1 - \exp(-\hbar\omega/k_B T)}, \tag{27}$$

also scales linearly in the MIR spectral region, it should be well suited for a correction. In contrast, two other proposed correction factors[29] given by $D_2(\omega) = \exp(\hbar\omega/2k_B T)$ and by $D_3(\omega) = 2/[1 + \exp(-\hbar\omega/k_B T)]$, respectively, exhibit a completely different frequency dependence in the MIR (cf. the inset).

Indeed, Fig. 8 reveals that at $T = 300$ K the thus corrected spectrum

$$\alpha'_{FTTCF}(\omega) = D_1(\omega)\alpha_{FTTCF}(\omega) \tag{28}$$

matches the TF-GV spectrum quite nicely. This result agrees with recent studies, which have shown that the correction

factor (27) yields the best agreement for comparing MM-MD of water with experimental data[30,32] and for a theoretical analysis of model systems.[31]

IV. SUMMARY

In Paper I of this work,[18] the attractive idea to calculate the MIR vibrational spectra of molecules in polar solvents from DFT/MM-MD trajectories by an analysis of classical atomic fluctuations—instead from established methods such as INMA of FTTCF—had led us to the proposal of the so-called generalized virial approach. Unfortunately, the GV spectra calculated for FA in water turned out to be sizably affected by systematic and artificial frequency underestimates, which were assigned by us to inevitable anharmonicities and temporal fluctuations of the molecular Hamiltonian.

To check this preliminary analysis and to remove the artifacts, we now applied a bandpass filtering procedure to the normal coordinate trajectories of FA. The resulting F-GV frequencies turned out to be quite accurate proving that our original hypotheses concerning the time scales and the effects of fluctuations as well as of anharmonicities are correct. Furthermore, we can state that slow fluctuations of (i) the equilibrium geometry and (iii) directions of the normal modes as well as (iv) diagonal and (v) off-diagonal anharmonicities of the molecular potential, which are sampled by the classical MD trajectory, can be identified and removed by bandpass filtering. Note that the filtering method requires that the spectral signatures of the various artifacts are clearly separated from the fundamental band in frequency space. This restriction limits the applicability of F-GV to small molecules with well separated, nonoverlapping vibrational bands. As a further consequence, GV frequency underestimates, which are caused by fluctuations of (ii) the force constants, cannot be removed by bandpass filtering.

Although the influence of this latter effect (ii) on the GV frequencies has been shown to be small ($\approx 1\ \mathrm{cm}^{-1}$) in the MIR, the aim to remove also this artifact has motivated a time-resolved calculation of TF-GV frequencies (14), which yields the trajectory $\omega_j(t|\tau)$ of the vibrational frequency of each mode at a resolution $\tau > \tau_j$ limited by the corresponding oscillation time $\tau_j = 2\pi/\omega_j$.

By analyzing the TF-GV trajectories of the vibrational frequencies resulting from our DFT/MM simulation of FA in water we found that large parts of the fluctuations are induced by fast librational motions of the molecule and of its solvent shell. Slower fluctuations are caused by conformational transitions within that shell. Investigating the sizes and time scales of the associated fluctuations we have shown that the faster fluctuations are in the motional narrowing limit and, therefore, contribute only weakly to the total linewidths, whereas the slower fluctuations are well outside that limit and generate the inhomogeneous broadenings. The inhomogeneously broadened TF-GV line shapes then agree very well with the FTTCF results.

This result and our comparison of INM (Refs. 19 and 20) and INMA (Ref. 9) linewidths further suggest that the overestimate of linewidths by INM, which is caused by the neglect of motional narrowing, can be partially corrected if the energy of the molecule is minimized in its solvent cage before frequency calculation, as required by INMA. Due to minimization the molecule can avoid strong collisional interactions with its environment and, therefore, the fluctuations of the INMA frequencies are smaller than those of INM.

Due to the good agreement between TF-GV and INMA concerning band positions and IR intensities, the corresponding MIR spectra of formaldehyde in aqueous solution calculated with these methods differ only in the linewidths and, consequently, also in the maximum absorption. In contrast, the harmonic approximation quantum correction factor (27) and a temperature scaling factor (cf. Sec. II B), which solely affect the IR intensities but neither the band positions nor the linewidths turned out to be necessary in order to obtain a good agreement also for the FTTCF spectrum (22). Concerning FTTCF we finally note that the use of large integration time steps in the Verlet algorithm, by which MD trajectories are commonly generated, causes sizable frequency overestimates.

Summarizing we may state that all three methods for computing MIR spectra of molecules in polar solvents from DFT/MM calculations have their merits:

INMA performs well at band positions and IR intensities, allows a rapid analysis of isotope effects, and may be extended towards a quantum mechanical treatment of anharmonic effects.

FTTCF automatically includes dynamical effects such as motional narrowing and therefore yields excellent line shapes. Upon inclusion of appropriate corrections for various artifacts (Verlet integration, classical treatment of nuclear motion, incomplete thermal equilibration) it can yield accurate harmonic spectra.

TF-GV also yields accurate harmonic spectra and enables a detailed analysis of the dynamical processes by which the fluctuating environment modulates the force constants of the solute molecule. Based on such an analysis one can then distinguish homogeneous and inhomogeneous contributions to the line shapes.

ACKNOWLEDGMENTS

Support by the Volkswagen Stiftung (Project No. I/73224) and by the Deutsche Forschungsgemeinschaft (Grant No. SFB 533/C3) is gratefully acknowledged. The authors also thank R. Ramirez, T. Lopez-Ciudad, P. Kumar, and D. Marx for communicating their results on quantum correction factors prior to publication.

APPENDIX: VERLET DESCRIPTION OF A HARMONIC OSCILLATOR

We consider the Verlet algorithm[35]

$$a(t_{n+1}) = 2a(t_n) - a(t_{n-1}) + (\Delta t)^2 f(t_n) \tag{A1}$$

with a time step Δt for a harmonic oscillator of restoring force $f(t) = -\omega_0^2 a(t)$, where $a(t)$ is a mass-weighted normal coordinate and ω_0 the associated harmonic constant. Interestingly, the discretized trajectory

$$a(t_n) = \cos \omega t_n \tag{A2}$$

generated by the Verlet algorithm for $a(t)$ is characterized by a frequency

$$\omega = \omega_0\{1 + (1/24)(\omega_0\Delta t)^2 + O[(\omega_0\Delta t)^4]\}, \tag{A3}$$

which is larger than ω_0 for finite Δt. This systematic Verlet overestimate of ω_0 immediately follows by inserting Eq. (A2) into Eq. (A1) and by expanding $a(t_{n+1})$ into a Taylor series around $a(t_n)$ to the fifth order in $(\omega\Delta t)$. Note that we have detected this error of the Verlet algorithm by a Fourier transformation of a numerically integrated harmonic oscillator.

However, the above numerical inaccuracy only hampers the FTTCF results but leaves GV frequencies untouched by a fortunate cancellation of errors. For a trajectory $a(t_n)$ determined by the Verlet algorithm the GV frequency is

$$\omega_{\mathrm{GV}}^2 = \frac{\langle [a(t_{n+1}) - a(t_n)]^2 \rangle}{(\Delta t)^2 \langle a^2(t_n) \rangle}. \tag{A4}$$

Inserting once again Eq. (A2) and applying the same Taylor expansion one finds

$$\omega_{\mathrm{GV}} = \omega\{1 - (1/24)(\omega\Delta t)^2 + O[(\omega\Delta t)^4]\}, \tag{A5}$$

where ω is the frequency generated by the Verlet algorithm. Using the Verlet overestimate (A3) one finds

$$\omega_{\mathrm{GV}} = \omega_0\{1 + O[(\omega\Delta t)^4]\}, \tag{A6}$$

i.e., the errors of order $O[(\omega\Delta t)^2]$ cancel as mentioned above. As we have noticed in test calculations for a one-dimensional harmonic oscillator (not shown here), this cancellation of errors even pertains to time steps of a size, for which the Taylor expansion becomes inapplicable. We have observed, for instance, an agreement $(\omega_0 - \omega_{\mathrm{GV}})/\omega_0 \leqslant 0.001$ between the GV frequency ω_{GV} and the harmonic frequency ω_0 for time steps as large as $\Delta t = \pi/2\omega_0$.

[1] J. Breton, J.-R. Burie, C. Berthomieu, Biochemistry **33**, 4953 (1994).
[2] C. Allin, M. R. Ahmadian, A. Wittinghofer, and K. Gerwert, Proc. Natl. Acad. Sci. U.S.A. **98**, 7754 (2001).
[3] A. Barth and C. Zscherp, Q. Rev. Biophys. **35**, 369 (2002).
[4] K. Hauser, M. Engelhard, N. Friedman, M. Sheves, and F. Siebert, J. Phys. Chem. A **106**, 3553 (2002).
[5] R. Vogel, F. Siebert, G. Mathias, P. Tavan, G. Fan, and M. Sheves, Biochemistry **42**, 9863 (2003).
[6] K. G. R. Pachler, F. Matlok, and H.-U. Gremlich, *Merck FTIR Atlas* (VCH, Weinheim, 1988).
[7] F. Siebert, Methods Enzymol. **246**, 501 (1995).
[8] M. Nonella, G. Mathias, M. Eichinger, and P. Tavan, J. Phys. Chem. B **107**, 316 (2003).
[9] M. Nonella, G. Mathias, and P. Tavan, J. Phys. Chem. A **107**, 8638 (2003).
[10] M. Schmitz, P. Tavan, and M. Nonella, Chem. Phys. Lett. **349**, 342 (2001).
[11] J. Neugebauer and B. A. Hess, J. Chem. Phys. **118**, 7215 (2003).
[12] M. Eichinger, P. Tavan, J. Hutter, and M. Parrinello, J. Chem. Phys. **110**, 10452 (1999).
[13] A. Warshel and M. Levitt, J. Mol. Biol. **103**, 227 (1976).
[14] M. Klähn, G. Mathias, C. Kötting, M. Nonella, J. Schlitter, K. Gerwert, and P. Tavan, J. Phys. Chem. A **108**, 6186 (2004).
[15] R. A. Wheeler, H. Dong, and S. E. Boesch, ChemPhysChem **4**, 382 (2003).
[16] R. A. Wheeler and H. Dong, ChemPhysChem **4**, 1227 (2003).
[17] Q. Cui and M. Karplus, J. Chem. Phys. **112**, 1133 (2000).
[18] M. Schmitz and P. Tavan, J. Chem. Phys. **121**, 12233 (2004) (preceding paper).
[19] R. M. Stratt, Acc. Chem. Res. **28**, 201 (1995).
[20] T. Keyes, J. Phys. Chem.**101**, 2921 (1997).
[21] H. H. Nielsen, Rev. Mod. Phys. **23**, 90 (1951).
[22] A. Willets, N. C. Handy, W. H. Green, Jr., and D. Jayatilaka, J. Phys. Chem. **94**, 5608 (1990).
[23] S. Mukamel, *Principles of Nonlinear Optical Spectroscopy* (Oxford University Press, Oxford, 1995).
[24] R. Kubo, M. Toda, and N. Hashitsume, *Statistical Physics II—Nonequilibrium Statistical Mechanics* (Springer, Heidelberg, 1985).
[25] R. Kubo, Adv. Chem. Phys. **15**, 101 (1969).
[26] D. W. Oxtoby, Adv. Chem. Phys. **40**, 2 (1979).
[27] D. A. McQuarrie, *Statistical Mechanics* (Harper and Row, New York, 1976), Chap. 21.
[28] P. L. Silvestrelli, M. Bernasconi, and M. Parrinello, Chem. Phys. Lett. **277**, 478 (1997).
[29] J. Borysow, M. Moraldi, and L. Frommhold, Mol. Phys. **56**, 913 (1985).
[30] H. Ahlborn, B. Space, and P. B. Moore, J. Chem. Phys. **112**, 8083 (2000).
[31] R. Ramirez, T. Lopez-Ciudad, P. Kumar, and D. Marx, J. Chem. Phys. **121**, 3973 (2004).
[32] M. P. Gaigeot and M. Sprik, J. Phys. Chem. B **107**, 10344 (2003).
[33] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C*, 2nd ed. (Cambridge University Press, Cambridge, 1992).
[34] D. A. Clabo, W. D. Allen, R. B. Remington, Y. Yamaguchi, and H. F. Schaefer III, Chem. Phys. **123**, 187 (1988).
[35] L. Verlet, Phys. Rev. Lett. **159**, 98 (1967).
[36] S. Tolosa and J. A. Sanson, Chem. Phys. **213**, 203 (1996).
[37] S. Tolosa and J. A. Sanson, Chem. Phys. **223**, 251 (1997).

4 Ausführliche Zusammenfassung und Ausblick

"THIS IS FINISH BUT NOT THE END"
(*Ripuli* [104])

Wie aus den Argumenten in der Einleitung hervorgeht, sind, um Reaktionsintermediate in experimentellen zeitaufgelösten Infrarot- (IR-) Spektren eindeutig identifizieren zu können, genau berechnete IR Spektren aller möglichen Intermediate notwendig.

Als ersten Schritt auf dem Weg zu solchen Rechnungen für die lichtinduzierte Reparatur von Cyclopyrimidindimer- (CPD-) Defekten in der DNS durch die Photolyase habe ich die Gasphasen IR Spektren von Modellstrukturen verschiedener in der Literatur vorgeschlagener Intermediate dieser Reparatur mit Standard DFT Methoden berechnet. Bei der Analyse der IR Spektren habe ich mich auf den Frequenzbereich 1500–1800 cm^{-1} konzentriert, der die IR starken C=O und die weniger IR starken C=C Streckschwingungen dieser Moleküle umfasst.

Die gute Übereinstimmung des berechneten IR Spektrums von Thymin mit experimentellen Gasphasenspektren in diesem Spektralbereich zeigt, dass die gewählte DFT Methode zur genauen Berechnung solcher IR Spektren geeignet ist. Die berechneten IR Spektren zeigen ferner, dass verschiedene chemische Modifikationen der Moleküle jeweils charakteristische Frequenzverschiebungen der C=O und der C=C Schwingungen bewirken: Die Dimerisierung zweier Pyrimidinbasen bewirkt Verschiebungen der C=O und der C=C Schwingungsfrequenzen, die anhand von elektronischen Resonanzstrukturen erklärt werden können. Das Monomer und das Dimer des 1-Methyl-5-Flour-Uracils weisen in den Rechnungen in Übereinstimmung mit experimentellen Beobachtungen jeweils höhere Carbonylfrequenzen auf als die entsprechenden Strukturen des 1-Methyl-Uracils und des 1-Methyl-Thymins. Schließlich bewirkt die Besetzung eines anti-bindenden Orbitals durch ein Radikalelektron in den anionischen Strukturen eine erhebliche Schwächung der C=O Bindung und dadurch eine starke Rotverschiebung der Frequenzen der Carbonylmoden um 70–100 cm^{-1} im Vergleich zu den entsprechenden neutralen Strukturen.

Anhand dieser Resultate kann man feststellen, dass die IR starken Carbonylmoden der Thyminbasen in der DNS sehr empfindlich auf jene chemischen Modifikationen reagieren, die sie während der CPD Reparatur durch Photolyase erfahren.

Durch gezielte Isotopenmarkierungen können die IR Spektren verschiedener Intermediate noch deutlicher voneinander unterscheidbar gemacht werden. Aus dieser Beobachtung habe ich den Schluss gezogen, dass die Reaktionsintermediate der CPD Reparatur in zeitaufgelösten, experimentellen IR Spektren anhand ihrer Carbonylmoden identifiziert werden können und dass man auf diese Weise den zeitlichen Verlauf der Reaktion rekonstruieren kann.

Als weiteren Schritt auf dem Weg zu einer genauen Berechnung von Infrarot- (IR-) Spektren habe ich in der Einleitung von Abschnitt 3.1 die instantane Normalmodenanalyse (INMA) und die Fouriertransformation der Zeitkorrelationsfunktion (FTT-CF) des Dipolmoments zur Berechnung von Mittelinfrarot- (MIR-) Spektren einzelner Moleküle in einem polaren Lösungsmittel mittels DFT/MM-MD Simulationen aus der linearen quantenmechanischen Störungstheorie („Fermis Goldene Regel“ [98]) abgeleitet und die Auswirkungen der jeweils zugrundeliegenden Näherungen und Annahmen diskutiert. Da sich viele der dabei verwendeten Argumente auch auf den allgemeineren Fall eines Moleküls in kondensierter Phase übertragen lassen, stellt diese Herleitung eine wichtige Diskussionsgrundlage für zukünftige Vergleiche von berechneten und experimentellen MIR Spektren dar.

Diese Diskussion bildet ferner die Grundlage für die Analyse der *principal mode analysis* (PMA) zur Berechnung von MIR Spektren [100, 101]. Bei dieser Analyse habe ich die fünf Bedingungen bestimmt, die erfüllt sein müssen, damit die PMA exakte Ergebnisse liefert: (α) man muss die Kovarianzanalyse auf die massengewichteten Auslenkungen des Moleküls anwenden, (β) das Molekül darf in dem verwendeten Bezugssystem keinen Gesamtimpuls oder -drehimpuls haben, (γ) das Kernpotential muss harmonisch sein, (δ) die kinetische Energie des Moleküls muss gleichmäßig auf alle Schwingungsmoden verteilt sein, (ϵ) der molekulare Hamiltonoperator muss zeitunabhängig sein. Ferner habe ich gezeigt, wie man die IR Intensität einer Schwingungsmode aus einer MD Trajektorie berechnen kann.

Die Bedingungen (γ)–(ϵ) sind in DFT/MM Simulationen von Molekülen in polaren Lösungsmitteln üblicherweise verletzt: Das Kernpotential ist in der Regel anharmonisch, das thermische Gleichgewicht wird in den kurzen DFT/MM-MD Simulationen üblicherweise nicht erreicht und der molekulare Hamiltonoperator fluktuiert aufgrund der Wechselwirkung des gelösten Moleküls mit seiner zeitabhängigen Umgebung. Die Auswirkungen dieser Verletzungen habe ich im einzelnen anhand einfacher analytischer Modelle und einer DFT/MM-MD Trajektorie eines Formaldehydmoleküls in der Gasphase und in wässriger Lösung getestet. Dieses Molekül habe ich verwendet, weil es das kleinste Molekül ist, das eine der im Kap. 2.1 betrachteten Carbonylgruppe besitzt.

Bei diesen Tests hat sich herausgestellt, dass man den riesigen Fehler, den die Verletzung von (δ) bei den Frequenzen verursacht, vermeiden kann, wenn man für jede Mode j eine modenspezifische Temperatur T_j berechnet und die Frequenz ω_j nicht über PMA, sondern über den „generalisierten Virial“ (GV) Ausdruck $\omega_j^2 = k_B T_j / \mathcal{V}(a_j)$ aus der Varianz der massengewichteten Normalkoordinate a_j ge-

winnt. Ferner führen Verletzungen der Bedingungen (γ) und (ϵ), die sich weiter in Fluktuationen (i) der Ruhelage, (ii) der internen Kraftkonstanten und (iii) der Normalmoden, bzw. in (iv) diagonale und (v) nicht-diagonale Anharmonizitäten aufteilen lassen, zu systematischen Unterschätzungen von Frequenzen sowohl bei der PMA als auch bei der GV Methode.

Da bei kleinen Molekülen die Fehlerquellen (i) und (iii)–(v) für die GV Frequenz ω_j einer Mode j im Frequenzraum in der Regel weit von der fundamentalen MIR Oszillation der Normalkoordinate a_j entfernt liegen, können sie durch einen Bandpassfilter beseitigt werden. Bei großen Molekülen mit stark überlappenden Schwingungsbanden ist der Fehler (iii) auf diese Weise nicht eliminierbar. Der Fehler (ii), der bei den GV Frequenzen durch fluktuierende Kraftkonstanten verursacht wird, kann in keinem Fall durch einen Filter beseitigt werden. Obwohl ich gezeigt habe, dass dieser Fehler nur etwa 1 cm^{-1} beträgt, hat mein Bemühen, diesen Fehler zu vermeiden, die TF-GV Methode hervorgebracht. Diese Methode erlaubt es, die Schwingungsfrequenzen eines Moleküls zeitaufgelöst auf der Zeitskala zu berechnen, die durch die Schwingungsdauer der betrachteten Mode gegeben ist.

Als Anwendungsbeispiel habe ich diese TF-GV Methode auf die oben schon erwähnte DFT/MM-MD Trajektorie des Formaldehydmoleküls in Wasser angewandt und dadurch für jede Mode dieses Moleküls eine Trajektorie der Schwingungsfrequenz mit einer Zeitauflösung von 10–30 fs erhalten. Auf diese Weise konnte ich die Ursachen für die beobachteten Frequenzfluktuationen genau untersuchen: Durch eine Analyse aller Trajektorien habe ich herausgefunden, dass ein großer Teil der Frequenzfluktuationen durch die translatorischen und rotatorischen Librationsbewegungen des gelösten Moleküls in dem Lösungsmittelkäfig verursacht wird. Die Zeitskalen dieser Fluktuationen sind dabei jedoch so kurz ($\sim$ 100 fs) und ihre Amplitude so gering ($\sim$ 10 cm^{-1}), dass sie aufgrund des *motional narrowing* Effektes [105–108] nur schwach zur MIR Linienbreite beitragen.

Analog zu diesem linienverschmälernden Effekt kann durch Vergröberung der Zeitauflösung der TF-GV Frequenzen auf 250 fs über die schnellen Frequenzfluktuationen gemittelt werden. In der Trajektorie der daraus resultierenden TF-GV Frequenzen werden langsamere Frequenzfluktuationen sichtbar (siehe Abb. 6 in Abschnitt 3.2), die durch Konformationsübergänge in der Lösungsmittelhülle verursacht werden. Diese Fluktuationen sind langsam ($\sim$ 2 ps) und stark ($\sim$ 10 cm^{-1}) genug, um vollständig zur „inhomogenen“ MIR Linienbreite beizutragen.

Anhand aller hier erwähnten Analysen und Resultate können die verschiedenen Methoden INMA, FTTCF und TF-GV zur Berechnung von MIR Spektren einzelner Moleküle in polaren Lösungsmitteln unter verschiedenen Gesichtspunkten bewertet werden. Die in Tabelle 4.1 aufgelisteten Bewertungen + (gut), o (befriedigend) und − (schlecht) werde ich im folgenden begründen:

Effizienz: Die INMA Methode hat gegenüber anderen Methoden den Vorteil, dass ein numerisch vergleichsweise schnell zu berechnendes MM Kraftfeld zum Generie-

	INMA	FTTCF	TF-GV
Effizienz	+	−	−
Anwendbarkeit	+	+	−
Frequenzen	−/+	+	+
IR Intensität	+	+	+
Linienbreiten	− /o	+	+
Isotopeneffekte	+	−	−
Zeitauflösung	o	−	+

Tabelle 4.1: Stärken und Schwächen der verschiedenen Methoden zur Berechnung von MIR Spektren einzelner Moleküle in polaren Lösungsmitteln auf der Skala gut (+), befriedigend (o), schlecht (−). Zu weiteren Erläuterung siehe Text.

ren des thermischen Ensembles von Lösungsmittelstrukturen verwendet werden kann [81, 95]. Das numerisch aufwendigere DFT/MM Kraftfeld wird dann nur für eine kurze Equilibrierung, zum Minimieren und zur Berechnung der Hesse Matrix benötigt. Diese drei Schritte können zudem für jede Lösungsmittelkonfiguration unabhängig von anderen Konfigurationen durchgeführt werden, so dass diese DFT/MM Rechnungen auf triviale Weise parallelisierbar sind.

Sowohl FTTCF als auch TF-GV benötigen für die Berechnung des IR Spektrums eine zusammenhängende Trajektorie, so dass eine solche Vereinfachung hier nicht möglich ist. Ihre spektrale Auflösung $\delta\omega_j = 1/L$ ist zudem durch die Länge L dieser Trajektorie begrenzt [109]. Demnach ist eine zusammenhängende 30 ps Trajektorie notwendig, um das IR Spektrum eines Moleküls mit einer Genauigkeit von 1 cm^{-1} zu bestimmen.

Anwendbarkeit: Weder für die Anwendbarkeit von INMA noch für die von FTTCF gibt es bezüglich der Größe des Moleküls methodisch bedingte Einschränkungen. Zur Anwendung der TF-GV Methode ist jedoch ein Frequenzfilter notwendig, der die fundamentale Oszillation einer Normalkoordinate z.B. von störenden Einflüssen anderer Moden bereinigt. Daher ist diese Methode nur für Moleküle anwendbar, deren Moden nicht stark mischen und deren Schwingungsbänder dadurch nicht stark überlappen. Diese Bedingung erfüllen jedoch nur kleine Moleküle mit maximal 10–20 Atomen.

Durch die Entwicklung einer Filtermethode, die intelligenter ist als der bisher verwendete Rechteckfilter, indem sie zum Beispiel die bisher ignorierte Phaseninformationen unterschiedlicher Schwingungen ausnutzt, könnte diese Einschränkung jedoch behoben werden.

Frequenzen: Da das Molekül während der Trajektorie Konfigurationen einnehmen kann, an denen das molekulare Potential konvex ist, kann eine instantane Berechnung der Hesse Matrix ohne vorherige Energieminimierung zu imaginären INMA Frequenzen ω_j führen, deren physikalische Bedeutung fraglich ist [96, 110, 111]. Minimiert man dagegen die Energie des Moleküls vor der Frequenzberechnung gemäß dem in den Artikeln [91, 93] beschriebenen Protokoll, so liefert die INMA Methode genaue harmonische und reelle Frequenzen.

Ausserdem ist INMA als einzige Methode um eine störungstheoretische Berechnung anharmonischer Quantenkorrekturen erweiterbar. Aus meinen Ergebnissen geht hervor, dass die FTTCF und die TF-GV Methode dagegen anharmonische Verschiebungen nur insoweit liefern, wie die anharmonischen Teile der Potentialoberfläche während der klassischen Trajektorie abgetastet werden. Man beachte, dass das Molekül bei Raumtemperatur nur geringe thermische Auslenkungen um die Gleichgewichtslage erfährt, an denen das Potential in guter Näherung harmonisch ist. Daher sind hier sind die resultierenden klassischen anharmonischen Verschiebungen sehr viel kleiner als die experimentell beobachtbaren quantenmechanischen Verschiebungen.

Schließlich habe ich gezeigt, dass die FTTCF Frequenzen artifiziell hoch sind, wenn die zugrundeliegende MD Trajektorie mit dem Verlet-Algorithmus [112] und einer großen Schrittweite erzeugt wurde. Dieses Artefakt kann jedoch nachträglich näherungsweise korrigiert werden.

IR Intensität: Die IR Intensität wird von allen Methoden gleichermaßen gut reproduziert. In Übereinstimmung mit fast gleichzeitig geführten Studien [99, 111, 113, 114] habe ich gezeigt, dass dazu jedoch der harmonische Korrekturfaktor

$$D(\omega) = \frac{\hbar\omega}{k_B T} \frac{1}{1 - \exp(-\hbar\omega/k_B T)}, \tag{4.1}$$

auf das FTTCF Spektrum multipliziert werden muss, wenn die Kernbewegung — wie in MD Simulationen üblich — klassisch beschrieben wird. Außerdem habe ich gezeigt, dass die FTTCF Intensitäten weiterhin eine Temperaturskalierung benötigen, wenn das Molekül während der MD Simulation das thermische Gleichgewicht nicht erreicht.

Linienbreite: Die INMA Methode beruht auf einer Separation der Dynamik des gelösten Moleküls und der des Lösungsmittels. Dadurch ignoriert sie dynamische, linienverschmälernde Effekte, z.B. das *motional narrowing* [105–108], was zu einer Überschätzung der INMA Linienbreiten führt. Ich habe gezeigt, dass diese Überschätzung der INMA Linienbreiten zum Teil dadurch näherungsweise kompensiert werden kann, dass die Energie des Moleküls vor der Frequenzberechnung gemäß dem in den Artikeln [91, 93] beschriebenen Protokoll im festgehaltenen Lösungsmittelkäfig minimiert wird. Eine solche Minimierung erlaubt dem Molekül, starken Wechselwirkungen mit diesem Käfig auszuweichen und so die Fluktuationen der Frequenzen zu verringern.

Die TF-GV Methode beruht ebenfalls auf einer entkoppelten Analyse der Dynamik des gelösten Moleküls. Bei hoher TF-GV Zeitauflösung führt dies analog zur INMA zu einer Überschätzung der Linienbreite durch die Standardabweichung der TF-GV Frequenzen. Diese Überschätzung kann durch eine Vergröberung der Zeitauflösung behoben werden, da hierdurch die schnellsten Fluktuationen, die aufgrund des *motional narrowing* nur vernachläsigbar zur Linienbreite beitragen, herausgemittelt werden. Bei höchster Zeitauflösung erlaubt die TF-GV Methode eine genaue Analyse der Frequenzfluktuationen und von ihren Ursachen. Wie ich in Kapitel 3.2 gezeigt habe, kann dadurch bei ausreichender statistischer Grundlage das *motional narrowing* analysiert werden und die Linienform exakt aus der Korrelationsfunktion der Frequenzen berechnet werden.

Die FTTCF Methode basiert dagegen nicht auf einer Separation der Dynamik des gelösten Moleküls und der des Lösungsmittels, so dass hier alle dynamischen Effekte vollständig berücksichtigt werden.

Isotopeneffekte: Die Hesse Matrizen, die bei INMA für jeden Lösungsmittelkäfig bestimmt werden, erlauben bei verschwindend geringem Rechenaufwand einen direkten Zugang zu den in der experimentellen Praxis sehr wichtigen Isotopeneffekten [87, 88]. Um Isotopeneffekte mit FTTCF oder TF-GV zu berechnen, müsste zum jetzigen Stand der Technik eine neue MD Trajektorie des isotopenmarkierten Moleküls berechnet werden.

Eine Erweiterung der GV Methode zu einer schnellen Berechnung von Isotopeneffekten könnte an der Hesse Matrix ansetzen, die näherungsweise aus der Kovarianzmatrix der atomaren Auslenkungen folgt (siehe Abschnitt 3.1).

Zeitauflösung: Durch die Berechnung des IR Spektrums aus einem globalen Erwartungswert und einer anschließenden Fouriertransformation bietet die FTTCF keinen Zugang zu expliziten Zeitabhängigkeiten der Schwingungsfrequenzen. Die INMA Frequenzen des gelösten Moleküls werden dagegen einzeln in verschiedenen Lösungsmittelkäfigen berechnet und erlauben so, den Einfluss der Struktur der Umgebung auf das IR Spektrum des Moleküls sehr genau zu untersuchen [76, 81]. Für eine hohe Zeitauflösung sind jedoch sehr viele Frequenzberechnungen und damit ein hoher Rechenaufwand notwendig. Insbesondere lassen sich die Umgebungseffekte, die zur homogenen bzw. zur inhomogenen Linienbreite beitragen, nicht unterscheiden. Mit der TF-GV Methode lassen sich dagegen die Frequenzfluktuationen auf einfache Weise bis zu einer Zeitskala auflösen, die durch die Schwingungsfrequenz der betrachteten Moden gegeben ist. Damit eröffnet sie den Weg zur Bestimmung der Ursachen für Frequenzfluktuationen sowie für die Analyse der Linienbreiten (siehe oben: **Linienbreiten**).

Ausblick

Zukünftige Anwendungsmöglichkeiten der TF-GV Methode bieten all jene Systeme, bei denen experimentell beobachtbare Frequenzfluktuationen oder -verschiebungen mit dynamischen Prozessen (z.B. Konformationsdynamik [16, 72, 115] oder Energierelaxation [116]) in Zusammenhang gebracht werden. Zusammen mit den in MD Simulationen zugänglichen Details der molekularen Dynamik kann die TF-GV Methode hier dazu dienen, die Ursachen für diejenigen Phänomene zu untersuchen, die in experimentell gemessenen IR Spektren beobachtet werden.

Des weiteren ermöglicht die vor kurzem ermittelte Kristallstruktur des DNS/Photolyase Komplexes [45] DFT/MM-MD Simulationen, mit denen nicht nur die katalysierte Reparatur der CPD Defekte durch die Photolyase nachvollzogen und deren treibende Kräfte untersucht werden kann. Auch die IR Spektren möglicher Intermediate der Reparatur können jetzt aufgrund der neuerdings bekannten Bindungsgeometrie der DNS Basen in der Bindungstasche der Photolyase genau *in situ* berechnet werden. Für eine solche Berechnung von IR Spektren müssen die hier vorgestellten Methoden jedoch erweitert werden, um die Kopplung z.B. der Carbonylmoden der DNS Basen an resonante Schwingungsmoden der Umgebung (Amidmoden des Proteins, Carbonylmoden des FADH) korrekt zu berücksichtigen.

A Über die in der DFT/MM Trajektorie beobachteten Gradientensprünge

Der folgende Abschnitt ist ein Abdruck des Artikels[1]

> Matthias Schmitz and Paul Tavan: „Supplementary Material to: Vibrational spectra from atomic fluctuations in dynamics simulations: I. Theory, limitations, and a sample application“, EPAPS Document No. E-JCPSA6-121-511447.

Er enthält Informationen über die Ursache, die Folgen und die Behebung der artifiziellen Gradientensprünge, die in der in Kapitel 3 diskutierten DFT/MM Trajektorie innerhalb des DFT Fragments aufgetreten sind.

[1] Ein direkter Link auf eine elektronische Version dieses Dokuments befindet sich in dem HTML Literaturverzeichnis des Onlineartikels [2]. Das Dokument kann auch über die EPAPS Homepage (http://www.aip.org/pubservs/epaps.html) oder über ftp.aip.org im Verzeichnis /epaps/ erreicht werden. Auf der EPAPS Homepage befinden sich weitere Informationen. Mit freundlicher Genehmigung des American Institute of Physics.

Supplementary Material to: Vibrational spectra from atomic fluctuations in dynamics simulations: I. Theory, limitations, and a sample application

Matthias Schmitz and Paul Tavan
Theoretische Biophysik, Lehrstuhl für Biomolekulare Optik,
Ludwig-Maximilians-Universität München, Oettingenstraße 67, 80538 München, Germany

Gradient jumps affecting the DFT fragment in the DFT/MM trajectory

As noted in Section IV B 2 of our paper, the formaldehyde molecule (FA, DFT-fragment) experienced occasional, large and artificial force jumps during the 35 ps trajectory. Apart from heating up the molecule this computational artifact had no further consequences and, therefore, was detected by us only after completion of the simulation through an analysis of the energy distribution among the normal coordinates.

The unusually large and artificial gradient jumps are rare events. For documentation figure 1 shows the histogram $h(|\nabla E|^2)$ of the squared gradients of the energy of the DFT-fragment FA as calculated by our DFT/MM-MD program at each of the 70000 steps of the trajectory (the gradients are taken with respect to all nuclear coordinates in the fragment). As is typical for the expected Maxwell-like behavior, the distribution exponentially decays with increasing gradients up to $|\nabla E|^2 \approx 25000$ kcal^2/mol^{2}Å^2. However, an unexpected plateau appears at gradients beyond that value, from which we conclude that gradients with $|\nabla E|^2 > 25000$ kcal^2/mol^{2}Å^2 are due to computational artifacts of our DFT/MM program. Some of the large gradients fall outside the depicted range, which has been chosen for clear visualization of the exponential behavior. We have

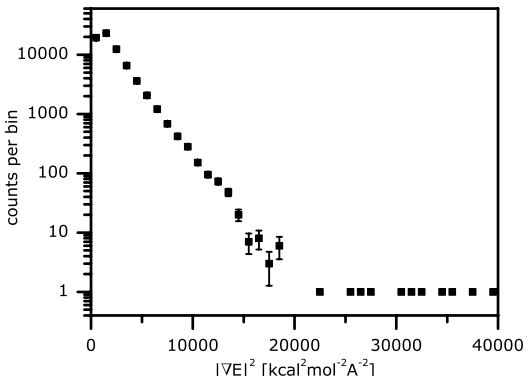

FIG. 1: Logarithmic plot of the histogram $h(|\nabla E|^2)$ of squared gradients observed in the 35 ps DFT/MM trajectory of FA in solution. The bin size is 1000 kcal^2/mol^{2}Å^2. For bins with $h(|\nabla E|^2) > 1$, the error bars denote the expected statistical error.

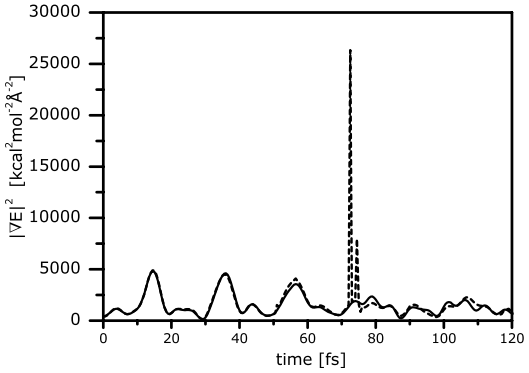

FIG. 2: Trajectory $|\nabla E|^2(t)$ of squared gradients after a restart shortly before one of the force jumps. Dashed line: Original convergence criterion for the Kohn-Sham orbitals (10^{-4}). Solid line: Tightened criterion (10^{-5}).

counted 31 such artifacts, corresponding to an abundance of about once every 2000 simulation steps, or every picosecond of simulation time.

The force jumps solely induce peaks into the second time derivative $\ddot{a}_j(t)$ of the normal coordinates $a_j(t)$, but neither into the trajectories of the normal coordinates $a_j(t)$, nor into their time derivatives $\dot{a}_j(t)$, nor into the GV frequencies derived from these trajectories. Similarly, the FTTCF computations presented in part II of this work and the INMA results discussed here are unaffected from the artifact. In the case of INMA this statement is trivial, because here the gradients are minimized before the computation of the Hessians.

Analysis of trajectories not suffering from gradient jumps

Figure 2 compares two trajectories $|\nabla E|^2(t)$, which both start at identical conditions shortly before one of the force jumps occurred in the original simulation. The dashed line represents that original simulation and exhibits jumps. The solid line is calculated from a simulation, in which the convergence criterion for the DFT wave function has been tightened by a factor of ten. This trajectory is smooth. Force jumps do not occur proving that our original choice of the convergence criterion (10^{-4}) had been too sloppy and that this choice is the

cause of the observed computational artifact.

During the refereeing and the revision process we have computed a new 15 ps DFT/MM-MD trajectory (NT) using the tightened convergence criterion (10^{-5}). In this extended NT force jumps were absent. As a result the standard deviation of the temperatures of the vibrational modes decreased significantly to 60 K as compared to the standard deviation of 177 K observed for our original 35 ps DFT/MM-MD trajectory (OT). Despite this improvement the FA molecule remained far from thermal equilibrium with its 300 K environment in the NT as witnessed by the corresponding average temperature of only 119 K. Therefore one may now ask, whether this incomplete thermal equilibrium is caused by other computational artifacts, for instance by the slight violation of the *actio equals reactio* principle in our hybrid method, which is due to the partial charge approximation for the charge distribution of the DFT fragment.[1]

To check this suspicion, we have computed a 200 ps MM-MD trajectory (MT) of FA in aqueous solution. Also in this comparably long trajectory we have observed an average mode specific temperature of only 207 K for the vibrational modes. Therefore we conclude that the incomplete thermal equilibrium of FA in the NT cannot be attributed to an artifact of the DFT/MM technology, but instead is caused by a very slow thermal equilibration of the solute molecule. The slowness of that equilibration is readily understood considering the coupling of the vibrational modes to the solvent, which should be extremely weak as the solvent translational and rotational dynamics is far off resonance with the FA vibrations.

This conclusion is supported by the much better thermal equilibrium observed for the translational and rotational modes of the FA molecule, which can exchange energy with their environment more easily: For these modes the average mode temperature is 340 K at 124 K standard deviation (s.d.) in OT, is 382 K (65 K s.d.) in NT, and is 312 K (9 K s.d.) in MT.

The vibrational frequencies calculated by the various PMA, GV, and FTTCF methods from the NT slightly differ from the results of the OT presented in this paper and in part II of this work. These differences can be attributed to the different parts of the configuration space sampled during the respective trajectories and to the more limited statistical basis provided by the shorter NT. Nevertheless, the conclusions drawn in the two papers are unaffected by these differences.

---

[1] M. Eichinger, P. Tavan, J. Hutter, and M. Parrinello, J. Chem. Phys. **110**, 10452 (1999).

Literaturverzeichnis

[1] Schmitz, M., P. Tavan und M. Nonella. Vibrational Analysis of Carbonyl Modes in Different Stages of Light-Induced Cyclopyrimidine Dimer Repair Reactions. *Chem. Phys. Lett.* **349**, 342–348 (2001).

[2] Schmitz, M. und P. Tavan. Vibrational spectra from atomic fluctuations in dynamics simulations: I. Theory, limitations, and a sample application. *J. Chem. Phys.* **121**, 12233–12246 (2004).

[3] Schmitz, M. und P. Tavan. Vibrational spectra from atomic fluctuations in dynamics simulations: II. Solvent-induced frequency fluctuations at femtosecond time-resolution. *J. Chem. Phys.* **121**, 12247–12258 (2004).

[4] Simpson, B. (aus: „The Simpsons“, Folge 2F20).

[5] Stryer, L. *Biochemistry*. W. H. Freeman, San Francisco (1975).

[6] Oesterhelt, D. und W. Stoeckenius. Functions of a new photoreceptor membrane. *Proc. Natl. Acad. Sci.* **70**, 2853–2857 (1973).

[7] Itoh, H., A. Takahashi, K. Adachi, H. Noji, R. Yasuda, M. Yoshida und K. Kinosita Jr. Mechanically driven ATP synthesis by F1-ATPase. *Nature* **427**, 465–468 (2004).

[8] Sancar, A. Structure and Function of DNA Photolyase. *Biochemistry* **33**, 2–9 (1994).

[9] Barbacid, M. Ras Genes. *Annu. Rev. Biochem.* **56**, 779–827 (1987).

[10] Baldwin, R. L. The Nature of Protein-Folding Pathways – the Classical Versus the New View. *J. Biomol. NMR* **5**, 103–109 (1995).

[11] Dill, K. und H. Chan. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19 (1997).

[12] Frauenfelder, H., P. G. Wolynes und R. H. Austin. Biological Physics. *Rev. Mod. Phys.* **71**, 419–430 (1999).

[13] Carstens, H. *Konformationsdynamik lichtschaltbarer Peptide: Molekulardynamiksimulationen und datengetriebene Modellbildung*. Dissertation, Ludwig-Maximilians Universität München (2004).

[14] Byler, D. M. und H. Susi. Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers* **25**, 469–487 (1986).

[15] Susi, H. und D. M. Byler. Resolution-enhanced fourier-transform infrared-spectroscopy of enzymes. *Methods in Enzymology* **130**, 290–311 (1986).

[16] Barth, A. und C. Zscherp. What vibrations tell us about proteins. *Quarterly Reviews of Biophysics* **35**, 369–430 (2002).

[17] Vogel, R., F. Siebert, G. Mathias, P. Tavan, G. Fan und M. Sheves. Deactivation of Rhodopsin in the transition from the signaling state Meta II to Meta III involves a thermal isomerization of the retinal chromophore C=N double bond. *Biochemistry* **42**, 9863–9874 (2003).

[18] Wallace, B. A. Gramicidin Channels and Pores. *Ann. Rev. Biophys. Biophys. Chem* **19**, 127–157 (1990).

[19] Frauenfelder, H., S. G. Sligar und P. G. Wolynes. The Energy Landscapes and Motions of Proteins. *Science* **254**, 1598–1603 (1991).

[20] Boyer, P. D. The ATP synthase a splendid molecular machine. *Annu. Rev. Biochem.* **66**, 717–749 (1997).

[21] Blundell, T. L. und L. N. Johnson. *Protein Crystallography*. Academic Press, London (1976).

[22] Giacovazzo, C., H. L. Monaco, D. Viterbo, F. Scordari, G. Gilli, G. Zanotti und M. Catti. *Fundamentals of Crystallography*, Kapitel 8. Oxford University Press, Oxford (1992).

[23] Berman, H. M. *et al*. The Protein Data Bank. *Nucleic Acids Research* **28**, 235–242 (2000). Http://www.rcsb.org/pdb/.

[24] Park, H.-W., S.-T. Kim und A. Sancar. Crystal Structure of DNA Photolyase from Escherichia Coli. *Science* **268**, 1866–1872 (1995).

[25] Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts und J. D. Watson. *Molekularbiologie der Zelle*. VCH, Weinheim (1990).

[26] Zinth, W. und D. Oesterhelt. Primary Photochemical Processes in Bacteriorhodopsin. In E. Riklis, (Herausgeber) *Photobiology, the Science and its Applications*, Seiten 531–535. Plenum Press, New York (1990).

[27] Siebert, F., W. Mäntele und W. Kreutz. Evidence for the protonation of two internal carboxylic groups during the photocycle of bacteriorhodopsin. *FEBS Letters* **141**, 82–87 (1982).

[28] Whittal, R. M., H. L. Ball, F. E. Cohen, A. L. Burlingame, S. L. Prusiner und M. A. Baldwin. Copper binding to octarepeat peptides of the prion protein monitored by mass spectrometry. *Protein Sci.* **9**, 332–343 (2000).

[29] Schmidt, M. Protein Kinetics: Structures of Intermediates and reaction mechanism from time-resolved x-ray data. *Proc. Natl. Acad. Sci. USA* **101**, 4799–4804 (2004).

[30] Canet, D. *NMR – Konzepte und Methoden*. Springer, Berlin (1994).

[31] Neuhaus, D. und M. Williamson. *The nuclear Overhauser effect in structural and conformational analysis*. VCH, Weinheim (1989).

[32] Riek, R., S. Hornemann, G. Wider, M. Billeter, R. Glockshuber und K. Wüthrich. NMR structure of the mouse prion protein domain PrP(121-231). *Nature* **382**, 180–182 (1996).

[33] Garcia, F. L., R. Zahn, R. Riek und K. Wüthrich. NMR structure of the bovine prion protein. *Proc. Natl. Acad. Sci. USA* **97**, 8334–8339 (2000).

[34] Spörlein, S., H. Carstens, H. Satzger, C. Renner, R. Behrendt, L. Moroder, P. Tavan, W. Zinth und J. Wachtveitl. Ultrafast spectroscopy reveals sub-nanosecond peptide con-formational dynamics and validates molecular dynamics simulation. *Proc. Natl. Acad. Sci.* **99**, 7998–8002 (2002).

[35] Nilsson, L., M. Karplus, A. Brünger, M. Clore und A.Gronenbornt. Determination of 3D molecular structures from NMR/NOE distance data: application to nucleic acids. *J. Mol. Graphics* **4**, 182 (1986).

[36] Watson, J. D. und F. H. C. Crick. A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).

[37] McAteer, K., Y. Jing, J. Kao, J. S. Taylor und M. A. Kennedy. Solution-state structure of a DNA dodecamer duplex containing a cis-syn Thymine Cyclobutane Dimer, the major UV photoproduct of DNA. *J. Mol. Biol.* **282**, 1013 (1998).

[38] Kim, S.-T. und A. Sancar. Photochemistry, photophysics, and mechanism of pyrimidine dimer repair by DNA photolyase. *Photochem. Photobiol.* **57**, 895–904 (1993).

[39] Lemaire, D. G. E. und B. P. Ruzsicska. Quantum yields and secondary photoreactions of the photoproducts of dTpdT, dTpdC, and dTpdU. *Photochem. Photobiol.* **57**, 755–769 (1993).

[40] Harm, W. *Biological Effects of ultraviolet radiation*. Cambridge University Press, Cambridge (1980).

[41] Begley, T. P. Photoenzymes: A Novel Class of Biological Catalysts. *Acc. Chem. Res.* **27**, 394–401 (1998).

[42] Sancar, A. und G. B. Sancar. DNA Repair Enzymes. *Annu. Rev. Biochem.* **57**, 29–67 (1988).

[43] Okamura, T., A. Sancar, P. F. Heelis, T. P. Begley, Y. Hirata und N. Mataga. Picosecond Laser Photolysis Studies on the Photorepair of Pyrimidine Dimers by DNA Photolyase. 1. Laser Photolysis of Photolyase-2-Deoxyuridine Dinucleotide Photodimer Complex. *J. Am. Chem. Soc.* **113**, 3143–3145 (1991).

[44] Tamada, T., K. Kitadokoro, Y. Higuchi, K. Inaka, A. Yasui, P. E. de Ruiter, A. P. M. Eker und K. Miki. Crystal structure of DNA photolyase from Anacystis nidulans. *Nature Struct. Biol.* **4**, 887–890 (1997).

[45] Mees, A., T. Klar, P. Gnau, U. Hennecke, A. P. M. Eker, T. Carell und L.-O. Essen. Crystal Structure of a Photolyase Bound to a CPD-Like DNA Lesion After in Situ Repair. *Science* **306**, 1789–1793 (2004).

[46] Voityuk, A. und N. Rösch. Ab Initio Study on the Structure and Splitting of the Uracil Dimer Anion Radical. *J. Phys. Chem. A* **101**, 8335–8338 (1997).

[47] Pezeshk, A., I. D. Podmore, P. F. Heelis und M. C. R. Symons. Electron Addition to Thymine Dimers and Related Compounds: A Mimic of Natural Repair. *J. Phys. Chem.* **100**, 19714–19718 (1996).

[48] Saettel, N. J. und O. Wiest. DFT Study of the [2+2] Cycloreversion of Uracil Dimer Anion Radical: Waters Matter. *J. Am. Chem. Soc.* **123**, 2693–2694 (2001).

[49] Carell, T. und R. Epple. Repair of UV Light Induced DNA Lesions: A Comparative Study with Model Compounds. *Eur. J. Org. Chem.* Seiten 1245–1258 (1998).

[50] Brünger, A. T. Crystallographic refinement by simulated annealing. In N. W. Isaacs und M. R. Taylor, (Herausgeber) *Cyrstallographic Computing 4: Techniques and New Technologies*. Clarendon Press, Oxford (1988).

[51] McCammon, J. A., B. R. Gelin und M. Karplus. Dynamics of folded proteins. *Nature* **267**, 585–590 (1977).

[52] Grubmüller, H., B. Heymann und P. Tavan. Ligand-receptor binding: Molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* **271**, 997–999 (1996).

[53] Böckmann, R. A. und H. Grubmüller. Conformational dynamics of the F1-ATPase β-subunit: A molecular dynamics study. *Biophys. J.* **85**, 1482–1491 (2003).

[54] de Groot, B. L., T. Frigato, V. Helms und H. Grubmüller. The mechanism of proton exclusion in the aquaporin-1 water channel. *J. Mol. Biol.* **333**, 279–293 (2003).

[55] Stork, M., A. Giese, H. A. Kretzschmar und P. Tavan. Molecular dynamics simulations indicate a possible role of parallel β-helices in seeded aggregation and amyloid formation of poly-Gln (2004). Submitted.

[56] Mathias, G., M. Eichinger, H. Carstens, M. Stork, A. Weiss *et al*. EGO-MMII users guide (2004). Lehrstuhl für BioMolekulare Optik, Ludwig Maximilians Universität München, Oettingenstrasse 67, D-80538 München.

[57] Lindahl, E., B. Hess und D. van der Spoel. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Mod.* **7**, 306–317 (2001).

[58] Pearlman, D., D. Case, J. Caldwell, W. Ross, T. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel und P. Kollman. AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules. *Comp. Phys. Comm.* **91**, 1–41 (1995).

[59] Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan und M. Karplus. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **4**, 187–217 (1983).

[60] Ponder, J. W. und D. A. Case. Force Fields for Protein Simulations. *Advances in Protein Chemistry* **66**, 27–85 (2003).

[61] Tavan, P., H. Carstens und G. Mathias. Molecular dynamics simulations of proteins and peptides: Problems, achievements, and perspectives. In J. Buchner und T. Kiefhaber, (Herausgeber) *Handbook of Protein Folding*. Wiley-VCH, Weinheim (2004). Im Druck.

[62] Mathias, G., B. Egwolf, M. Nonella und P. Tavan. A fast multipole method combined with a reaction field for long-range electrostatics in molecular dynamics simulations: The effects of truncation on the properties of water. *J. Chem. Phys.* **118**, 10847–10860 (2003).

[63] McCammon, J. A. Protein Dynamics. *Rep. Prog. Phys.* **47**, 1–46 (1984).

[64] Eaton, W. A., V. Munoz, S. J. Hagen, G. S. Jas, L. J. Lapidus, E. R. Henry und J. Hofrichter. Fast kinetics and mechanisms in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 327–359 (2000).

[65] Qiu, L. L., S. A. Pabit, A. E. Roitberg und S. J. Hagen. Smaller and faster: The 20-residue Trp-cage protein folds in 4 μs. *J. Am. Chem. Soc.* **124**, 12952–12953 (2002).

[66] Geibel, S., J. H. Kaplan, E. Bamberg und T. Friedrich. Conformational Dynamics of the Na+/K+-ATPase probed by voltage clamp flourometry. *Proc. Natl. Acad. Sci.* **100**, 964–969 (2003).

[67] Wilson, E. B., J. C. Decius und P. C. Cross. *Molecular Vibrations*. McGraw Hill, London (1955).

[68] Woodward, L. A. *Introduction to the theory of molecular vibrations and vibrational spectroscopy*. At the Clarendon Press, Oxford (1972).

[69] Atkins, P. W. *Molecular Quantum Mechanics*. Oxford University Press, Oxford, zweite Auflage (1983).

[70] Pachler, K. G. R., F. Matlok und H.-U. Gremlich. *Merck FTIR Atlas*. VCH, Weinheim (1988).

[71] Watson, T. M. und J. D. Hirst. Influence of Electrostatic Environment on the Vibrational Frequencies of Proteins. *J. Phys. Chem. A* **107**, 6843–6849 (2003).

[72] Schultheis, V. Dissertation, Ludwig-Maximilians Universität München, Lehrstuhl für Biomolekulare Optik, AG Theoretische Biophysik, in Vorbereitung.

[73] Hauser, K., M. Engelhard, N. Friedman, M. Sheves und F. Siebert. Interpretation of amide I difference bands observed during protein reactions using site-directed isotopically labeled bacteriorhodopsin as a model system. *J. Phys. Chem. A* **106**, 3553–3559 (2002).

[74] Breton, J. *et al*. Binding-sites of quinones in photosynthetic bacterial reaction centers investigated by light-induced FTIR difference spectroscopy. *Biochemistry* **33**, 4953–4965, 12405–12415, 14378–14386 (1994).

[75] Brudler, R., H. J. M., de Groot, W. B. S. van Liemt, W. F. Steggerda *et al*. Asymmetric binding of the 1-C=O and 4-C=O groups of Q(A) in rhodobacter-sphaeroides-R26 reaction centers monitored by fourier-transform infrared-spectroscopy using site-specific isotopically labeled ubiquinone-10. *EMBO J.* **13**, 5523–5530 (1994).

[76] Nonella, M., G. Mathias, M. Eichinger und P. Tavan. Structures and vibrational frequencies of the quinones in Rb. Sphaeroides derived by a combined density functional / molecular mechanics approach. *J. Phys. Chem. B* **107**, 316–322 (2003).

[77] Murgida, D. H., E. Schleicher, A. Bacher, G. Richter und P. Hildebrandt. Resonance Raman spectroscopic study of the neutral flavin radical complex of DNA photolyase from Escherichia Coli. *J. Ram. Spec.* **32**, 551–556 (2001).

[78] Schleicher, E., G. Richter und A. Bacher. IR spectra of Photolyase with DNS substrate bound (2000). Unpublished.

[79] Allin, C., M. R. Ahmadian, A. Wittinghofer und K. Gerwert. Monitoring the GAP catalyzed H-Ras GTPase reaction at atomic resolution in real time. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7754–7759 (2001).

[80] Martin, C. B., M.-L. Tsao, C. M. Hadad und M. S. Platz. The Reaction of Triplet Flavin with Indole. A Study of the Cascade of Reactive Intermediates Using Density Functional Theory and Time Resolved Infrared Spectroscopy. *J. Am. Chem. Soc.* **124**, 7226–7234 (2002).

[81] Babizka, G. *Dichtefunktionaltheorie kombiniert mit einem molekülmechanischen Kraftfeld: Die Berechnung der Schwingungsspektren des Retinalchromophores von Bakteriorhodopsin*. Diplomarbeit, Ludwig-Maximilians Universität München, Lehrstuhl für Biomolekulare Optik, AG Theoretische Biophysik (2004).

[82] Hohenberg, P. und W. Kohn. Inhomogeneous Electron Gas. *Phys. Rev. B* **136**, 864–870 (1964).

[83] Kohn, W. und L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev. A* **140**, 1133–1138 (1965).

[84] Neugebauer, J. und B. A. Hess. Fundamental vibrational frequencies of small polyatomic molecules from density-functional calculations and vibrational perturbation theory. *J. Chem. Phys.* **118**, 7215–7225 (2003).

[85] Nielsen, H. H. The Vibration-Rotation Energies of Molecules. *Rev. Mod. Phys.* **23**, 90–136 (1951).

[86] Ahlrichs, R., M. Bär, M. Häser, H. Horn und C. Kölmel. Electronic Structure Calculations on Workstation Computers: The Program System TURBOMOLE. *Chem. Phys. Lett.* **162**, 165 (1989).

[87] Frisch, M. *et al*. GAUSSIAN 98 *Revision A.5*. Gaussian Inc., Pittsburgh PA (1999).

[88] Schmidt, M. W. *et al*. General atomic and molecular electronic-structure system. *J. Comput. Chem.* **14**, 1347–1363 (1993).

[89] Hutter, J., A. Alavi, T. Deutsch, M. Bernasconi, S. Goedecker, D. Marx, T. Tuckermann und M. Parinello. CPMD Version 3.7.2. MPI für Festkörperforschung Stuttgart and IBM Zürich Research Laboratory.

[90] Mathias, G. *Anwendung von QM/MM Hybridmethoden zur Berechnung der Schwingungsspektren von biologischen Chromophoren in situ*. Diplomarbeit, Ludwig-Maximilians Universität München, Lehrstuhl für Biomolekulare Optik, AG Theoretische Biophysik (2000).

[91] Nonella, M., G. Mathias und P. Tavan. The infrared spectrum of 1,4-benzoquinone in water obtained from a hybrid molecular dynamics simulation. *J. Phys. Chem. A* **107**, 8638–8647 (2003).

[92] Warshel, A. und M. Levitt. Theoretical Study of Enzymatic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lyzosyme. *J. Mol. Biol.* **103**, 227–249 (1976).

[93] Eichinger, M., P. Tavan, J. Hutter und M. Parrinello. A hybrid method for solutes in complex solvents: Density functional theory combined with empirical force fields. *J. Chem. Phys.* **110**, 10452–10467 (1999).

[94] Cui, Q. und M. Karplus. Molecular properties from combined QM/MM methods. I. Analytical second derivative and vibrational calculations. *J. Chem. Phys.* **112**, 1133–1149 (2000).

[95] Klähn, M., G. Mathias, C. Kötting, M. Nonella, J. Schlitter, K. Gerwert und P. Tavan. IR spectra of phosphate ions in solution: Predictions of a DFT/MM approach compared with observations. *J. Phys. Chem. A* **108**, 6186–6194 (2004).

[96] Stratt, R. M. The Instantaneous Normal Modes of Liquids. *Acc. Chem. Res.* **28**, 201–207 (1995).

[97] Keyes, T. Instantaneous Normal Mode Approach to Liquid State Dynamics. *J. Chem. Phys.* **101**, 2921–2930 (1997).

[98] McQuarrie, D. A. *Statistical Mechanics*. Harper and Row, New York (1976).

[99] Ramirez, R., T. Lopez-Ciudad, P. Kumar und D. Marx. Quantum Corrections to Classical Time-Correlation Functions: Hydrogen Bonding and Anharmonic Floppy Modes. *J. Chem. Phys.* **121** (2004). In print.

[100] Wheeler, R. A., H. Dong und S. E. Boesch. Quasiharmonic Vibrations of Water, Water Dimer, and Liquid Water from Principal Component Analysis of Quantum or QM/MM Trajectories. *Chem. Phys. Chem.* **4**, 382–384 (2003).

[101] Wheeler, R. A. und H. Dong. Optimal Spectrum Estimation in Statistical Mechanics. *Chem. Phys. Chem.* **4**, 1227–1230 (2003).

[102] Giese, B., B. Carl, T. Carl, T. Carell, C. Behrens, U. Hennecke, O. Schiemann und E. Feresin. Excess electron transport through DNA: A single electron repairs more than one UV-induced lesion. *Angew. Chem. Int. Ed.* **43**, 1848–1851 (2004).

[103] Haas, C., K. Kräling, M. Cichon, N. Rahe und T. Carell. Weak Dependence of Excess Electron Transfer in DNA on Electron Transfer Direction. *Angew. Chem. Int. Ed.* **43**, 1842–1844 (2004).

[104] Ripuli. aus dem Lied: „Helsinki is hell“ (1999).

[105] Kubo, R. A Stochastic Theory of Line Shape. *Adv. Chem. Phys.* **15**, 101–127 (1969).

[106] Mukamel, S. *Principles of Nonlinear optical spectroscopy*. Oxford University Press, Oxford (1995).

[107] Kubo, R., M. Toda und N. Hashitsume. *Statistical Physics II - Nonequilibrium Statistical Mechanics*. Springer, Heidelberg (1985).

[108] Oxtoby, D. W. Dephasing of molecular vibrations in liquids. *Adv. Chem. Phys.* **40**, 2–48 (1979).

[109] Press, W. H., S. A. Teukolsky, W. T. Vetterling und B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, zweite Auflage (1992).

[110] Ladanyi, B. M. und R. M. Stratt. Short-Time Dynamics of Solvation: Linear Solvation Theory for Polar Solvents. *J. Phys. Chem.* **99**, 2502–2511 (1995).

[111] Ahlborn, H., B. Space und P. B. Moore. The effect of isotopic substitution and detailed balance on the infrared spectroscopy of water: A combined time correlation function and instantaneous normal mode analysis. *J. Chem. Phys.* **112**, 8083–8088 (2000).

[112] Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev. Lett.* **159**, 98–103 (1967).

[113] Gaigeot, M. P. und M. Sprik. Ab Initio Molecular Dynamics Computation of the Infrared Spectrum of Aqueous Uracil. *J. Phys. Chem. B* **107**, 10344–10358 (2003).

[114] Lawrence, C. P., A. Nakayama, N. Makri und J. L. Skinner. Quantum dynamics in simple fluids. *J. Chem. Phys.* **120**, 6621–6624 (2004).

[115] Nonella, M. *et al*. The infrared spectrum of ubiquinone in water obtained from a hybrid molecular dynamics simulation. In Vorbereitung.

[116] Schrader, T., A. Sieg, F. Koller, W. Schreier, Q. An, W. Zinth und P. Gilch. Vibrational relaxation following ultrafast internal conversion: comparing IR and Raman probing. *Chem. Phys. Lett.* **392**, 358–364 (2004).

[117] Simpson, B. (aus: „The Simpsons", Folge 5F09).

Danksagung

"I WILL NOT MESS WITH THE CLOSING CREDITS"
(nach *B. Simpson* [117])

Alljenen, die zum Gelingen meiner Promotion beigetragen haben, möchte ich an dieser Stelle ganz herzlich danken:

Allen voran meinem Betreuer Paul Tavan, der mich stets mit seinem persönlichen Einsatz und seinen hohen wissenschaftlichen Ansprüchen motiviert und diese Arbeit mit wertvollen Anregungen bereichert hat. Danke auch für die Freiheit der Wahl des Schwerpunkts meiner Arbeit, die harte Schule des wissenschaftlichen Schreibens und die großzügige Finanzierung meiner Promotion im Rahmen des SFB 533.

Marco Nonella gilt mein Dank für seine stete Hilfsbereitschaft und die großzügige Finanzierung meiner Promotion im Rahmen des VW Projekts.

Nicht hoch genug einzuschätzen ist die unübertroffene Arbeitsatmosphäre der Arbeitsgruppe, die von fruchtbaren wissenschaftlichen Diskussionen und von kompromissloser gegenseitiger Unterstützung geprägt ist. Mein ausdrücklicher Dank dafür gilt Bernhard Egwolf, Verena Schultheis, Martina Stork, sowie allen Diplomanden der Arbeitsgruppe.

Zusätzlich dazu gilt meine höchste Anerkennung für ihre hervorragenden Verdienste um ein stabil laufendes Netzwerk verschiedenster Computerplattformen den Computeradministratoren Heiko Carstens, Tassilo Christ, Thomas Hirschberger, Gerald Mathias, Rudolf Reichold, Paul Strodel und Andreas Weiss.

Ein großer Dank gilt auch sämtlichen Mitarbeitern der AG Zinth und der AG Riedle, die für meine laienhaften Fragen zur experimentellen biophysikalischen Forschung stets ein offenes Ohr hatten und mir so einen Blick über meinen theoretischen Tellerrand ermöglichten.

Danke auch an Frau Barbara Podolski, Frau Alexandra Michaelis, Frau Nicole Klemradt und Frau Ines Haame, die mir als Sekretärinnen des Lehrstuhls bzw. des SFB 533 ausnahmslos hilfsbereit und unkompliziert bei allen organisatorischen Tätigkeiten mit Rat und Tat zur Seite gestanden haben.

Für die regelmäßige Erinnerung an ein Leben „da draußen“ bedanke ich mich bei der

Spvgg. Okulatja 2000 mit einem dreifachen Hip-Hip-Hurra!, bei allen anderen, nicht namentlich erwähnten Freunden, vor allem aber bei den Jungs von bank – ROCK 04/06!!!

Unvorstellbar wäre das Zustandekommen dieser Dissertation ohne diejenigen Lieben, die mir stets das Gefühl gaben, dass sie immer, egal was komme, hinter mir stehen. Dank dafür (sowie für sonntäglichen Extremsport, Abendessen und „Tatort“-schauen, äh, -gucken) den vier Rosenheimern, obrigado tambem á minha menina por me suportares (as vezes isso deve ser mais difícil do que ter escrito esta tese), unermeßlicher Dank jedoch zu allererst meinen Eltern.

Lebenslauf

Name:	Matthias Schmitz
Geburtsdatum:	20.07.1976
Geburtsort:	Königswinter

seit 2000	Wissenschaftlicher Mitarbeiter, LMU München
2000	Diplom in Physik
1996 – 2000	Studium der Physik, RWTH Aachen
1995 – 1996	Grundwehrdienst, Budel/NL und Erndtebrück
1994 – 1995	Studium der Physik, RWTH Aachen
1994	Abitur
1986 – 1994	Gymnasium Nonnenwerth, Remagen