Models for the Calculation of Peptide Vibrational Spectra

DISSERTATION
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich
Chemische und Pharmazeutische Wissenschaften
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von
Roman D. Gorbunov
aus Dniprodzerzhinsk

Frankfurt am Main
2007
(DF1)

vom Fachbereich Chemische und Pharmazeutische Wissenschaften der
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan:	Prof. Dr. Harald Schwalbe
1. Gutachter:	Prof. Dr. Gerhard Stock
2. Gutachter:	Prof. Dr. Josef Wachtveitl
Datum der Disputation:	....

Contents

1 Introduction 11

2 Ab initio models of amide I vibrations 17

- 2.1 Exciton Model . . . 17
  - 2.1.1 Terms Definition . . . 17
  - 2.1.2 Choice of Theory Level and Basis Set . . . 19
- 2.2 Parameterization Schemes . . . 23
  - 2.2.1 Finite Difference Method . . . 24
  - 2.2.2 Construction of Amide I Local Modes . . . 24
  - 2.2.3 Step for the Numerical Differentiation . . . 32
  - 2.2.4 Hessian Matrix Reconstruction Method . . . 36
  - 2.2.5 Usage of the NMA-Based Local Modes . . . 38
  - 2.2.6 Carbonyl Coordinate Displacement Method . . . 41
  - 2.2.7 Comparison of the Parameterization Schemes . . . 42
- 2.3 Amide I Anharmonicities . . . 47
  - 2.3.1 Anharmonicities in NMA . . . 47
  - 2.3.2 Anharmonicities in the GD: Conformational Dependency . . . 51
  - 2.3.3 Explanation of the Disagreement Between FD and HMR Methods . 54
  - 2.3.4 Effect of the Anharmonicities on the Transition Frequencies . . . 56
- 2.4 Building Block Model . . . 56
  - 2.4.1 Methods . . . 59
  - 2.4.2 First-Neighbor Couplings . . . 59
  - 2.4.3 Second-Neighbor Coupling . . . 61
  - 2.4.4 Site Energies of the Terminal Residues . . . 62
  - 2.4.5 Site Energies of the Inner Peptide Unit . . . 63
  - 2.4.6 Concluding remarks . . . 64

- 2.5 Transferability of Maps 64
  - 2.5.1 Effect of Side Chains 64
  - 2.5.2 Effect of End Groups (Different Protonation States) 66
  - 2.5.3 Comparison of GD and AAA Maps 66

3 Calculation of Infrared Spectra 69

- 3.1 Trialanine 69
  - 3.1.1 Semiclassical Line Shape Theory 70
  - 3.1.2 Solvent-Induced Frequency Shift 73
  - 3.1.3 Distributions of Vibrational Frequencies 74
  - 3.1.4 Calculation of the Absorption Spectrum 77
  - 3.1.5 Summary 83
- 3.2 Photoswitchable Bicyclic Azobenzene Octapeptide 84
  - 3.2.1 Methods 84
  - 3.2.2 frequencies Distributions 86
  - 3.2.3 Frequencies Correlation Functions 90
  - 3.2.4 Spectra 99

Conclusions 104

References 109

Acknowledgments 118

Zusammenfassung 120

Lebenslauf 124

Publikationen 125

List of Figures

1.1 "Glycine dipeptide" (GD) 14

1.2 Scheme and atom labeling of trialanine cation A_3^+. 15

2.1 Comparison of the amide I vibrational coupling k_{12} and force constant k_1 of GD as obtained for $\phi = -60°$ at three levels of theory: Hartree Fock, density functional theory, and Møller-Plesset perturbation theory. 20

2.2 Amide normal mode frequencies of the NMA and GD molecules as functions of the basis set. GD molecule is in the parallel sheet conformation β_P (-119,113). 23

2.3 Scheme of the "glycine dipeptide analog" (GD) molecule, introducing the two local coordinate systems, which are employed to define NMA-based amide I local modes of the system. 25

2.4 The force and coupling constant as function of the step used in the numerical differentiation. 33

2.5 Dependence of the functions $\xi_j\left(\Delta q\right)$ on the accuracy of the minimum definition. 34

2.6 Examples of the functions $\xi_j\left(\Delta q\right)$. 35

2.7 Difference between the exact potential energy and the fitting polynomial of different order. 37

2.8 Force constant calculated by numerical differentiation of the *ab initio* potential energy and the fitting cubic polynomial. 37

2.9 Coupling constants, difference between the force constants, and average force constant calculated by FD and HMR methods. 43

2.10 The coupling and difference between the force constants calculated by the HMR method in combinations with the NMA-based local modes and CCD method. 45

2.11 The coupling constant, difference between the force constant, and the average force constant calculated for the fully optimized and restricted geometries. 46

2.12 Transition frequencies of the first two excited states as well ground state energy as functions of size of the basis set. 48

2.13 Difference between the first two transition frequencies $((E_2 - E_1) - (E_1 - E_0))$ 49

2.14 $(E_2 - E_1) - (E_1 - E_0)$ as a function of the normal mode coordinate range used during the fitting of the amide I potential energy. Different curves correspond to different order of polynomial used during fitting. 50

2.15 Values of the local coordinate which correspond to fully optimized geometry of GD as well as mesh used for calculation of cubic anharmonicities in GD. 52

2.16 Average frequencies, frequency splitting and site energies obtained with the usage of full optimized, restricted and restricted tuned geometries. 55

2.17 Transition frequencies shifted by the anharmonicities are shown as functions of the not shifted values. 57

2.18 Blocked glycine peptides under consideration: **A** N-methylacetamide (NMA), **B** "glycine dipeptide" (GD), and **C** "glycine tripeptide" (GT). 58

2.19 Amide I local-mode frequencies ε_n (panels B and C) and associated vibrational couplings β_{nm} (panels A) of glycine tripeptide. Compared are results obtained directly from DFT calculations ("Reference") and from various approximate schemes ("Model"), see text. 60

2.20 Transferability of the amide I vibrational constants for Ac-Gly-NHCH$_3$ (GD) to peptides with a hydrophilic side chain Ac-Asp-NHCH$_3$ (Asp) and a hydrophobic side chain Ac-Phy-NHCH$_3$ (Phy), respectively. Shown are (in cm^{-1}) the intersite coupling β, the site energies ε_1 and ε_2, as well as the frequency gap $\Delta\omega = \omega_+ - \omega_-$. 65

2.21 Transferability of the amide I vibrational constants for Ac-Gly-NHCH$_3$ (GD) to trialanine in its zwitterionic (A$_3^{+-}$), cationic (A$_3^{+}$), and anionic (A$_3^{-}$) state. Shown are (in cm^{-1}) the intersite coupling β, the site energies ε_1 and ε_2, as well as the frequency gap $\Delta\omega = \omega_+ - \omega_-$. 67

2.22 (ϕ, ψ)-maps of the mean $\bar{\omega}$ (top) and the splitting $\Delta\omega$ (bottom) of the two amide I frequencies, as obtained for isolated glycine dipeptide (left), isolated trialanine (middle), and trialanine in D$_2$O (right). 68

3.1 Distribution of the amide I normal-mode frequencies obtained for the three conformational states of trialanine (a) in the gas phase and (b) in solution. Panel (c) shows the corresponding absorption bands calculated within the cumulant approximation. 75

3.2 Distribution of the frequency splitting $\Delta\omega$ without (left) and with (right) the inclusion the solvent contribution, as obtained for the three conformational states of glycine dipeptide (top) and trialanine (bottom). 76

3.3 Correlation functions $\langle M_k(t)\rangle_s$ of the total transition dipole moment, shown for both amide I normal modes ($k = 1, 2$) and the conformations $s = \alpha_{\mathrm{R}}$, β, and P_{II}. 78

3.4 Amide I absorption bands $\sigma_s(\omega)$ of trialanine obtained for the conformations $s = \alpha_{\mathrm{R}}$, β, and P_{II}. Compared are results calculated directly from semiclassical line shape theory (via Eq. (3.5), thick black lines), by invoking only the adiabatic approximation (via Eq. (3.8, thin red lines), and by invoking adiabatic and cumulant approximations (via Eq. (3.8), blue dashed lines). . 79

3.5 Comparison of experimental (Ref. [1], green dashed line) and calculated amide I absorption spectra of trialanine. The latter were obtained directly from semiclassical line shape theory (via Eq. (3.5), thick black line) and by invoking adiabatic and cumulant approximations (via Eq. (3.8), thin red line). 81

3.6 Structure and amino acid labeling of the bicyclic azobenzene peptide bcAMPB. 84

3.7 Frequency distributions corresponding to the beginning and end of the time evolution of the system. 87

3.8 Dependency of the frequency distribution on time. 88

3.9 Dependencies of the fitting coefficients on frequency. 88

3.10 Correlation between coefficients of the fitting. 89

3.11 Time dependent changes of the frequency distribution. 90

3.12 Time dependent changes of the frequency distribution. The time dependency is calculated with the usage of the fitting functions. 91

3.13 Averaged [over time-segments and MD trajectories] normal mode frequencies as functions of time. 93

3.14 Averaged [over time-segments and MD trajectories] site energies as functions of time. 95

3.15 Contributions to the first tree site energies as functions of time. 96

3.16 *Ab initio* maps of the ε_n combined with the distributions of the dihedral angles ϕ and ψ for the beginning and end of the time evolution. 97
3.17 *Ab initio* maps of the ε_c combined with the distributions of the dihedral angles ϕ and ψ for the beginning and end of the time evolution. 98
3.18 Fitting parameter of the correlation functions of the 8 normal mode frequencies as functions of time. 100
3.19 Fitting parameter of the correlation functions of the normal modes frequencies obtained by the averaging over time (red curves) as well as by linear fit of the time dependencies (blue curves). 101
3.20 Time dependent spectra of photoswitchable peptide. 102
3.21 Changes of the vibrational spectrum of photoswitchable peptide. 102

List of Tables

2.1 Basis-set dependency of the diagonal force constants $k1$ and k_2 and the vibrational coupling k_{12} as obtained from DFT calculations on glycine dipeptide. Data are shown for the following conformations and Ramachandran angles (ϕ, ψ): Parallel β-sheet β_P (-119,113), antiparallel β-sheet β_{AP} (-139,135), right-handed α-helix α_R (-57,-47), and left-handed α-helices α_{L1} (57,47) and α_{L2} (90,-90). The root mean square deviation (RMSD) is computed with respect to the 6-311+G(2df,2p) data, using besides the five listed structures 12 additional conformations with $(\phi, \psi) = (-n\cdot 90, -m\cdot 90)$, with $n, m = 1, 2, 3$. Units are mdyn/Åu. 22

Chapter 1

Introduction

Owing to significant progress in multidimensional infrared (IR) laser techniques, [2, 3] the investigation of the vibrational bands of peptides and proteins has recently received considerable attention. Most studies have focused on the strongly IR-active amide I mode (mainly C=O stretch), which is a sensitive probe of hydrogen bonding, dipole-dipole interactions, and the conformation of the peptide backbone. [4, 5] Amide I modes possess several features which make them suitable conformational probes. The first advantage of the amide I modes is that they are IR-active. This property is conditioned by a large value of the transitional dipole moment associated with the amide I modes. Because of that property the amide I peaks are very high and can be easily distinguished in vibrational spectra. The second advantage of the amide I modes is that they are well separated in frequency from other normal modes. Because of that fact the overlap of the amide I peaks with other ones is quite small. This also facilitate the extraction of the amide I peaks from total vibrational spectra. And finally, the main advantage is that the amide I modes are structurally sensitive. This means that different characteristic of the amide I peaks (such us position, intensity and shape) strongly depend on the conformational structure and dynamics of peptides.

Some information about the structure of peptides can be obtained from one-dimensional infrared spectra. Employing empirical rules that relate frequencies of amide bands to secondary structures, structural motives such as α- and 3_{10}-helices as well as parallel and antiparallel β-sheets have been identified. For example, Schweitzer-Stenner *et al.* have investigated the conformation of small peptides by using a combination of vibrational spectroscopies including Fourier-Transform IR, polarized Raman and vibrational circular dichroism. [6–9] Recently, the advent of multidimensional IR techniques has revealed a wealth of novel and quite detailed information on the structure and dynamics of biomolecules. [2, 10]

Beautiful examples are the two-dimensional infrared studies of various small peptides by Hamm and Hochstrasser and their coworkers. [1, 11–15] The interpretation of these experiments, however, is far more involved than for simple linear IR absorption and clearly requires substantial theoretical support. A first-principle theoretical description gives a relation between the peptides geometry and their vibrational spectra and, as a consequence, is required for the interpretation of the IR-experiments in terms of the conformational structure and dynamics.

A theoretical description of the IR response of a peptide in aqueous solution represents a considerable challenge. Since a direct *ab initio* molecular dynamics description [16–18] of a solvated peptide is in general computationally too expensive, usually some mixed quantum-classical strategy [19–26] is chosen, which contains the following parts. (i) First, we have to perform a state-of-the-art molecular dynamic (MD) simulation of the complete system in order to get the correct thermal populations of the existing conformational states. The MD trajectories has to be sufficiently long to get a comprehensive conformational sampling and to reach the convergence in the statistical averaging of the spectra calculation. (ii) Precise *ab initio* calculations, giving the vibrational properties of the peptide, have to be performed. The large numerical effort of *ab initio* calculations as well as the high flexibility of peptides restricts to the calculation of small peptides. (iii) It is *a priori* not clear which approximations are applicable to calculation of vibrational spectra of a solvated peptide. As a consequence, additional analysis is often required. In this work we are concerned with the latter two parts of the calculation.

To this end, we first calculate the vibrational frequencies of the gas-phase system for all conformational structures visited by the MD trajectory. Since the vibrational frequency splitting of interest are relatively small, their calculation requires accurate *ab initio* methods using large basis sets. Because of that, direct *ab initio* calculations of vibrational frequencies for the entire molecule are restricted to small systems and/or only few conformations. The computational costs dramatically increase with the size of the system. First, the numerical effort of the *ab initio* geometry optimization and normal mode analysis, giving the vibrational properties of the system, exponentially increases with the system size. Moreover, peptides are very flexible molecules (the relative orientation of adjacent peptide units is given by two very flexible dihedral angles) and, as a consequence, the number of possible conformations exponentially depends on the number of peptide units.

The complexity of the *first principle* modeling of the vibrational properties requires a set of approximations. As we have already mentioned, we focused our study on the amide I modes. We assume that — to a first approximation — these modes do not interact with the remaining vibrational degree of freedom of the peptide. This allows us to construct an exciton model in which only the amide I excitations are considered. The assumption of a separable amide I subspace reduces the full system including all degrees of freedom to an N-dimensional vibrational problem, where N is the number of peptide units. Recent *ab initio* calculations of the amide I anharmonic couplings indicate that this basis assumption of separability is surprisingly well fulfilled [27–30]. As the second approximation, we consider only harmonic terms in the amide I potential energy of the systems. The accuracy of the harmonic approximation has been studied in the present work by direct *ab initio* calculations of higher order terms in the amide I potential energy.

In the next approximation, we suppose that all amide I normal modes can be presented as superposition of local vibrations. For the sake of simplicity, the local vibrations can be imagined as the C=O stretch vibrations of the corresponding peptide unit. The local modes, as a basis for the amide I vibrational motion, are used in the construction of the vibrational Hamiltonian of polypeptides. The construction of the polypeptides Hamiltonian, called “building block model”, is based on the assumption that the local vibrational properties of a polypeptide depend on the local conformational structure in the same way like in smaller peptides (called building blocks). The necessity to use the building block model is due to the fact that a direct *ab initio* calculation of the vibrational properties is possible only for small peptides and/or several conformations.

To parameterize the vibrational Hamiltonian of the building block, we have to perform *ab initio* calculations of its vibrational properties. In the case of dipeptides the dependency of the vibrational properties on the dihedral angles ϕ and ψ, giving conformation of the system, can be found directly (so called *ab initio* parameterization of the Hamiltonian). Several groups have considered the amide I vibrations of Ac-Gly-NHCH$_3$ (CH$_3$-CONH-CH$_2$-CONH-CH$_3$), often referred to as glycine dipeptide (GD) [28, 31–34] (see Fig. 1.1). Employing various approaches, the off-diagonal vibrational coupling between the two peptide units as well as the diagonal force constants have been calculated as a function of the (ϕ, ψ) dihedral angles of the peptide backbone. To calculate these maps, several choices of local amide I modes have been suggested, including the C=O stretch vibrations and the normal modes of N-methylacetamide (CH$_3$-CONH-CH$_3$). Various ways to calculate the vibrational couplings and force constants for a given set of local modes have been proposed, including finite-difference differentiation [31] and the so-called Hessian matrix reconstruction method. [28, 32] It has to be noted that the question of the Hamiltonian

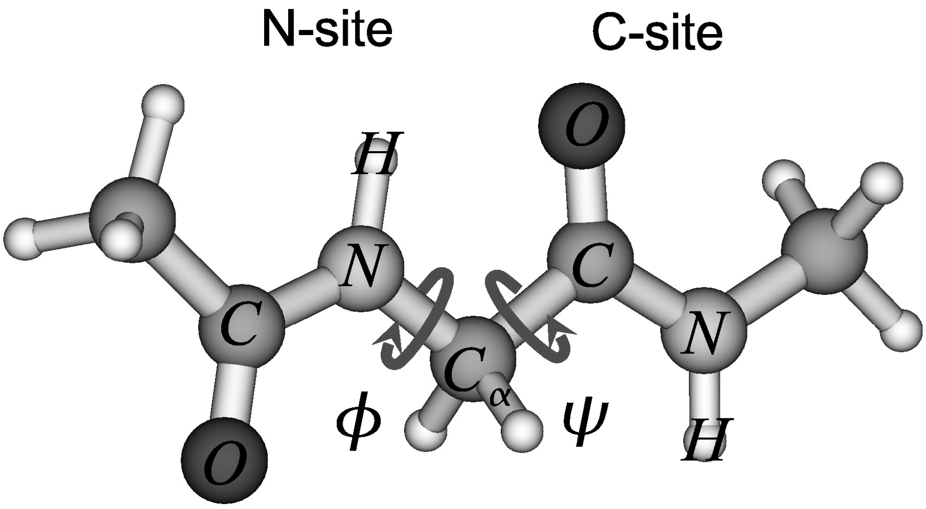

Figure 1.1: "Glycine dipeptide" (GD)

parameterization is not trivial. Different parameterization schemes were found to give qualitatively different force constant. Therefore a more detailed investigation of the parameterization is required and will be performed in this work. Next question which is related with the parameterization of dipeptides is to which extend *ab inito* maps of coupling and force constants obtained for one dipeptide are transferable to other ones. In particular it would be interesting to study to what extent the (ϕ, ψ) maps are transferable to peptide with other side chains or other end-groups. In the present work we answer these questions.

To complete the theoretical modeling of the vibrational Hamiltonian of the system, we have to take the solvent effects into account. The solvent-induced frequency shift can be obtained via electrostatic models based on empirical relations between the electric field produced by the surrounding solvent and the induced amide I frequency shift. Since the influence of the solvent on the vibrational spectra is considerable, an accurate parameterization of the frequency shift is required. [28, 29, 35–42]. This consists of the following steps. To this end, representative MD snap shots of a peptide unit including the first solvation shell are adopted and geometry optimization of the peptide with fixed solvent and a subsequent normal mode analysis are performed. Based on these data an empirical expression relating the amide I frequency shift to the values of the electric potentials at the atoms of the peptide unit is derived, which can be used to calculate the solvent induced frequency shift directly from the MD trajectories. Another way to include solvent effects is

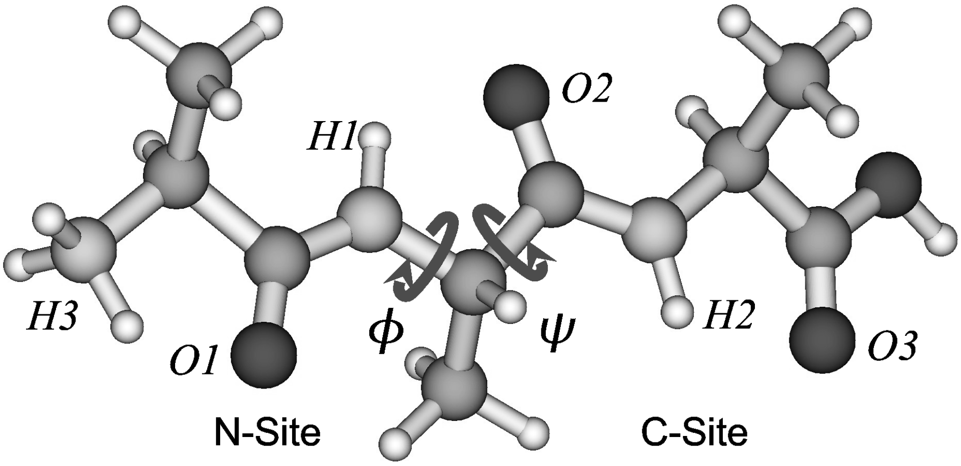

Figure 1.2: Scheme and atom labeling of trialanine cation A_3^+.

a direct calculation of the frequency shift via *ab initio* calculations of MD snapshots of the solvated peptide including the surrounding water. However, in many cases this strategy is not applicable, since the MD trajectory has to be sufficiently long to reach convergence in the spectra calculations.

The construction of the vibrational Hamiltonian is the first of the two main parts of the present work. In the second part, we implement the Hamiltonian to the calculation of vibrational spectra. The main goal of this part is to study the applicability of different spectroscopic approximation to some particular systems as well as to demonstrate a practical way to relate a time-dependency of vibrational spectra to conformational changes.

Assuming that the fluctuations of the peptide and the surrounding solvent molecules result in a classical time-dependence of the vibrational frequencies and dipoles moments, the spectral line shape of the system is calculated using semiclassical line shape theory. [43, 44] However, in many cases the applicability of spectroscopic approximations (such as the Franck-Condon approximation [45], the second-order cumulant expansion, [46, 47] and the adiabatic approximation. [20,34,48]) has to be studied additionally. In the present work, we analyzed the accuracy of the mentioned approximations for the cationic trialanine (A_3^+). Trialanine is a small peptide with two peptide bonds and one set of backbone dihedral angles (ϕ, ψ), see Fig. 1.2. Driven by a number of experimental [1, 6, 13–15, 49] and theoretical [20, 50–54] studies, trialanine has emerged as a paradigm to study conformational dynamics of a small peptide in aqueous solution.

The second system considered in the present work, is a photoswitchable bicyclic azobenzene octapeptide. The considered system is an example of a molecular photoswitch included into biomolecule. The considered system has been well characterized experimentally [55–59]. In photoswitchable peptides, the light-induced structural changes of the chromophore upon photoisomerization around the central N=N double bond are directly transferred into the peptide chain. By photoexciting the system by an ultrashort laser pulse, the subsequent conformational dynamics of the peptide is investigated by optical [56, 57] or infrared [58, 59] spectroscopy. These types of experiments, especially in combination with two-dimensional infrared probing [60], provide a new and promising way to study the folding and unfolding of peptides in unprecedented detail. This system is also interesting from a theoretical point of view. First, we can test the building block model on a real system. Second, we show a clear example of how the vibrational Hamiltonian can be used for the interpretation of the time-dependent vibrational spectra in terms of conformational changes. Third, we demonstrate that even weak spectroscopic changes can be, in principle, reproduced with the formulated vibrational model. An finally, we develop a strategy of the treatment of non-equilibrium processes.

Chapter 2

Ab initio models of amide I vibrations

2.1 Exciton Model

2.1.1 Terms Definition

As we have explained in Introduction the vibrational exciton model is a suitable choice for the description of the amide I vibrations in polypeptides. Let us summarize the main definitions associated with this model. For simplicity, we restrict the discussion to the special case of a peptide with two interacting amide I vibrations. Within the harmonic approximation, the Hamiltonian of two coupled oscillators can be written as

$$H = \frac{p_1^2}{2} + \frac{p_2^2}{2} + \frac{1}{2}k_1 q_1^2 + \frac{1}{2}k_2 q_2^2 + k_{12} q_1 q_2 \,, \tag{2.1}$$

where q_1 and q_2 are mass-weighted local modes residing on the first and second peptide unit, respectively, and p_1 and p_2 are the corresponding conjugate momenta. The potential energy is characterized by the force constants k_1 and k_2 of the local modes as well as by the bilinear vibrational coupling k_{12}. In this work we are not concerned with anharmonic potential terms of the exciton model, which have been introduced by several authors on an empirical [11] as well as on an *ab initio* [29] level, respectively.

To consider the normal modes $Q_\mp$ of the model, we introduce the unitary transformation

$$\begin{pmatrix} Q_- \\ Q_+ \end{pmatrix} = \begin{pmatrix} \cos\Theta & \sin\Theta \\ -\sin\Theta & \cos\Theta \end{pmatrix} \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}, \tag{2.2}$$

with the mixing angle

$$\Theta = \frac{1}{2}\arctan[2k_{12}/(k_1 - k_2)], \tag{2.3}$$

which diagonalizes Hamiltonian (2.1). This yields

$$H = \frac{P_-^2}{2} + \frac{P_+^2}{2} + \frac{1}{2}K_- Q_-^2 + \frac{1}{2}K_+ Q_+^2, \tag{2.4}$$

where Q_- and Q_+ are the two amide I normal modes with corresponding momenta P_- and P_+ and the force constants

$$K_\mp = \tfrac{1}{2}(k_1 + k_2) \mp \frac{1}{2}\sqrt{(k_1 - k_2)^2 + 4k_{12}^2}. \tag{2.5}$$

To make contact with the state representation commonly used in exciton theory, we introduce harmonic-oscillator creation and annihilation operators

$$b_j^\dagger = (k_j^{1/4} q_j - \mathrm{i} k_j^{-1/4} p_j)/\sqrt{2\hbar}, \tag{2.6}$$

$$b_j = (k_j^{1/4} q_j + \mathrm{i} k_j^{-1/4} p_j)/\sqrt{2\hbar}, \tag{2.7}$$

where $j = 1, 2$. The operators $b_j^\dagger$ and b_j create and destruct a localized vibration in the jth peptide unit with frequency $\sqrt{k_j}$, respectively, and satisfy the bosonic commutation relations $[b_i, b_j^\dagger] = \delta_{ij}$. Insertion into Eq. (2.1) yields

$$H = \varepsilon_1 b_1^\dagger b_1 + \varepsilon_2 b_2^\dagger b_2 + \beta\left(b_1^\dagger b_2 + b_2^\dagger b_1 + b_1^\dagger b_2^\dagger + b_2 b_1\right), \tag{2.8}$$

where

$$\varepsilon_i = \hbar\sqrt{k_i}, \tag{2.9}$$

$$\beta = \frac{\hbar k_{12}}{2(k_1 k_2)^{1/4}} \tag{2.10}$$

denote the energy of the ith site and the intersite coupling, respectively. Neglecting the nonresonant terms $b_1^\dagger b_2^\dagger$ and $b_2 b_1$, Hamiltonian (2.8) reduces to the Frenkel exciton model [44]

$$H = \begin{pmatrix} b_1 & b_2 \end{pmatrix} \begin{pmatrix} \varepsilon_1 & \beta \\ \beta & \varepsilon_2 \end{pmatrix} \begin{pmatrix} b_1^\dagger \\ b_2^\dagger \end{pmatrix}, \tag{2.11}$$

which conserves the number of excitations. Diagonalizing Hamiltonian (2.11), we obtain the normal mode frequencies

$$\Omega_\mp = \tfrac{1}{2}(\varepsilon_1 + \varepsilon_2) \mp \frac{1}{2}\sqrt{(\varepsilon_1 - \varepsilon_2)^2 + 4\beta^2}. \tag{2.12}$$

Note that the frequencies $\Omega_\mp$ obtained from Eq. (2.12) and $\omega_\mp = \hbar K_\mp$ obtained from Eq. (2.5) are not exactly the same, which is a consequence of neglecting the nonresonant terms $b_1^\dagger b_2^\dagger$ and $b_2 b_1$ in Eq. (2.8). For amide I vibrations with $\beta/\varepsilon \ll 1$, these deviations are quite small, though, and can be safely neglected.

The system described by the Hamiltonian (2.11) separates into blocks of the ground state, the one-exciton Hamiltonian, the two-exciton Hamiltonian, and higher terms (which are not of interest for the description of 2D-IR spectra). This results into the following model (ψ, ϕ)-dependent Hamiltonian

$$H(\psi,\phi) = \begin{pmatrix} 0 & & & & & \\ & \epsilon_1 & \beta(\psi,\phi) & & & \\ & \beta(\psi,\phi) & \epsilon_2 & & & \\ & & & 2\epsilon_1 - \Delta_{11} & 0 & \sqrt{2}\beta(\psi,\phi) \\ & & & 0 & 2\epsilon_2 - \Delta_{22} & \sqrt{2}\beta(\psi,\phi) \\ & & & \sqrt{2}\beta(\psi,\phi) & \sqrt{2}\beta(\psi,\phi) & \epsilon_1 + \epsilon_2 - \Delta_{12} \end{pmatrix} \tag{2.13}$$

In the above written Hamiltonian we have have additionally introduced the anharmonic terms Δ_{ij}. The anharmonicity of the amide I vibratio is very weak. The experimentally measured difference between $(0 \rightarrow 1)$ and $(1 \rightarrow 2)$ absorption frequencies in NMA [11] is 16 cm^{-1} , and given that ω=1600 cm^{-1} the ration $\Delta/\omega = 0.01$. However, the anharmonic terms must be taken into account since the nonlinear response of a harmonic system vanishes exactly, and anharmonicity is a pivotal conditions for 2D IR spectroscopy.

2.1.2 Choice of Theory Level and Basis Set

The density function theory (DFT) was selected for all further ab initio calculations since they are less demanding than other post-Hartree-Fock methods but yields the comparable results for peptides. Of all available functional, the gradient-corrected hybrid functionals provide the best overall coherence with MP2 calculations and experimental data. All calculations were performed using the Gaussian98 [61] and Gaussian03 [62] programs at the B3LYP level of theory. The notation B3 indicates a three-parameter Becke exchange functional in which a part of exchange contribution has been calculated in the same fashion as that in Hartree-Fock approximation but using Kohn-Sham orbitals instead Hartree-Fock ones [63, 64]. LYP indicates the Lee-Yang-Parr correlation functional [65].

To study the importance of electron correlation for the calculation of amide I vibrations in peptides and check the selected level of theory, we compare computations of GD obtained for three levels of theory: Hartree Fock (HF), density functional theory (DFT) with B3LYP functional, [64] and Møller-Plesset perturbation theory (MP2). In all cases, the 6-31+G(d) basis set is used. As a representative example, Fig. 2.1 shows the conformational dependency of the force constants k_1, k_2 and the vibrational coupling k_{12} as a function of the backbone dihedral angle ψ. The vibrational constants are calculated for NMA-based

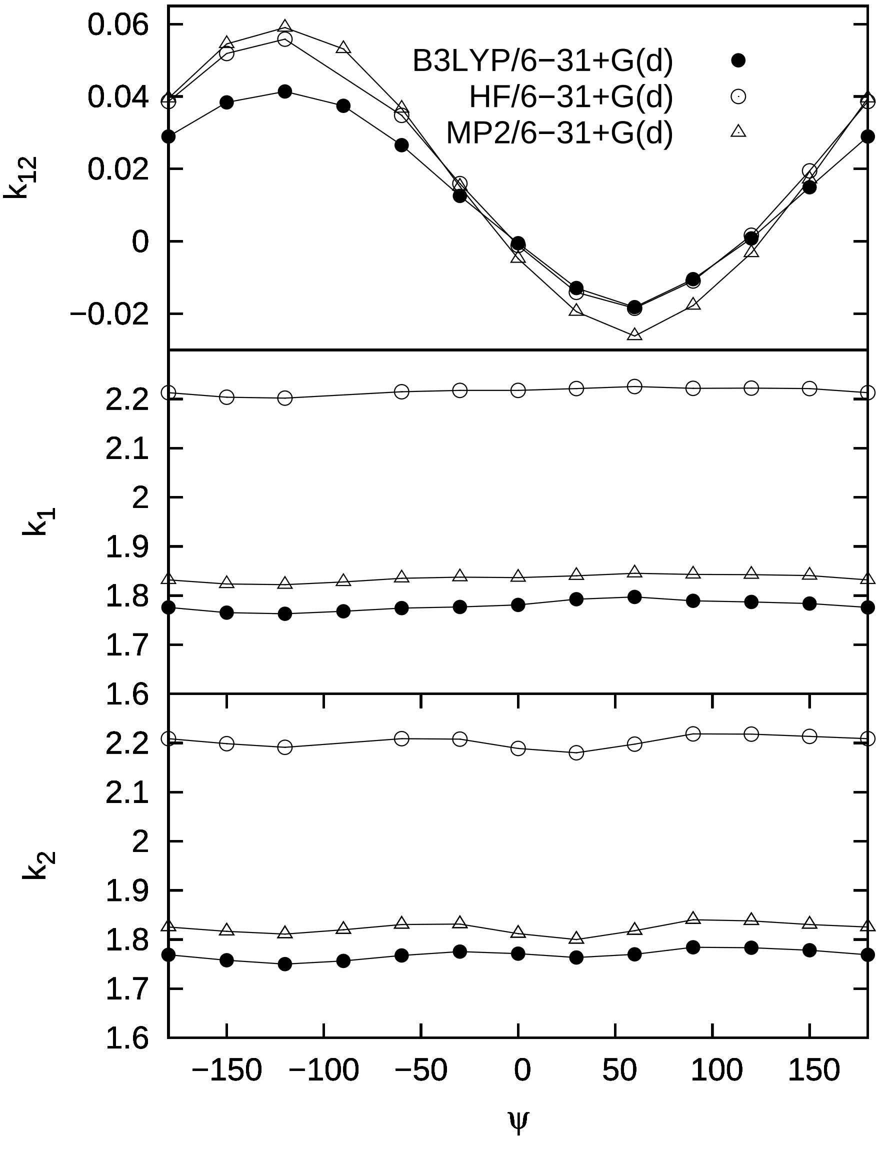

Figure 2.1: Comparison of the amide I vibrational coupling k_{12} and force constant k_1 of GD as obtained for $\phi = -60°$ at three levels of theory: Hartree Fock, density functional theory, and Møller-Plesset perturbation theory.

local modes, using the finite-differences method. Choosing $\phi = -60°$, the plots monitors the transition from a α_R-helical structure at $\psi = -60°$ to a poly(Gly)II (P_{II}) structure at $\psi = 140°$.

The diagonal and off-diagonal vibrational constants are seen to depend in a different way on the theory level and on the peptide conformation, respectively. The vibrational coupling k_{12} is only little affected by the level of theory chosen, but is a sensitive probe of the peptide conformation. The force constants k_1 and k_2, on the other hand, vary only weakly as a function of ψ. Depending on the level of theory, the curves are significantly shifted to each other, thus reflecting the degree of electron correlation taken into account by the method. Employing the recommended frequency scale factors (0.90 for HF, 0.99 for B3LYP, and 0.94 for MP2), however, all three levels of theory give similar results for the force constants.

To study the basis-set dependency of the amide I vibrations of GD, we chose DFT/B3LYP as level of theory. Table 2.1 compares the vibrational constants for selected conformations of GD (parallel and antiparallel β-sheet as well as right- and left-handed α-helices), as obtained for eight different basis sets from 6-31G(d) to 6-311++G(3df,2pd). Furthermore the root mean square deviation (RMSD) of each basis set is given with respect to the 6-311++G(3df,2pd) data, using in total 17 conformations (see caption). On the basis of RMSD results, two main observation can be made. First, it is noted that already with the smallest basis set 6-31G(d) the vibrational coupling k_{12} is given with a relative accuracy of ≈ 1 %. The force constants k_1 and k_2, on the other hand, are found to be much more sensitive to the choice of the basis set. Second, while the addition of polarization functions does not significantly improve matters (compare, e. g., 6-31G(d) and 6-31G(d,p) results), even moderate basis sets supplemented by diffuse s and p orbitals to C, N and O atoms yield results that are comparable with the 6-311++G(3df,2pd) reference calculations. In the Fig. 2.2 We have plotted amide I normal mode frequencies of the NMA and GD molecules as functions of the basis set.

Due to these findings, DFT calculations using a 6-31G+(d) basis set and the B3LYP functional are chosen as a compromise between high accuracy and low computational effort. Based on the observation that the combination of the B3LYP functional with a 6-31+G(d) basis set leads to maximally planar peptide group in NMA, this level of theory was also recently employed to compute amide vibrations of the alanine and glycine dipeptides. [34, 66]

Table 2.1: Basis-set dependency of the diagonal force constants $k1$ and k_2 and the vibrational coupling k_{12} as obtained from DFT calculations on glycine dipeptide. Data are shown for the following conformations and Ramachandran angles (ϕ, ψ): Parallel β-sheet β_P (-119,113), antiparallel β-sheet β_{AP} (-139,135), right-handed α-helix α_R (-57,-47), and left-handed α-helices α_{L1} (57,47) and α_{L2} (90,-90). The root mean square deviation (RMSD) is computed with respect to the 6-311+G(2df,2p) data, using besides the five listed structures 12 additional conformations with $(\phi, \psi) = (-n \cdot 90, -m \cdot 90)$, with $n, m = 1, 2, 3$. Units are mdyn/Åu.

basis set		β_P	β_{AP}	α_R	α_{L1}	α_{L2}	RMSD
6-31G(d)	k_{12}	0.00676	0.01079	0.01792	0.01793	-0.00424	$5 * 10^{-5}$
	k_1	1.8667	1.8697	1.8551	1.8550	1.8667	0.2
	k_2	1.8714	1.8721	1.8537	1.8538	1.8702	0.2
6-31G(d,p)	k_{12}	0.00636	0.01036	0.01791	0.01790	-0.00455	$6 * 10^{-5}$
	k_1	1.8615	1.8645	1.8500	1.8502	1.8617	0.1
	k_2	1.8667	1.8660	1.8492	1.8489	1.8659	0.1
6-31+G(d)	k_{12}	0.00637	0.01073	0.01999	0.01999	-0.00453	$3 * 10^{-6}$
	k_1	1.7827	1.7868	1.7740	1.7746	1.7852	0.003
	k_2	1.7901	1.7918	1.7707	1.7703	1.7884	0.003
6-31+G(d,p)	k_{12}	0.00574	0.01008	0.01999	0.01987	-0.00461	$6 * 10^{-6}$
	k_1	1.7795	1.7814	1.7701	1.7705	1.7807	0.002
	k_2	1.7859	1.7879	1.7666	1.7667	1.7836	0.002
6-311+G(d,p)	k_{12}	0.00619	0.01036	0.02088	0.02069	-0.00434	$3 * 10^{-6}$
	k_1	1.7635	1.7667	1.7556	1.7553	1.7666	0.0005
	k_2	1.7708	1.7723	1.7527	1.7520	1.7694	0.0006
6-311+G(2d,p)	k_{12}	0.00672	0.01082	0.02044	0.02026	-0.00352	$8 * 10^{-7}$
	k_1	1.7456	1.7482	1.7376	1.7379	1.7481	0.009
	k_2	1.7542	1.7553	1.7352	1.7341	1.7519	0.009
6-311+G(2df,2p)	k_{12}	0.00651	0.01048	0.0204	0.02024	-0.00370	$1 * 10^{-7}$
	k_1	1.7579	1.7618	1.7511	1.7504	1.7611	0.002
	k_2	1.7669	1.7681	1.7473	1.7468	1.7650	0.002
6-311++G(3df,2pd)	k_{12}	0.00654	0.01051	0.02037	0.02028	-0.00379	0.0
	k_1	1.7686	1.7719	1.7612	1.7612	1.7709	0.0
	k_2	1.7774	1.7778	1.7580	1.7572	1.7755	0.0

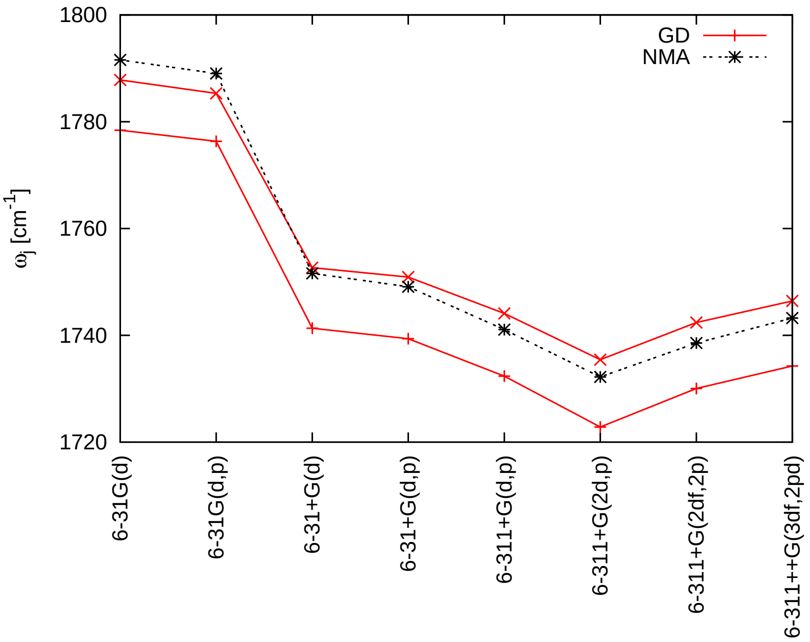

Figure 2.2: Amide normal mode frequencies of the NMA and GD molecules as functions of the basis set. GD molecule is in the parallel sheet conformation β_P (-119,113).

2.2 Parameterization Schemes

The fist part of the present work was dedicated to the calculation of the parameters of the vibrational Hamiltonian (2.1) or (2.11). The ab initio parameterization of the Hamiltonian is based on one of two main ideas. The first one uses definitions of the force and coupling constants (k_j and k_{ij}, respectively) as second order derivatives of the potential energy over the corresponding vibrational coordinates. If the potential energy is known as a function of the local coordinates, the force and coupling constants can be calculated by finite-difference differentiation [31]. The second method is based on the unitary relation between the vibrational properties of the normal and local modes. In this method the site energies ε_j of the local oscillations and coupling β can be calculated if the normal mode frequencies ω_j and the localization angle Θ (see formula (2.3)) are known. The ω_j and Θ can be obtained from a common normal mode analysis. This method is known as Hessian matrix reconstruction [28, 32]

We have found that the two parameterization schemes give a qualitatively different conformational dependencies of the force constants and, as a consequence, site energies. Thereby the separate part of the project was dedicated to the comparison of the two mentioned parameterization schemes and understanding of the reasons of the disagreement.

2.2.1 Finite Difference Method

In this section we describe the finite difference method which was one of the two parameterization schemes analyzed and used in the present work. According to the vibrational Hamiltonian written in the form (2.1), the force and coupling constants can be expressed as the second order derivatives of the potential energy over the local coordinates q_j:

$$k_j = \frac{\partial^2 E}{\partial q_j^2}, \quad k_{ij} = \frac{\partial^2 E}{\partial q_i \partial q_j}. \tag{2.14}$$

In more details the method performs as follows.

1. We define the local modes (shifts of atoms of corresponding peptide unit) as functions of the local coordinates q_j.
2. For different values of the local coordinates (and as a consequence different positions of atoms) we perform ab initio calculations of the potential energy.
3. The obtained potential energy as a function of the local coordinates is used for the calculation of the second-order derivatives (via the usage of finite-difference formula).

All the above mentioned steps are considered in more details further.

2.2.2 Construction of Amide I Local Modes

As we have already mentioned, in order to apply the finite difference method for the calculation of the force and coupling constants we have first to define the local modes. In other words we have to specify dependecy of the GD geometry on the local coordinates q_j. For simplicity of the model, the local modes are defined to be independent on the normal modes, peptide units and conformations of the peptide. The independence on the normal modes means that for different normal modes we can use the same set of the local modes to construct an appropriate linear combination representing the given normal mode. The independence on the peptide unit allows us to treat all local modes in the same manner. And finally, because of the independence of the local modes on the conformation we can keep the conformational dependency of the Hamiltonian in its parameters ε_j and β_{ij} (and not, for example, in the creation and annihilation operators).

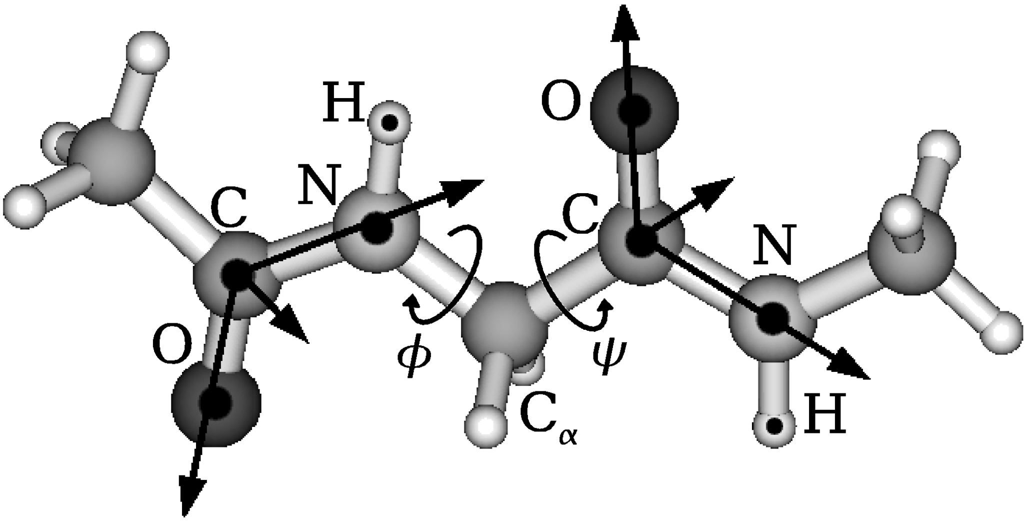

Figure 2.3: Scheme of the "glycine dipeptide analog" (GD) molecule, introducing the two local coordinate systems, which are employed to define NMA-based amide I local modes of the system.

One of the simplest way to define the local modes is to use the amide I normal mode of the NMA molecule which is often considered as a model of the single peptide unit. In other words we project the amide I normal mode motion of the NMA molecule into peptide units of GD. In more details the projection procedure can be defined as follows.

1. We begin with a full geometry optimization and a subsequent normal mode analysis of the NMA molecule. Based on these calculations, the relative positions of atoms in the CONH peptide units and the methyl groups of GD are constrained to their NMA values. Moreover, the Ramachandran angles (ϕ, ψ) are fixed to represent the peptide conformation of interest.

2. A partial geometry optimization of GD with fixed (ϕ, ψ), CONH peptide units, and the methyl groups is performed, i. e., only the coordinates of two hydrogen atoms bonded to C_α and the lengths of the bonds N-C_α and C_α-C are optimized.

3. The positions of atoms in the peptide units of GD are given in terms of the two local coordinate systems shown in Fig. 2.3. For each CONH peptide unit, the origin of the coordinate system is located at the C atom, the first two axes are given by the C-O and C-N bonds, and the third axis is obtained by the a cross product of the first two basis vectors. Since the CONH peptide units are constrained to their NMA

geometry, it is straightforward to project amide I vibrations of NMA into peptide units of GD.

Let us consider the construction of the local modes on basis of the output of the normal mode analysis generated by the Gaussian program [61, 62]. In particular we will consider question of the local modes normalization.

We start from the terms definition. Matrix of transition from normal coordinates to Cartesian displacements will be denoted as $\hat{l}^m_{CART}$. Upper indexes m indicates that columns of the matrix are orthonormalized in the mass-weighted coordinates *i.e.*

$$\sum_{i=1}^{3N} m_i l^{m\ 2}_{CART\ i,j} = 1, \tag{2.15}$$

where m_i are atomic masses and N is number of atoms in the molecule. Here we need to emphasize that columns of the introduced matrix are not orthogonal in the Cartesian coordinates since orthogonalization cannot be fulfilled in both Cartesian and mass-weighted coordinates. According to the definition, each column of $\hat{l}^m_{CART}$ is a normal mode given in Cartesian coordinates and orthonormalized in the mass-weighted ones. Or in other words, the n-th column is Cartesian displacements of atoms (with respect to their equilibrium positions) at n-th normal coordinate equal to 1 and other normal coordinates are zeros. To obtain Cartesian displacements corresponding to a given set of values of the normal coordinates one needs to multiply the considered matrix $\hat{l}^m_{CART}$ by column consisting of the values of the normal coordinates.

The introduced matrix is not printed by the Gaussian program. However, in the output files one can find a matrix obtained by normalization of the $\hat{l}^m_{CART}$. The normalization procedure is relatively straightforward. Each of the $3N$ elements of a given column of the l^m_{CART} is scaled by a normalization factor N_i, which is defined as follows

$$N_i = \sqrt{\left(\sum_{k}^{3N} l^m_{CART\ k,i}\right)^{-1}}. \tag{2.16}$$

In other words, each element of a given column i of l^m_{CART} is multiplied by the factor N_i. After this procedure columns of the new matrix

$$l^c_{CART\ i,j} = l^{m\ 2}_{CART\ i,j} N_j \tag{2.17}$$

are normalized in the Cartesian coordinates

$$\sum_{i=1}^{3N} l^{c\ 2}_{CART\ i,j} = 1. \tag{2.18}$$

Upper index c is used to indicate that columns of the matrix are normalized in Cartesian coordinates. One needs to emphasize that modes consisted in $l^c_{CARTi,j}$ are normalized but still not orthogonal in Cartesian coordinates.

It will be observed that the above introduced normalization factor is related with the so called reduced masses μ_i of the normal modes (μ_i can also be found in the Gaussian output files)

$$\mu_i = N_i^2 = \left(\sum_k^{3N} l^{m\ 2}_{CART\ k,i} \right)^{-1}. \tag{2.19}$$

Therefore, in order to calculated the Cartesian displacements corresponding to some values of the normal coordinates n_j, matrix l^c_{CART} printed out by the Gaussian should be multiplied by column consisting of the normal coordinates divided by the corresponding reduced masses

$$\Delta x_i = \sum_{j=1}^{3N} l^c_{CART\ i,j} \left(\frac{n_j}{\sqrt{\mu_j}} \right) = \sum_{j=1}^{3N} l^m_{CART\ i,j} n_j. \tag{2.20}$$

So, we have the explicit definition of the reduced mass (2.19). However, the reduced masses are defined through Cartesian displacements of the normal modes which are normalized in the mass-weighted coordinates. Let us find an expression which does not rely on this specific property and can be used in the general case. Using the definition of the reduced masses (2.19) as well as its relation with normalization coefficients, we can obtain the following relation involving the reduced mass

$$l^c_{CART\ i,j} = l^m_{CART\ i,j} \sqrt{\mu_j}. \tag{2.21}$$

The equation (2.21) can be considered as another definition of the reduced masses. Using (2.21) and (2.15) it is easy to express μ_j through atomic masses and Cartesian displacements

$$\sum_{i=1}^{3N} m_i l^{m\ 2}_{CART\ i,j} = 1 = \frac{1}{\mu_j} \sum_{i=1}^{3N} m_i l^{c\ 2}_{CART\ i,j} \tag{2.22}$$

$$\mu_j = \sum_{i=1}^{3N} m_i l^{c\ 2}_{CART\ i,j}$$

Further, we express the displacement normalized in the Cartesian coordinates through arbitrary ones

$$l^c_{CARTj,j} = \frac{l_{CART\ i,j}}{\sqrt{\sum_{i=1}^{3N} l^2_{CART\ i,j}}}. \tag{2.23}$$

Substituting (2.23) in (2.22) we obtain the final result

$$\mu = \frac{\sum_{i=1}^{3N} m_i \, l^2_{CART \ i,j}}{\sum_{i=1}^{3N} l^2_{CART \ i,j}}. \tag{2.24}$$

Now we can see that derived expression (2.24) gives previous definition of the reduced masses (2.19) in the partial case of modes normalized in mass-weighted Cartesian coordinates. In such way we have obtained generalized definition of reduced masses, which can be applied to arbitrary mode. The most convenient way to calculate reduced masses of the normal modes is to use the formula (2.22), since elements $l^c_{CARTi,j}$ can be directly taken from the Gaussian output.

As we have already mentioned, we use the amide I normal mode of NMA to construct the local modes of GD (so called projection procedure). The local modes motion, considered in the present work, involves only four atoms of the peptide unit (O-C-N-H). It means that the projection procedure implies fixations of some atoms taking a part into the initial amide I normal mode motion in the NMA molecule. As a result, the mode obtained by the fixation will not possesses some properties inherent in the normal modes. Firstly, they will not be normalized in the mass-weighted coordinates. And secondly, the variation of the local coordinates, defined in such way, will shift center of mass of the system. The normal modes constructed as linear combination of such local modes will also not posses the required properties. Therefore, after the fixation of atoms we need to modify obtained modes to fulfill the two about mentioned requirements. On the first step a constant vector has to be added to each element of the initial local mode to eliminate the shift of the center of mass related with the variation of the local coordinate. And on the second step one needs to multiply each element of the redefined local mode by an appropriate scaling factor to normalize the mode in mass-weighted Cartesian coordinates. These two steps do not commute and should be performed in the above mentioned order. Really, if $\sum y_i^2 = 1$ in general case $\sum (y_i + \Delta)^2 \neq 1$, while from $\sum y_i = 0$ follows that $\sum \alpha y_i = 0$. In other words fixation of center of mass violates normalization while normalization does not effect on the movement of the center of mass.

One can calculate the force and coupling constants using the initial, not normalized, local modes and than correct the obtained values by a corresponding scaling factor. Let us consider this procedure in more details. With l^{*c} we will denote the initial local coordinate which was obtained by fixation of some atoms ($*$ is used to indicate the fixation of atoms) in the amide I normal mode which was initially normalized in the Cartesian coordinates (c index). Next mode is obtained form the previous one by adding a constant vector and subsequent multiplication of a constant. As the result of the above transformation the new mode does not shift the center of mass and is normalized in the Cartesian coordinates. The corresponding vibrational coordinate will be denoted as l^c. The upper index indicates the normalization in the Cartesian coordinates. And finally, the coordinate of the normalized in the mass-weighted coordinates mode will be denoted as l^m. Since the Cartesian atomic shifts in the three considered modes are linear proportional to each other one can find simple linear relation between the corresponding local coordinates (l^{*c}, l^c, and l^m). Doing that we will use the following notation

$$\begin{aligned}\Delta x_i^{*c}\left(l^{*c}\right) &= \sum_{j=1}^{3N} l^{*c}_{CART\ i,j}\ l^{*c}_j, \\ \Delta x_i^{c}\left(l^{c}\right) &= \sum_{j=1}^{3N} l^{c}_{CART\ i,j}\ l^{c}_j, \\ \Delta x_i^{m}\left(l^{m}\right) &= \sum_{j=1}^{3N} l^{m}_{CART\ i,j}\ l^{m}_j.\end{aligned} \tag{2.25}$$

Properties of the local modes, corresponding to the local coordinates l^c and l^m are set by the formula (2.15) and (2.18). The length of the initial local mode we will denote as L, i.e.

$$\sum_{i=1}^{3N} l^{*c\ 2}_{CARTi,j} = L_j^2 \tag{2.26}$$

Let suppose that only amide I local coordinate (indicated by j) is equal to 1 and all other coordinates are equal to zero. In this case the interrelationships among different types of the local coordinates can be found from the requirement of equality of the Cartesian displacements.

$$\begin{aligned}\Delta x_i^{*c}\left(l^{*c}\right) &= \Delta x_i^{c}\left(l^{c}\right) \\ l^{*c}_{CART\ i,j}\ l^{*c}_j &= l^{c}_{CART\ i,j}\ l^{c}_j \\ l_j^{*c\ 2}\sum_i l^{*c\ 2}_{CART\ i,j} &= l_j^{c\ 2}\sum_i l^{c\ 2}_{CART\ i,j} \\ l_j^{*c\ 2} L_j^2 &= l_j^{c\ 2}, \\ l_j^{*c} &= \frac{1}{L_j} l_j^c.\end{aligned} \tag{2.27}$$

The relation between l^c and l^m can be found In the same way

$$\begin{aligned}\Delta x_i^c\left(l^c\right) &= \Delta x_i^m\left(l^m\right) \\ l^c_{CART\ i,j}\ l_j^c &= l^m_{CART\ i,j}\ l_j^m \\ l_j^{c\ 2}\sum_i l^{c\ 2}_{CART\ i,j} &= l_j^{m\ 2}\sum_i l^{m\ 2}_{CART\ i,j} \\ l_j^{c\ 2} &= l_j^{m\ 2}/\mu, \\ l_j^c &= \frac{1}{\sqrt{\mu_j}} l_j^m\end{aligned} \tag{2.28}$$

Using (2.27) and (2.28) we get

$$\frac{\partial^2 E}{\partial l_j^{m\ 2}} = \frac{\partial^2 E}{\partial l_j^2}\frac{1}{\mu_j L^2} \tag{2.29}$$

From the output of the Gaussian we extract the normal modes which are normalized in the Cartesian coordinates. In order to obtain the second order derivatives corresponding to the mass-weighted coordinates we have to devide the derivatives over the Cartesian coordinates by reduced masse of the normal mode μ^n. The same procedure can be performed with the modes where some atoms are fixed. In this case we obtain the following relation between the two type of the derivatives

$$\frac{\partial^2 E}{\partial l_j^{m*\ 2}} = \frac{\partial^2 E}{\partial l_j^{c*\ 2}}\frac{1}{\mu_j^n}, \tag{2.30}$$

where $*$, as before, is used to indicate the fixation of the atoms. However, the above relation does not account for the correction related with the atoms fixation. It means that the relation between the uncorrected and corrected derivatives over the mass-weighted coordinates is the follows

$$\frac{\partial^2 E}{\partial l_j^{m\ 2}} = \frac{\partial^2 E}{\partial l_j^{m*\ 2}}\frac{\mu_j^n}{\mu_j}\frac{1}{L_j^2}. \tag{2.31}$$

With this equation we summarize the question of the local modes normalization.

In the final part of the given chapter we consider the kinetic energy term of the Hamiltonian. In general case, the Hamiltonian of the considered system can be written in terms of the local coordinates as follows

$$\hat{H} = \begin{pmatrix} p1 & p2 \end{pmatrix}\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{pmatrix}\begin{pmatrix} p_1 \\ p_2\end{pmatrix} + \begin{pmatrix} q1 & q2 \end{pmatrix}\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22}\end{pmatrix}\begin{pmatrix} q_1 \\ q_2\end{pmatrix}, \tag{2.32}$$

where q_i, as before, are mass-weighted local coordinates and p_i are conjugated momenta. Let us find values of the coeficients a_{ij} of the above Hamiltonian (all anharmonic terms are niglected). Let us strictly define momenta p_i. We will use standard the definition of momenta

$$p_i = \frac{\partial L}{\partial \dot{q}_j} = \frac{\partial}{\partial \dot{q}_j}(T - V) = \frac{\partial T}{\partial \dot{q}_j}, \tag{2.33}$$

where L is the Lagrangian of the system. We suppose that main contribution to the kinetic energy of the system is conditioned by classical motion of nuclei, *i.e.*

$$T = \sum_{i=1}^{N} \frac{m_i \vec{v}_i^2}{2} \tag{2.34}$$

To express $\vec{v}_i$ through $\dot{q}_i$ let us find a relation between the local coordinates q_i and the Cartesian displacements Δx_i.

$$\Delta x_i = l^m_{CART\ i,1} q_1 + l^m_{CART\ i,2} q_2 \tag{2.35}$$

In the above expression we use two local mode (denoted as q_1 and q_2) which correspond to the consideration of dipeptides. Expresions (2.35) can be straightforvardly used for the calculation of the atomic velocities

$$\Delta v_i = l^m_{CART\ i,1} \dot{q}_1 + l^m_{CART\ i,2} \dot{q}_2. \tag{2.36}$$

Let us substitute the previous formula in the expresions for the kinetic energy (2.34)

$$\begin{gathered} T = \sum_{i=1}^{N} \frac{m_i \vec{v}_i^2}{2} = \sum_{i=1}^{N} \frac{m_i}{2} \left(l^m_{CART\ i,1} \dot{q}_1 + l^m_{CART\ i,2} \dot{q}_2 \right)^2 = \\ \sum_{i=1}^{N} \frac{m_i}{2} l^{m\ 2}_{CART\ i,1} \dot{q}_1 + \sum_{i=1}^{N} \frac{m_i}{2} l^{m\ 2}_{CART\ i,2} \dot{q}_2 + \sum_{i=1}^{N} \frac{m_i}{2} \left(l^m_{CART\ i,1} l^m_{CARTi,2} \right) 2 \dot{q}_1 \dot{q}_2. \end{gathered} \tag{2.37}$$

The considered modes are localized on different peptide units and, as a consequence, there are no atoms wich take a part in both local modes. From this fact follows that for any atoms number i, at least one multiplier in the bracets of the last term is equal to zero. In such way the expresion for the kinetic energy can be rewriten in the form

$$T = \frac{\dot{q}_1^2}{2} \sum_{i=1}^{N} m_i l^{m\ 2}_{CART\ i,1} + \frac{\dot{q}_2^2}{2} \sum_{i=1}^{N} m_i l^{m\ 2}_{CART\ i,2}. \tag{2.38}$$

Using normalization condition (2.15) the above expresion can be transformet to very simple form

$$T = \frac{\dot{q}_1^2}{2} + \frac{\dot{q}_2^2}{2}. \tag{2.39}$$

In such way we have shown that elements of the matrix representing the kinetic energies are given by $a_{ij} = \delta_{ij}$ where δ_{ij} is the Kronecker symbol.

2.2.3 Step for the Numerical Differentiation

In this section we consider the finite difference method used for the calculation of the second derivatives of the potential energy over the local coordinates. In particular, we consider selection of the step for the numerical differentiation. The formulas used for the calculation of the second derivatives are given bellow.

$$\frac{\partial^2 E(0,0)}{\partial q_1^2} \approx \frac{E(\Delta q,0)+E(-\Delta q,0)-2E(0,0)}{\Delta q^2}, \tag{2.40}$$

$$\frac{\partial^2 E(0,0)}{\partial q_2^2} \approx \frac{E(0,\Delta q)+E(0,-\Delta q)-2E(0,0)}{\Delta q^2}, \tag{2.41}$$

$$\frac{\partial^2 E(0,0)}{\partial q_1 \partial q_2} \approx \frac{E(\Delta q_1,\Delta q_2)-E(-\Delta q_1,\Delta q_2)-E(\Delta q_1,-\Delta q_2)+E(-\Delta q_1,-\Delta q_2)}{4\Delta q_1 \Delta q_2}. \tag{2.42}$$

Where E is the energy of the system and Δq_j are the differentiation steps. In our case the two local coordinates are treated in the same way, so we set $\Delta q_1 = \Delta q_2 = \Delta q$.

The approximate values of the second derivatives calculated by formulae (2.40) - (2.42) dependent on Δq. As an example, in the Fig. 2.4 the coupling and force constants are shown as functions of Δq for several conformations. According to the Fig. 2.4, range of the step values Δq can be conditionally divided into two parts where second derivatives shows qualitatively different behavior. In the first range $\Delta q \in [0.001, 0.01]$ the force and coupling constants show unregular fluctuations, whose amplitude decreases with increasing value of Δq. The amplitude of the force constant fluctuations is comparable with conformational dependency, while the conformational dependency of the coupling constant dominates over the dependency on the step Δq. In the second range $\Delta q \in [0.01, 0.1]$ the force constant shows a regular dependency on the Δq, while the coupling constant remains constants.

Since dependency on Δq can exceed conformational dependency, we have performed an additional study to estimate the optimal step Δq. The first origin of the inaccuracy is the systematic dependency of the force constants on the differentiation step Δq. This effect could not be observed if the potential energy is a parabolic quadratic function of the local coordinates. So, we can conclude that anharmonic terms becomes important if Δq is sufficiently large and causes the systematic shift of the force constants. In order to evaluate parabolicity of the energy as a function of the local coordinates we constructed the following functional:

$$F[y(x), x_2, x_1] = \frac{ln[y(x_2)-y(x_{min})]-ln[y(x_1)-y(x_{min})]}{ln(x_2-x_{min})-ln(x_1-x_{min})}, \tag{2.43}$$

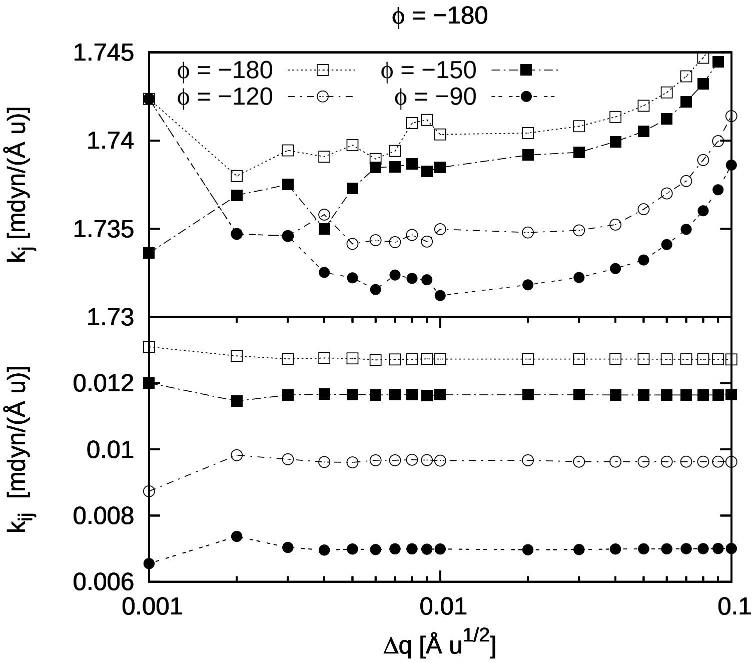

Figure 2.4: The force and coupling constant as function of the step used in the numerical differentiation.

where y is quasi-parabolic function with minimum located at $x = x_{min}$, x_1 and x_2 are two arbitrary points. The usage of (2.43) with respect to the function

$$y\left(x\right) = \alpha \left(x - x_{min}\right)^n \tag{2.44}$$

gives n. So the closer n to the 2, the more precisely point $[x_2, y\left(x_2\right)]$, minimum of the dependency $[x_{min}, y\left(x_{min}\right)]$ and some intermediate point $[x_1, y\left(x_1\right)]$ can be fitted by parabola. The functional (2.43) has been applied to four intersections of the energy E as a function of the local modes $Y_1\left(q\right) = E\left(q, 0\right)$, $Y_2\left(q\right) = E\left(0, q\right)$, $Y_3\left(q\right) = E\left(q, q\right)$, $Y_4\left(q\right) = E\left(q, -q\right)$. For the given functions $Y_j(q)$ functional (2.43) has been considered as a function of the x_2:

$$\xi_i\left(\Delta q_i\right) = \begin{cases} F\left[Y_j\left(q\right), \Delta q_i, \Delta q_{i-1}\right] & if \qquad \Delta qi < Deltaq_{min} \\ F\left[Y_j\left(q\right), \Delta q_i, \Delta q_{i+1}\right] & if \qquad \Delta q_{i-1} > Deltaq_{min} \end{cases} \tag{2.45}$$

In this way we can estimate the accuracy of the harmonic approximation for different ranges of the local coordinates.

We found out that high precision of the minimum finding is important for the calculation of ξ_i. In the Fig. 2.5 we show the above defined function ξ_1 calculated for $\phi = -180$ and $\psi = -180$. Different curves correspond to different ways of the minimum definition.

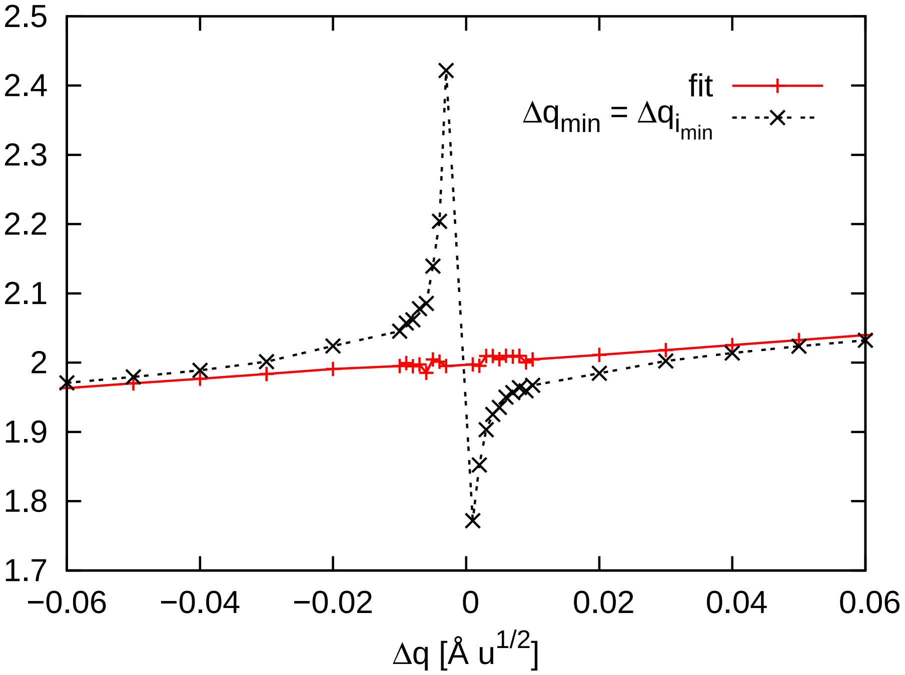

Figure 2.5: Dependence of the functions $\xi_j\left(\Delta q\right)$ on the accuracy of the minimum definition.

In the first one, as a position of minimum Δq_{min} of the function Y_1 we took a value of the step $\Delta q_{i_{min}}$ where a function $Y_1\left(\Delta q\right)$ was minimal in comparison with values corresponding to other steps. In the second case the localization of the minimum was found by minimization of the parabola obtained by fitting of the three nearest to the minimum points $[\Delta q_{i_{min}-1}, Y_j\left(\Delta q_{i_{min}-1}\right)]$, $[\Delta q_{i_{min}}, Y_j\left(\Delta q_{i_{min}}\right)]$, and $[\Delta q_{i_{min}+1}, Y_j\left(\Delta q_{i_{min}+1}\right)]$. In the Fig. 2.5 one can see that relatively small shift of the approximate value of the minimum used in the formula (2.45) from its real position causes an increasing systematic deviation from 2 near zero. Absence of a systematic deviation in the case of the usage of our minimum finding procedure tell us about its sufficient accuracy.

The above introduced functions ξ_j were calculated for different conformations of GD. The representative dependencies are shown in the Fig. 2.6. From the figure one can see that in the range $\Delta q \in [0.01, 0.1]$ ξ_j as a function of Δq shows a systematic dependency on Δq and $\Delta q \in [0.001, 0.01]$ non-regular fluctuation that increase by $\Delta q \rightarrow 0$ in the range. On this basis we can conclude that the optimal step for the calculation of the second derivatives by formulae (2.40)-(2.41) has to be around 0.01, what seams the be the best compromise between the systematic deviation and non-regular fluctuations.

In order to understand the mentioned behavior, the dependency of the second derivative

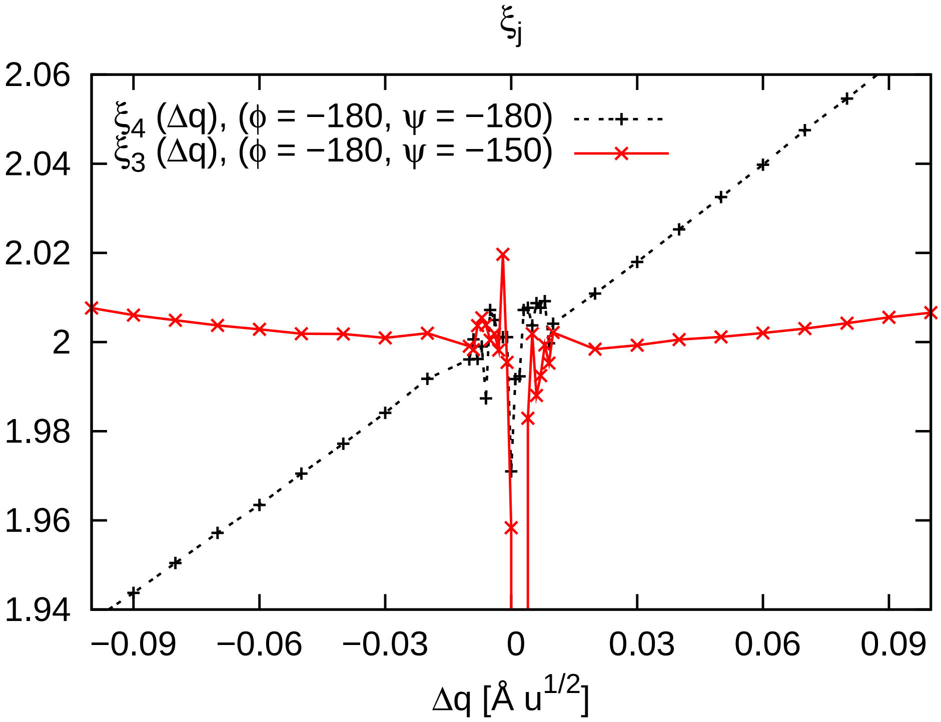

Figure 2.6: Examples of the functions $\xi_j(\Delta q)$.

on Δq have been studied in more details. If we apply the expressions (2.40) to the Taylor's expansion of the energy over the local coordinate Δq_1 and Δq_2 we obtain:

$$\begin{aligned}\frac{\partial^2 E\left(q_{1,min},q_{2,min}\right)}{\partial q_1^2} \approx \frac{\partial^2 E\left(0,0\right)}{\partial q_1^2} - \left[\frac{\partial^3 E\left(0,0\right)}{\partial q_1^3} q_{1,min} + \frac{\partial^3 E\left(0,0\right)}{\partial q_1^2 \partial q_2} q_{2,min}\right] + \\ \left[\frac{1}{2}\frac{\partial^4 E\left(0,0\right)}{\partial q_1^4} q_{1,min}^2 + \frac{1}{2}\frac{\partial^4 E\left(0,0\right)}{\partial q_1^2 \partial q_2^2} q_{2,min}^2 + \frac{\partial^4 E\left(0,0\right)}{\partial q_1^3 \partial q_2} q_{1,min} q_{2,min}\right] + \\ \frac{1}{12}\frac{\partial^4 E\left(0,0\right)}{\partial q_1^4}\Delta q^2 + \cdots\end{aligned} \tag{2.46}$$

For the mixed second derivative (2.42) the same procedure gives the next expressions:

$$\begin{aligned}\frac{\partial^2 E\left(q_{1,min},q_{2,min}\right)}{\partial q_1 \partial q_2} \approx \frac{\partial^2 E\left(0,0\right)}{\partial q_1 \partial q_2} - \left[\frac{\partial^3 E\left(0,0\right)}{\partial q_1^2 \partial q_2} q_{1,min} + \frac{\partial^3 E\left(0,0\right)}{\partial q_1 \partial q_2^2} q_{2,min}\right] + \\ \left[\frac{1}{2}\frac{\partial^4 E\left(0,0\right)}{\partial q_1^2 \partial q_2^2} q_{1,min}^2 + \frac{1}{2}\frac{\partial^4 E\left(0,0\right)}{\partial q_1 \partial q_2^3} q_{2,min}^2 \frac{\partial^4 E\left(0,0\right)}{\partial q_1^2 \partial q_2^2} q_{1,min} q_{2,min}\right] + \\ \frac{1}{6}\frac{\partial^4 E\left(0,0\right)}{\partial q_1^3 \partial q_2}\Delta q^2 + \cdots\end{aligned} \tag{2.47}$$

One can see from formulae (2.46) and (2.47) that there ara two types of inaccuracies involved in the in calculations of the second derivatives. The first one is related with the fact that second derivatives are calculated in the point $q_1 = q_2 = 0$, which is not exact the minimum of the energy. As we can see in the above given formulas (2.46) and (2.47), this type of inaccuracy does not depend on Δq and consequently cannot cause the above found systematic dependencies of the second derivatives on Δq.

The second type of inaccuracy is related with finite value of the step Δq. This contribution is proportional to the square of Δq and the fours order of the second derivatives of the energy. This inaccuracy causes systematic dependency of second derivatives and ξ_j on the Δq in the region $\Delta q \in [0.01, 0.1]$.

The non-regular fluctuations in the region $\Delta q \in [0.001, 0.01]$ can be explained as follows. In the case of a cubic function $y(q) = \alpha q^2 + \beta q^3$ the second derivative calculated by formula (2.40) is equal to $2\alpha + 6\beta q$. In order to explain observed behavior in the vicinity of zero we introduce an additional non-regular function $\Upsilon(q)$ so that $y(q) = \alpha q^2 + \beta q^3 + \Upsilon(q)$. In this case the second unmixed derivative will be equal to

$$\frac{\partial^2 E}{\partial q^2} = 2\alpha + 6\beta q + \frac{\Upsilon(\Delta q) + \Upsilon(-\Delta q) - 2\Upsilon(0)}{\Delta q^2} \tag{2.48}$$

To extract function $\Upsilon(q)$ from energy dependencies we performed its interpolation in the range $\Delta q \in [0.01, 0.1]$ by square function. Difference between the interpolating function and interpolated data is shown in the Fig. 2.7 (upper panel) where one can see systematic deviation of data form interpolating function. Systematic deviations disappears if one performs the same procedure with a cubic function (see Fig. 2.7, lower panel). It means that in the considered range of Δq one can distinguished only quadratic and cubic terms in the dependency of the energy on the local coordinates. In such way we proved that our assumption concerning $\Upsilon(q)$ was correct. The function $\Upsilon(q)$ appears in the expression () and causes above considered non-regular fluctuations. Devision by Δq in (2.48) causes the increase of the fluctuation amplitude if Δq decreases.

In order to estimate an influence of the function $\Upsilon(q)$ on the value of the second derivative calculated by formula (2.40) with the usage of $\Delta q = 0.01$ we calculated second derivative of the interpolating function analytically. Derivatives of the energy and its fitting cubic function are shown in the Fig. 2.8 as functions of configurations. From the fig. one can see that at $\Delta q = 0.01$ influence of the $\Delta(q)$ on the result is neglectable.

2.2.4 Hessian Matrix Reconstruction Method

In the above considered finite difference method, the approximate normal modes are defined as set of orthonormal linear combinations of the local modes, diagonalizing the potential

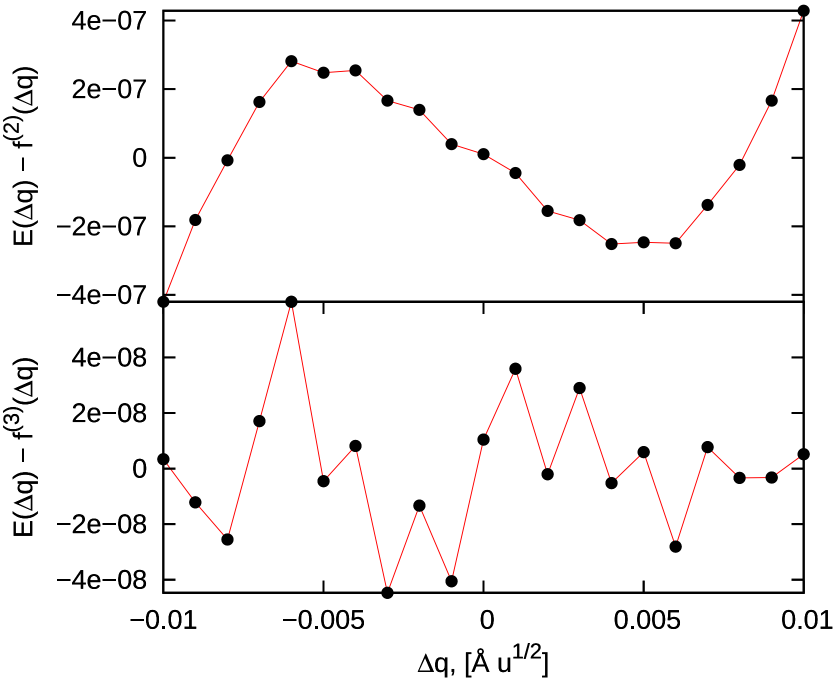

Figure 2.7: Difference between the exact potential energy and the fitting polynomial of different order.

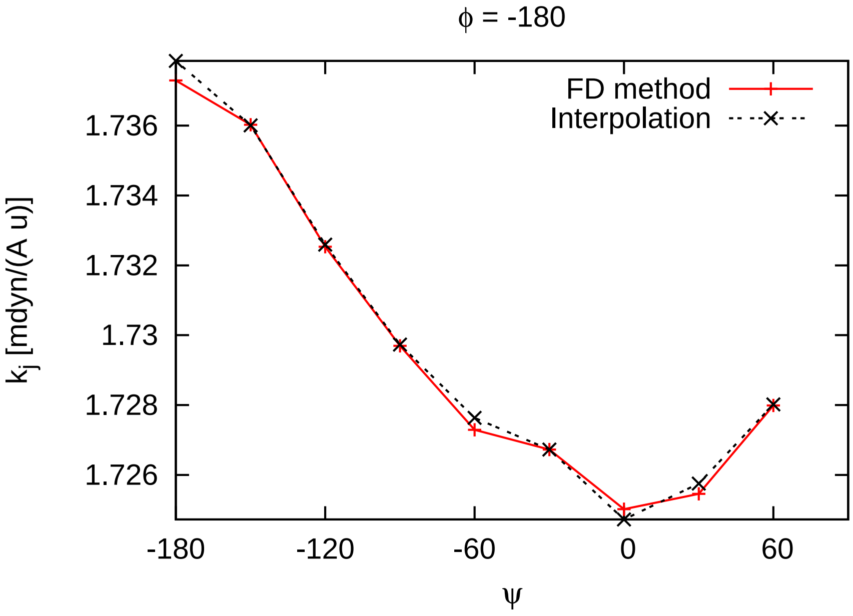

Figure 2.8: Force constant calculated by numerical differentiation of the *ab initio* potential energy and the fitting cubic polynomial.

energy matrix. As another possibility, we can define the approximate normal modes as such linear combinations of the local modes, which minimizes geometrical deviation from the exact normal modes. This definition is used as a basis in the so called Hessian matrix reconstruction (HMR) method [28, 32]. The existence of different definitions of the approximate normal modes is based on the fact that exact normal modes cannot be exactly expressed as linear combinations of the local ones. As a consequence, one can use different criteria to estimate to which extent a given linear combination of the local modes can be considered as an approximation to a normal mode. Different criteria used for the definition of the approximate normal modes require different quantities as the input. In the case of the FD method we have to know the potential energy as a function of the local coordinates. Diagonalization of the corresponding potential energy matrix gives us the localization angle Θ and the normal mode frequencies (see definition (2.3)).

In the HMR method we have to know the normal modes and their frequencies ω_j. The normal modes are used to find Θ which parameterize the unitary transformation relating the local and normal modes. Having Θ and the normal mode frequencies ω_j we can find the site energies ε_j and the coupling β between the local modes. The HMR method will be considered in more details in this sections.

2.2.5 Usage of the NMA-Based Local Modes

For simplicity we will describe the usage of the HMR method for parameterization of the vibrational Hamiltonian of dipeptides (systems which have only two amide I normal modes). All further derivations can be straightforwardly extended on the case of polypeptides. Let $\vec{n}_1$ and $\vec{n}_2$ be the amide I normal modes of dipeptide, which are given as $3N$-components vectors of atomic mass-weighted Cartesian displacements (where N is number of atoms). In this representation the normal modes can be expressed through the local modes $\vec{l}_j$ in the following way

$$\begin{pmatrix} \vec{n}_1 \\ \vec{n}_2 \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} \begin{pmatrix} \vec{l}_1 \\ \vec{l}_2 \end{pmatrix} + \begin{pmatrix} \vec{l}_1^{(t)} \\ \vec{l}_2^{(t)} \end{pmatrix}. \tag{2.49}$$

The first summand in the above expression corresponds to the fact that the normal modes $\vec{n}_j$ can, within some approximation, be presented as linear combinations of the local modes $\vec{l}_j$. The last summand (containing $\vec{l}_j^{(t)}$) counts for the fact that the above representation is not exact (the upper index t indicates that this component of the normal modes is transverse to the plane spanned by the local modes). The smaller $\vec{l}_j^{(t)}$ are, the better is the representation of the normal modes by linear combination of the local modes. If $\vec{l}_j^{(t)}$

are exactly equal to zero, we can find α_{ij} by the following formula

$$\alpha_{ij} = \left(\vec{n}_i, \vec{l}_j\right). \tag{2.50}$$

It should be noticed that the local modes $\vec{l}_j^{\,(t)}$ given in mass-weighted Cartesian coordinates depends on the orientation of the corresponding peptide unit. If vectors $\vec{l}_t^{\,(t)}$ are rather small the relation (2.49) can be written in the following form

$$\begin{pmatrix} \vec{n}_1^a \\ \vec{n}_2^a \end{pmatrix} = \begin{pmatrix} \cos(\theta_{11}) & \sin(\theta_{12}) \\ -\sin(\theta_{21}) & \cos(\theta_{22}) \end{pmatrix} \begin{pmatrix} \vec{l}_1 \\ \vec{l}_2 \end{pmatrix}, \tag{2.51}$$

where θ_{ij} are close to each other and the upper index a indicates that normal modes given by the expression are approximations to the exact modes. The vectors $\vec{n}_j^a$ given by the equation (2.49) are nothing else than projections of the exact normal modes on the plane spanned by the local modes $\vec{l}_j$. The projections are neither perpendicular nor normalized, whereas the local modes $\vec{l}_j$ are normalized by the construction. In order to obtain normalized normal modes we have to set $\theta_{j1} = \theta_{j2} \equiv \theta_j$. Under this condition θ_1 can be calculated as follows

$$\theta_1 \equiv \arctan\frac{\sin(\theta_1)}{\cos(\theta_1)} = \arctan\frac{\sin(\theta_{21})}{\cos(\theta_{11})} = \arctan\frac{\alpha_{12}}{\alpha_{11}} = \arctan\frac{\left(\vec{n}_1, \vec{l}_2\right)}{\left(\vec{n}_1, \vec{l}_1\right)}. \tag{2.52}$$

It should be noticed that normal modes calculated by the equation (2.51)

$$\vec{n}_1^a = \cos(\theta_{11})\,\vec{l}_1 + \sin(\theta_{12})\,\vec{l}_2 \tag{2.53}$$

are not equal to the modes calculated with the usage of θ_1

$$\vec{n}_1^a = \cos(\theta_1)\,\vec{l}_1 + \sin(\theta_1)\,\vec{l}_2. \tag{2.54}$$

Replacement of the equation (2.53) by the (2.54) is nothing but the normalization of the normal mode $\vec{n}_1^a$. The above normalization does not change angles between the considered normal mode and the two local ones. It should be also noticed that the angle θ_1, defined as above, does not require normalization of the initial (exact) normal modes $\vec{n}_1$. In the range $[-180, 180]$, the expression (2.52) gives two values for θ_1. However, only one of them gives correct sign of α_{11} and α_{12} which are equal to $\cos(\theta_1)$ and $\sin(\theta_1)$, respectively. It should be added that we defined the first normal mode as those which has the lower frequency.

The same procedure has been applied for the calculation of the θ_2.

$$\theta_2 = \arctan \frac{\sin(\theta_2)}{\cos(\theta_2)} = \arctan \frac{\sin(\theta_{21})}{\cos(\theta_{22})} = \arctan \left(-\frac{\alpha_{22}}{\alpha_{21}} \right) = \arctan \left[-\frac{\left(\vec{n}_2, \vec{l}_1\right)}{\left(\vec{n}_2, \vec{l}_2\right)} \right] \equiv \arctan \lambda_2. \tag{2.55}$$

As it was in the case of θ_1, the above expression gives two values for θ_2. In the case of θ_2 we select those value which is closer to the θ_1. We can do that since sign of the normal modes has no physical meaning. So we always can define it in the way giving the relation (2.51), where θ_{ij} are close to each other. In other words, the sign of the second normal mode was defined in such way that

$$[\vec{n}_1, \vec{n}_2]_3 > 0, \tag{2.56}$$

where $\vec{n}_j = (\alpha_{j1}, \alpha_{j,2}, 0)$.

Finally, after θ_1 and θ_2 are calculated, we replace them with their average value $\Theta = (\theta_1 + \theta_2)/2$. The latter reduction of the parameters number makes the approximate normal modes orthogonal to each other.

Here we have to mention that the difference between θ_1 and θ_2 was in the range $[-3.5°, 5.9°]$ and its quadratic mean was about $1.7°$. It means that projections of the normal modes on the plane spanned by the local modes can be considered as perpendicular to each other within a good approximation. In its turn it means that planes spanned by the local and normal modes, respectively, are close to each other. This proves that normal modes can be ruther accurately presented by linear combinations of the local modes.

The found angle Θ and the normal mode frequencies are then used to calculate the local mode frequencies and coupling. In order to find the relation between normal mode frequencies as well as the site energies and coupling we will use the following expression for the potential energy

$$V(l_1, l_2) = \begin{pmatrix} l_1 \\ l_2 \end{pmatrix} \begin{pmatrix} \varepsilon_1 & \beta \\ \beta & \varepsilon_2 \end{pmatrix} \begin{pmatrix} l_1 & l_2 \end{pmatrix} = \begin{pmatrix} n_1 \\ n_2 \end{pmatrix} \begin{pmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{pmatrix} \begin{pmatrix} n_1 & n_2 \end{pmatrix} \tag{2.57}$$

Using the relation between the normal and local coordinates

$$\begin{pmatrix} n_1^a \\ n_2^a \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} l_1 \\ l_2 \end{pmatrix}. \tag{2.58}$$

we can find the following relation between the vibrational properties of the normal and local modes

$$\begin{pmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} \varepsilon_1 & \beta \\ \beta & \varepsilon_2 \end{pmatrix} \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \tag{2.59}$$

This gives the following expressions for the normal modes frequencies

$$\begin{aligned}\epsilon_1 &= \omega_1 \cos^2(\theta) + \omega_2 \sin^2(\theta),\\ \epsilon_2 &= \omega_1 \sin^2(\theta) + \omega_2 \cos^2(\theta),\\ \beta &= \frac{1}{2}\left(\omega_1 - \omega_2\right)\sin\left(2\theta\right).\end{aligned} \tag{2.60}$$

2.2.6 Carbonyl Coordinate Displacement Method

In the above description of the HMR method, as an example, we used the local modes constructed on the bases of the normal modes of NMA (hereafter NMA-based local modes). As we have already noticed, the usage of such local modes requires the artificial geometry restrictions which can cause unphysical contribution to the vibrational properties of the peptides. This disadvantage of the FD method appears because of the following dilemma. On the one hand, in the FD method, the local modes have to involve sufficient number of atoms in order to give a reasonable dependency of the potential energy on the local coordinates. On the other hand, the more atoms are involved into the local mode motion, the stronger are the geometrical restrictions. This dilemma can be avoided in the HMR method since it uses the local modes for another purpose. They are used to calculate localization of the normal modes. It means that in the HMR method, the local modes may consist of such combination of internal coordinates which does not reproduce correct dependency of the potential energy on the local coordinates but still gives rather accurate localizations of the normal modes. In this section we will show that CO bonds of peptide units can be used as an accurate measure of the normal modes localizations. The main advantage of the CO bonds is that their usage does not require the above mentioned geometry restrictions. The usage of the CO bonds, which is known as carbonyl coordinate displacement (CCD) method [67], is based on the following calculation of the normal mode localization angles

$$\begin{aligned}\theta_1 &= \arctan\frac{CO_2\left(n_1,0\right) - CO_2\left(0,0\right)}{CO_1\left(n_1,0\right) - CO_1\left(0,0\right)},\\ \theta_2 &= \arctan\frac{CO_2\left(0,0\right) - CO_2\left(0,n_2\right)}{CO_1\left(0,n_2\right) - CO_1\left(0,0\right)}.\end{aligned} \tag{2.61}$$

Let us show that, if the normal modes can be presented as linear combination of the identical local modes, the NMA-based local modes and carbonyl coordinate displacement methods will give the same results under condition that the variation of the normal coordinates is sufficiently small. Under the two above mentioned conditions, the nonzero elements of the normal mode will form the following 3*8-dimensional vector

$$\vec{n}_1 = \left(\ \alpha_1\Delta\vec{r}_H,\ \alpha_1\Delta\vec{r}_N,\ \alpha_1\Delta\vec{r}_C,\ \alpha_1\Delta\vec{r}_O,\ \alpha_2\Delta\vec{r}_H,\ \alpha_2\Delta\vec{r}_N,\ \alpha_2\Delta\vec{r}_C,\ \alpha_2\Delta\vec{r}_O\ \right) \tag{2.62}$$

where α_j is the localization of the normal mode on the j_{th} peptide unit. The local modes are

$$\vec{l}_1 = \left(\ \Delta\vec{r}_H,\ \Delta\vec{r}_N,\ \Delta\vec{r}_C,\ \Delta\vec{r}_O,\ 0,\ 0,\ 0,\ 0\ \right), \tag{2.63}$$

$$\vec{l}_2 = \left(\ 0,\ 0,\ 0,\ 0,\ \Delta\vec{r}_H,\ \Delta\vec{r}_N,\ \Delta\vec{r}_C,\ \Delta\vec{r}_O\ \right). \tag{2.64}$$

In this case the localization angle θ_1 calculated with the usage of the expression (2.52) will be as follow

$$\theta_1 = \arctan\frac{\left(\vec{n}_1,\vec{l}_2\right)}{\left(\vec{n}_1,\vec{l}_1\right)} = \arctan\frac{\alpha_2}{\alpha_1} \tag{2.65}$$

In the carbonyl coordinate displacement method, the localization angle should be calculated by the formula (2.61) which, under the considered conditions, will give the following result.

$$\theta_1 = \arctan\frac{|\,(\vec{r}_C+\alpha_2\Delta\vec{r}_C)-(\vec{r}_0+\alpha_2\Delta\vec{r}_O)\,|-|\vec{r}_C-\vec{r}_0|}{|\,(\vec{r}_C+\alpha_1\Delta\vec{r}_C)-(\vec{r}_0+\alpha_1\Delta\vec{r}_O)\,|-|\vec{r}_C-\vec{r}_0|} \tag{2.66}$$

$$= \arctan\frac{\sum_{n=1}^{\infty}\Phi_n\alpha_2^n}{\sum_{n=1}^{\infty}\Phi_n\alpha_1^n} \approx \arctan\frac{\alpha_2}{\alpha_1}.$$

The results of the two last formulas are identical. In such way we have pointed the sufficient conditions of the agreement between the results obtained with the usage of the NMA-based local modes and the CCD method. We have to notice that the above comparison of the local modes can be applied only to the HMR method (and not to the FD one).

2.2.7 Comparison of the Parameterization Schemes

In this section we compare results obtained with the HMR and FD methods. The two methods are used in combination with the NMA-based local modes and, as a consequence, the restricted geometry. The usage of the same local modes and geometries allows us to concentrate on the difference related with the usage of the different criteria for the definition of the approximate normal modes. In the Fig. 2.9 we show the

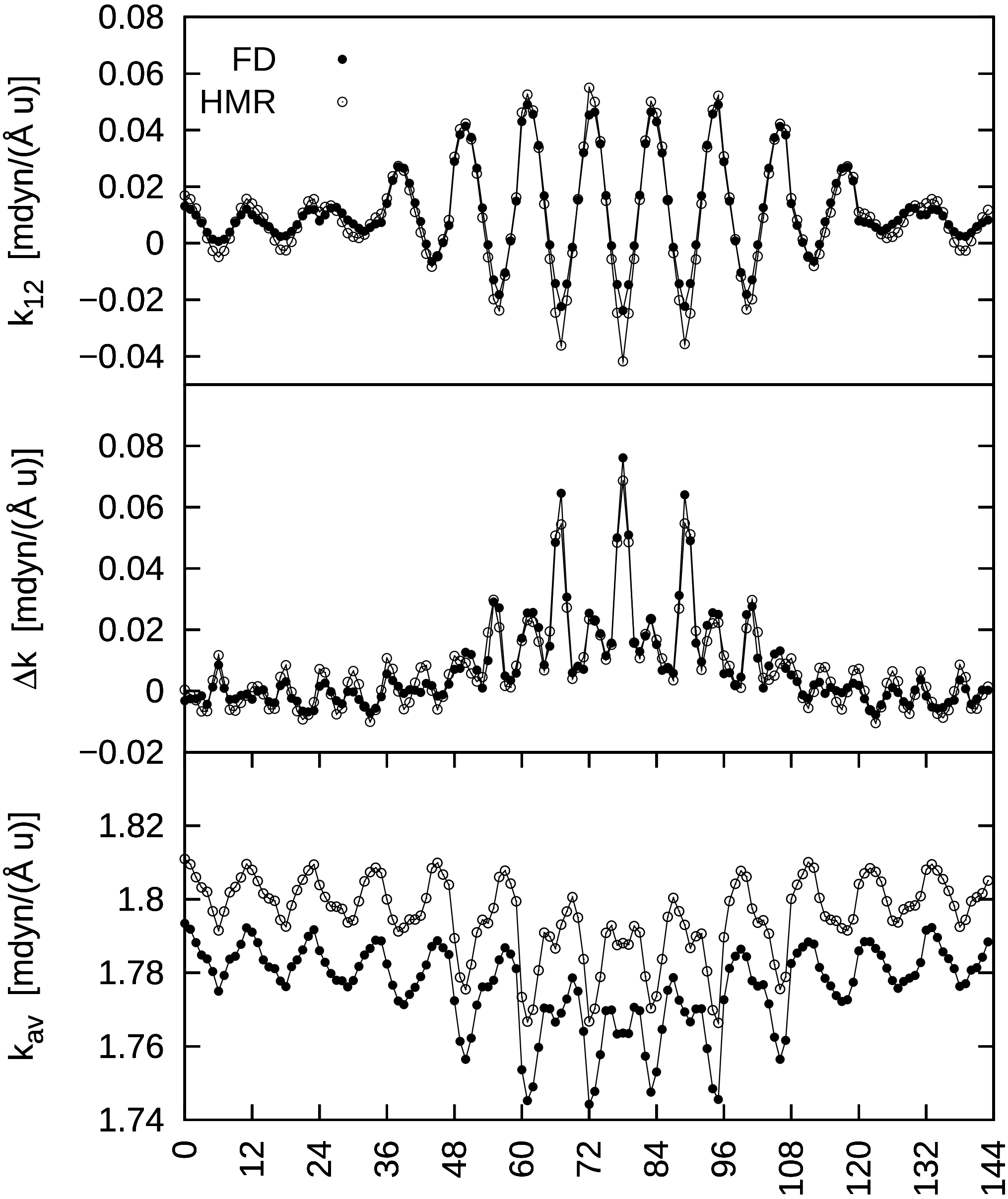

Figure 2.9: Coupling constants, difference between the force constants, and average force constant calculated by FD and HMR methods.

coupling, differnece between the force constants and the avergae force constant as functions of conformation. The shown data have been obtained with the usage of the two-dimensional grid of the dihedral angles ϕ and ψ. Each dihedral angles was variate in the range from -180° to 150° with the step 30° (12 values). All possible combinations of the ϕ and ψ angles were considered (144 points). For the simplicity of the comparison we numerated the $2D$ data by one integer parameter n in the following way. The data number n is related to the ϕ and ψ angles as $\phi = -180^\circ + 30^\circ \times int\left[(n-1)/12\right]$ and $\psi = -180^\circ + 30^\circ \times mod\left[(n-1), 12\right]$, i.e. conformations are arranged in the order $(\phi, \psi) = (-180^\circ, -180^\circ), (-180^\circ, -150^\circ), \ldots, (-180^\circ, 150^\circ), (-150^\circ, -180^\circ)$. Here, $int\,(x)$ stands for the largest integer which is $\leq x$ and $mod(x, y)$ is calculated as $x - int\,(x/y) \times y$.

As we can see in the Fig. 2.9 the two different parameterization schemes give very close conformational dependencies of the considered vibrational properties. The only difference is the average shift of the force constant, which is probably related with the cubic terms of the potential energy and the fact in that FD method, the second derivative are calculated not in the minimum of the potential energy over the local coordinates.

In this section we compare results obtained with the usage of the NMA-based local modes and CCD method. In both cases the HMR method was used. The first reason to usa this method is that it is less sensitive to the size of the local modes. As we have shown before, even CO bonds, used as measures of the normal modes localizations, can give accurate result if the HMR method is used. The latter is not true for the FD method. The CO bonds, used as local modes in the FD methods, will give rather inaccurate results. The second reason to use the HMR method is that it is simpler from the technical point of view. Id does not require the projection procedure in which we have to constrain the geometry depending on the values of the local coordinates. Instead of that, in the HMR method, only the full geometry optimization with the subsequent normal mode analysis are required.

In the Fig. 2.10 we can see that coupling and difference between the force constants obtained with the NMA-based local modes are very close to those obtained with the usage of the CCD method. The sum of the force constant in the two considered cases should be identical since the trace of the matrix is invariant with respect to the arbitrary unitary transformation. Since the both modes use the same initial diagonal matrix, constructed of the force constants of the normal modes, after the unitary transformation the sum of the diagonal elements has also to be identical.

In the previous two sections we have shown that FD and HMR methods give very close

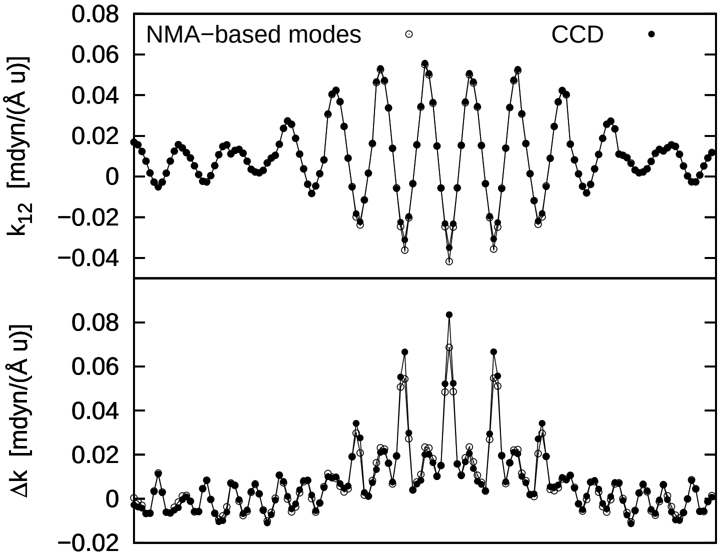

Figure 2.10: The coupling and difference between the force constants calculated by the HMR method in combinations with the NMA-based local modes and CCD method.

results under conditions that they use the same local modes and are applied to the same geometry. We have also shown that replacement of the NMA-based local modes by the CO bonds causes very small changes of the vibrational properties. In such way we have found a parameterization scheme (HMR + CCD) which is quite simple from technical point of view, does not require any geometrical restriction and still is rather accurate.

In this section we apply the HMR method, combined with the CCD ones, to the both the fully optimized geometry (only ϕ and ψ angles were fixed) and the restricted one (which is used if the NMA-based modes are used). The comparison has to show how importante is the influence of the geometrical restrictions, used in the FD method, on the vibrational properties. The results of the comparison are shown in the Fig. 2.11. As we can see in the figure, the force constants are very sensitive to the geometry restrictions while the coupling is rather insensitive to the change of the geometry. These results indicate that the NMA-based local modes, and as a consequence the FD method, are not suitable for the calculation of the force constants. The influence of the considered geometry restriction will be considered in more details in chapter where the amide I anharmonicities are considered.

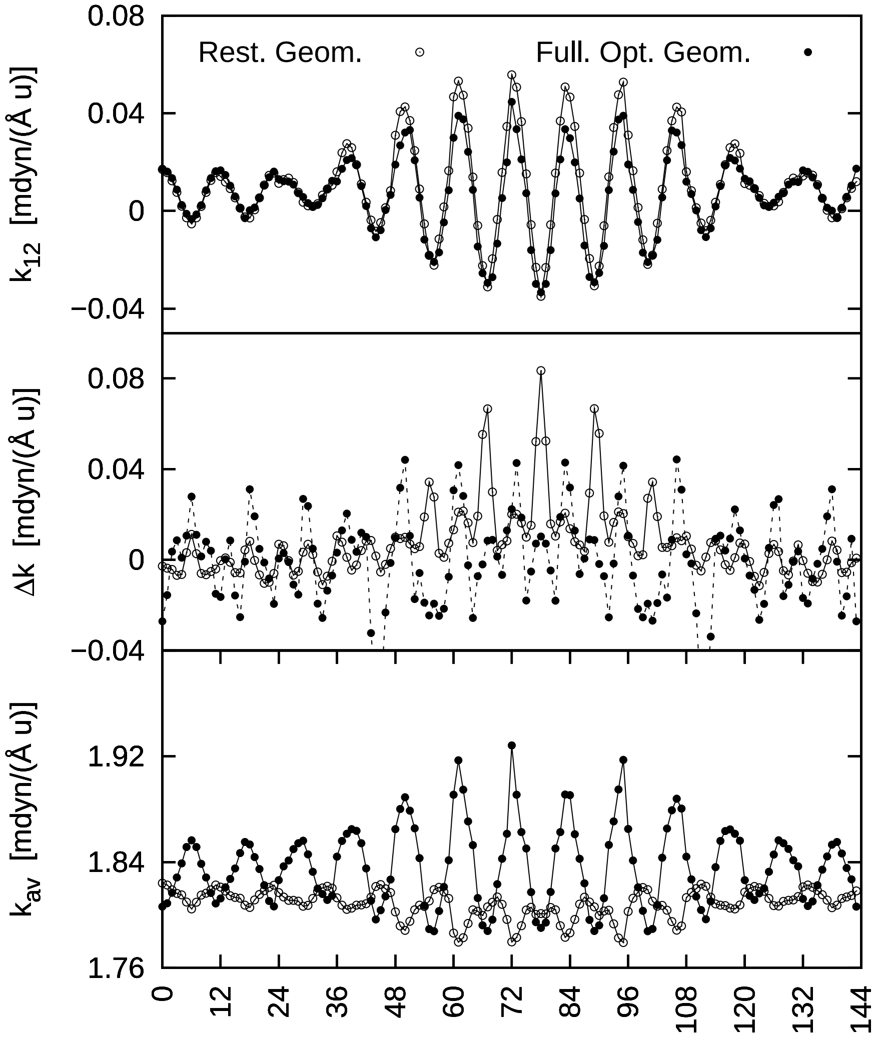

Figure 2.11: The coupling constant, difference between the force constant, and the average force constant calculated for the fully optimized and restricted geometries.

2.3 Amide I Anharmonicities

2.3.1 Anharmonicities in NMA

In the present paper we have calculated the amide I anharmonicities up to the 7_{th} order. For these purposes we have calculated potential energy of NMA as a function of the amide I normal coordinate which has been variated in the range $[-1.4\ u^{1/2}\mathring{A}, 1.4\ u^{1/2}\mathring{A}]$ with the step of 0.01 $u^{1/2}\mathring{A}$. Energies obtained on the described grid have been fitted by the polynomial functions of different order, which afterwards have been used in the calculation of energies of the ground and the first two excited states.

In more details, Hamiltonian of the system was written in the following form

$$\hat{\mathbf{H}} = \frac{1}{2}\hat{\mathbf{p}}^2 + \sum_{n=0}^{N} \frac{C_n}{n!}\hat{\mathbf{q}}^n, \tag{2.67}$$

where coefficients in the potential energy term (C_n) are fitting parameters and $\hat{\mathbf{q}}$ is the operator of the mass-weighted amide I normal mode coordinate (with the corresponding momenta operator $\hat{\mathbf{p}}$). N is the order of the polynomial which has been used for the fitting of the potential energy curve. Then the considered Hamiltonian has been rewritten in terms of the creation and annihilation operators with the usage of the following relations:

$$\begin{aligned} \hat{\mathbf{q}}_\mathbf{i} =& \frac{1}{C_2^{\frac{1}{4}}}\sqrt{\frac{\hbar}{2}}\left(\hat{\mathbf{a}} + \hat{\mathbf{a}}^+\right), \\ \hat{\mathbf{p}}_\mathbf{i} =& -iC_2^{\frac{1}{4}}\sqrt{\frac{\hbar}{2}}\left(\hat{\mathbf{a}} - \hat{\mathbf{a}}^+\right), \end{aligned} \tag{2.68}$$

where C_2 is the coefficient before the second order term. Hereafter it is called the force constant. Afterwards Hamiltonian was expressed through eigenvectors of harmonic oscillator by multiplication from the left and right sides on the unitary operator:

$$\hat{\mathbf{I}} = \sum_{n=0}^{N} |n><n|, \tag{2.69}$$

where $|n>$ are the eigenfunctions of the harmonic oscillator with the force constant C_2. And finally the eigen values of the Hamiltonian have been obtained by its diagonalized. Results obtained within the mentioned procedure can significantly depend on three parameters: range of the normal coordinate used in the fitting of the *ab initio* values of the potential energy, order of polynomial used for the fitting of the given *ab initio* values, and size of the basis set.

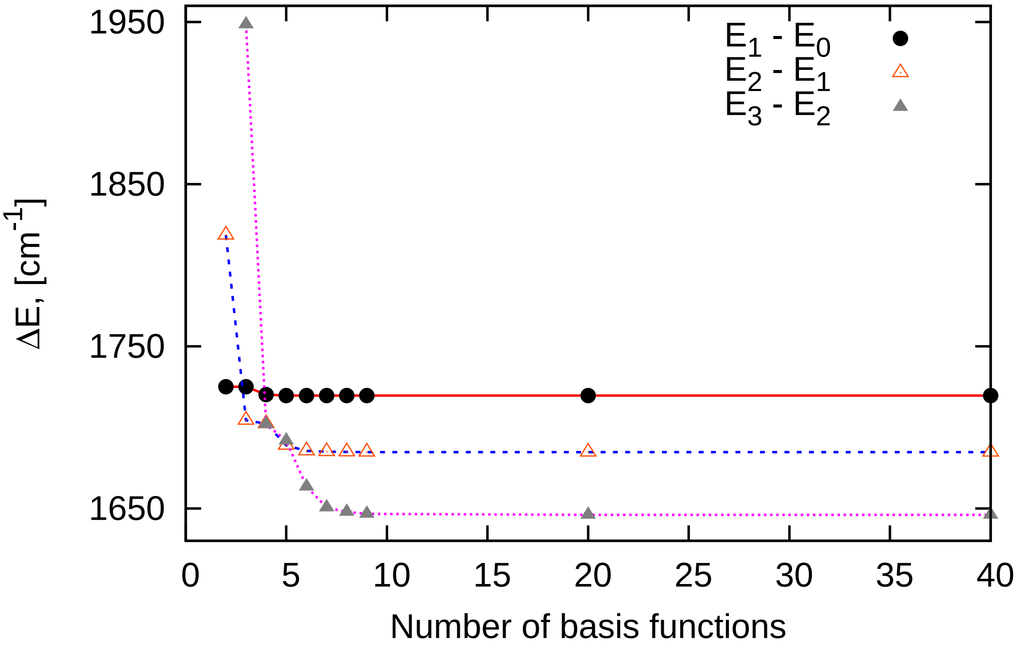

Figure 2.12: Transition frequencies of the first two excited states as well ground state energy as functions of size of the basis set.

Let us first consider convergence of the amide I excitation energies with respect to the number of the basis set functions. For this purposes we have considered cubic potential energy:

$$V(q) = \frac{1}{2}\left(k + \frac{a}{3}q\right)q^2, \tag{2.70}$$

for which energy of the ground and the first three excited states have been calculated with different number of basis set functions. For the mechanical anharmonicity a we took value $a = 3.8\ mdyn/(\mathring{A}^2\ u^{\frac{3}{2}})$ which has been obtained by fitting of the potential energy in the vicinity to zero. The energy differences between neighboring states $\Delta E_j = E_{j+1} - E_j$ $(j = \overline{0,2})$are shown on the the Fig. 2.12 as the functions of the basis set size. On the the Fig. 2.13 we have shown difference between first two transition frequencies. This quantity is zero if the considered anharmonicity a is zero. As one can see the needed number of the basis set function is the larger the higher excitation is considered. For all considered excitation 10 basis functions seams to be sufficient for a reliable energy calculations.

Let us consider other two parameters which can significantly influence on the energy levels. In order to get reasonable values of the energy levels the precise potential energy should be known in the range where wave functions of the considered states are localized. Moreover, order of the interpolating polynomial has to be sufficiently high to accurately

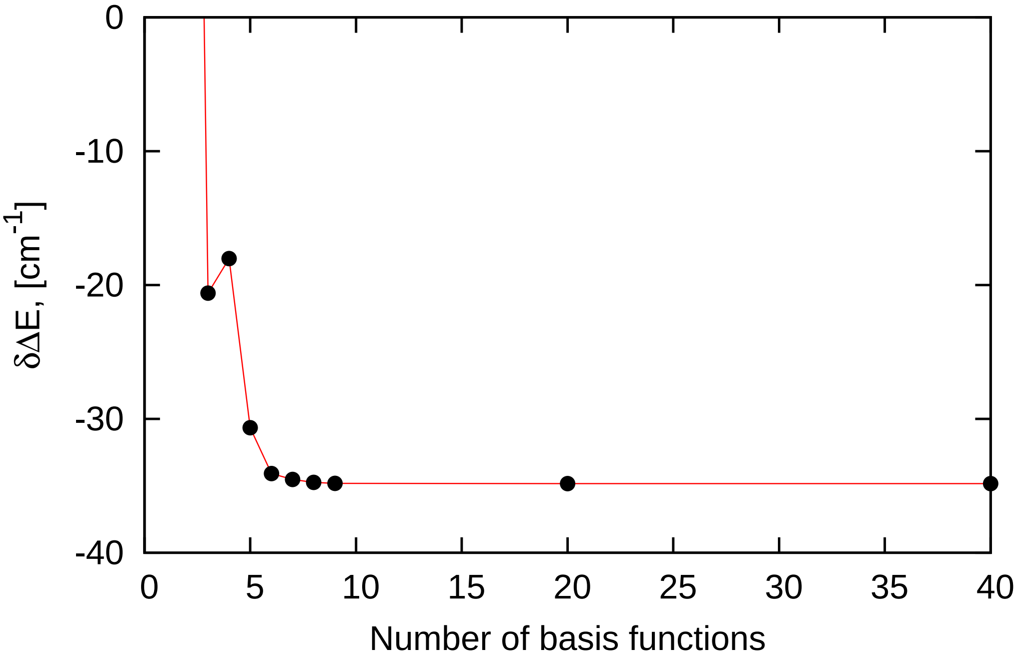

Figure 2.13: Difference between the first two transition frequencies $((E_2 - E_1) - (E_1 - E_0))$

reproduce anharmonicities presenting in the *ab initio* potential energy in the given range. The larger range of the normal coordinate is considered the higher order anharmonicity of the potential energy are visible and, as a consequence, the higher oder polynomial is needed. In another words results are the more accurate the larger range of the potential energy is considered and the higher oder polynomial is used. However, one needs to mention that under fixed order of the fitting polynomial results can become worse if the range of the normal mode coordinate is increased. Really, if one increases range of the normal coordinate the polynomial of a fixed order tries to fit potential energy in the hole range and particularly on the ends of the range (where probability distribution of the considered wave function can be small) and as a consequence in the vicinity to zero (where probability distribution is large) agreement between fitted data and fitting function becomes worse. It means that at fixed ranger results are the better the higher order polynomial is used and under fixed polynomial order there is an optimal range which will give the most accurate results. Having this in mind, the convergence of the eigenvalues on the range of the normal coordinate and order of fitting the polynomial should be carefully studied. In the Fig. 2.14 "frequency anharmonicity" $\delta\Delta E = \Delta E_{21} - \Delta E_{10} = (E_2 - E_1) - (E_1 - E_0)$ is plotted as a function of the range used in the fitting procedure. Different curves correspond to different order of the used polynomials. For each range of the normal coordinate (x-axis) where the

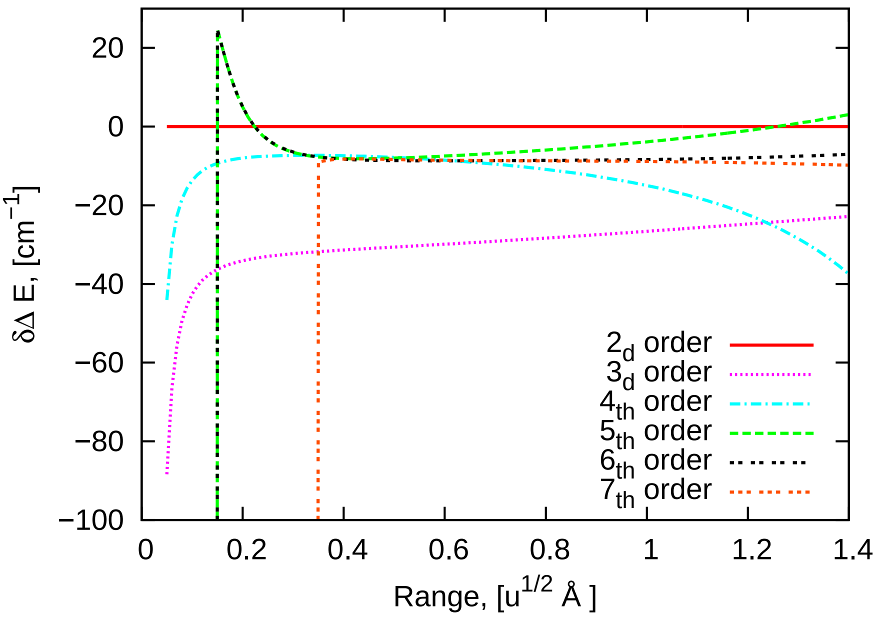

Figure 2.14: $(E_2 - E_1) - (E_1 - E_0)$ as a function of the normal mode coordinate range used during the fitting of the amide I potential energy. Different curves correspond to different order of polynomial used during fitting.

potential energy has been fitted one needs to reach convergence over polynomial order. In general, the larger range of the normal coordinate is used, the higher oder polynomial is needed. However, the oder of the polynomial has to be reasonably high. Superfluous high order terms of the fitting polynomial can slightly improve agreement with the fitted data in the range of the interpolation. However, if the range of the interpolation is not sufficiently large, the coefficients before the high order terms can be obtained with a bad accuracy. This can cause a strong disagreement between the fitted data and fitting polynomial beyond the range of the interpolation. In its turn it can significantly change wavefunctions whose range of the localization is comparable with the range of fitting. The above described effect can be observed in the figure. For the range of about 0.3 $u^{1/2}\AA$ we reached convergence at 5_{th}, 6_{th} order polynomial, however results obtained with the 7_{th} order polynomial show completely different behavior. At about 0.4 $u^{1/2}\AA$ we observe convergence over polynomial at about 4_{th} order (since replacement of 4_{th} oder polynomial by 5_{th} (as well as 6_{th} and 7_{th}) order one does not change significantly results). Moreover, we can say that at this point we have also reached convergence over range. Really, one can say that results have converged at given range of the normal coordinate only if upon further increase of the range results converged over polynomial order do not change significantly. In another words, at each range one needs to reach convergence over polynomial order and than converged in this sense results should be converged over increase of range of the potential energy interpolation. From the physical point of view the range of the wavefunction localization define range of convergence. The last one, together with the form of the potential energy, defines required polynomial order. In the considered range of the normal coordinate (up to the 1.4 $u^{1/2}\mathring{A}$) the 7th order polynomial can be considered as results converged over polynomial order (since in the considered range this results are close to results obtained with 6_{th} order polynomial). Therefore, range of about 0.4 $u^{1/2}\mathring{A}$, and 4_{th} order polynomial can be considered as "minimal" parameters which provide convergence of the first three energy levels. It should be emphasizes that the third order polynomial obtained by the fitting of the *ab initio* potential energy in the range where wave functions are localized cannot give enough accurate result since cubic potential cannot accurately represent *ab initio* energies in the hole range of the wave function localization. Therefore, $\delta\Delta E$ obtained within the cubic potential is about two times larger than corresponding values obtained with higher order polynomials. It means, that for an accurate description of the first two excited states corresponding to amide I vibrations in NMA one needs to take into account at least forth oder anharmonicities. However, in this case, for the consistency one also needs to take mixed forth order derivatives into account, which are not proved to be enough small with respect to the unmixed ones. In other words, in order to get reasonable *ab initio* value of frequencies splitting one needs to take into account forth oder interaction of amide I mode with other modes and in in such way to refuse the approximation of amide I subspace. In the present work we will confine ourself by consideration of the cubic anharmonicities and work within approximation of the amide I subspace.

2.3.2 Anharmonicities in the GD: Conformational Dependency

The building block model in combination with the harmonic approximation is a common method for the description of the amide I vibrations in polypeptides. In this model the local amide I vibrations are considered as harmonic oscillators which are coupled with each other through the second order interaction. It other words, in the expansion of the potential energy over the local coordinates only the second order terms remain (which depend on the peptide secondary structure). In the present work we made one step to the extension of the Hamiltonian by the inclusion of the mechanical anharmonicities to the potential energy. Namely, we have calculated all types of cubic anharmonicities and studied their conformational dependency and influence on the energy levels. Moreover, the introduced

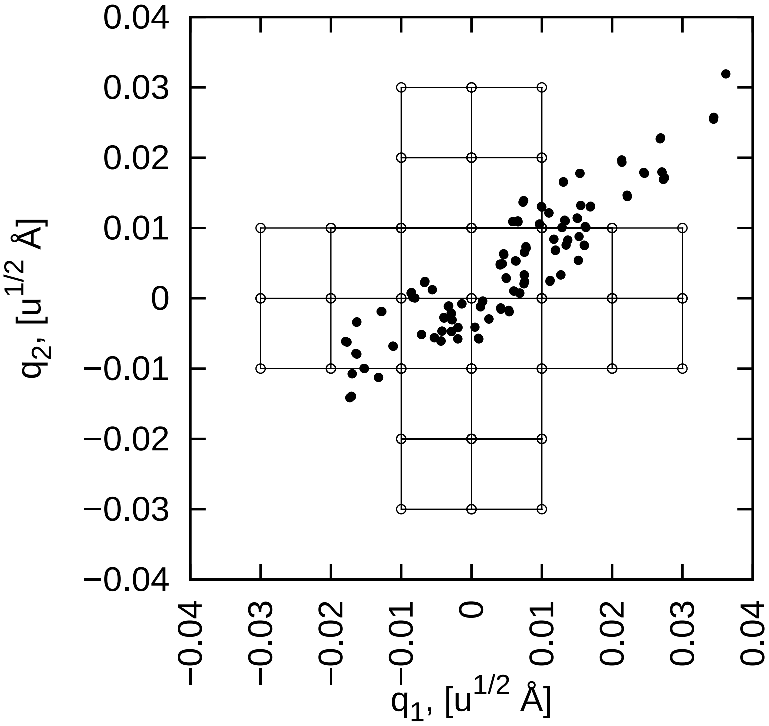

Figure 2.15: Values of the local coordinate which correspond to fully optimized geometry of GD as well as mesh used for calculation of cubic anharmonicities in GD.

cubic terms were used to explain earlier found quantitative disagreement between the local force constant calculated with the finite difference and the Hessian matrix reconstruction methods.

For each pare of the dihedral angles ϕ and ψ the third oder anharmonicities have been calculated by linear fitting of the force constants as functions of the local coordinates. In more details it has been done in the following way. For calculation of the force and coupling constants we have used the finite difference formulas given in our previous paper [53]. It means that in order to calculate all second oder derivatives in point (q_1, q_2) we need to know energy in 9 points $(q_1 + n\Delta q, q_2 + m\Delta q)$, where n and m can take on values $(-1, 0, 1)$ and Δq is the step used for the numerical differentiation. In the Fig. 2.15 the grid of the local coordinate where the potential energy has been calculated is shown.

The second derivative over the first local coordinate has been calculated for $q_1 = 0$ and $q_2 = \Delta q * n$ where $n = \overline{-3, 3}$, as well as for $q_2 = 0$ and $q_1 = \Delta q * n$, where $n = \overline{-2, 2}$. In this way we have obtained $k_1 = \partial^2 E/\delta q_1^2$ as function of q_1 and q_2. The same strategy has been applied to obtain the second order derivatives as functions of the local coordinates. Obtained dependencies of the force constants on local coordinates have been interpolated by the linear functions $k_j\left(q_1, q_2\right) = k_{0j} + \partial^2 E/\delta q_j^2 = a_{j1}q_1 + a_{j2}q_2$. Coefficients of proportionality a_{jm} between the local coordinates and force constants are nothing else as approximate values for the third oder derivatives.

$$a_{jm} = \frac{\partial^3 E}{\delta q_j^2 \delta q_m} \tag{2.71}$$

We have found that in the considered range of the local coordinates the force and coupling constants can within a good approximation be considered as linear functions of the local coordinates. Mean square deviations of the exact second order derivatives from the corresponding fitting values are only about 0.005 $mdyn/(\mathring{A}u)$. The unmixed third order derivatives calculated for different ϕ and ψ angles were found to slightly differ from each other. Moreover, conformational dependency has significantly random character. Taking this into account we have replace *ab initio* (ϕ, ψ)-maps of the cubic unmixed anharmonicities by constant numbers. Moreover, the average values of the two types of the unmixed derivatives were found to be very close to each other (3.8223 $mdyn/(\mathring{A}^2\ u^{\frac{3}{2}})$ and 3.8347 $mdyn/(\mathring{A}^2\ u^{\frac{3}{2}})$ for a_{11} and a_{22}, respectively) and, as a n approximation, we used one averaged value (3.8285 $mdyn/(\mathring{A}^2\ u^{\frac{3}{2}})$) for the two cases. It should be mentioned that this number is close to the corresponding anharmonicity found for the NMA molecules. The mean square deviation of the calculated anharmonicities from the above introduced average value is only about 0.13 $mdyn/(\mathring{A}^2\ u^{\frac{3}{2}})$ and 0.17 $mdyn/(\mathring{A}^2\ u^{\frac{3}{2}})$ for the a_{11} and a_{22}, respectively. The average values of the mixed third order derivatives (a_{12} and a_{21}) were found to be very close to zero (0.0126 $mdyn/(\mathring{A}^2\ u^{\frac{3}{2}})$ and 0.0139 $mdyn/(\mathring{A}^2\ u^{\frac{3}{2}})$, respectively). Moreover, as in the case of the unmixed derivatives, the conformational dependency was found to be small and rather random. To simplify the model we considered the unmixed derivatives to be equal to zero. The mean square deviation of the exact number from zero is only 0.083 $mdyn/(\mathring{A}^2\ u^{\frac{3}{2}})$ and 0.098 $mdyn/(\mathring{A}^2\ u^{\frac{3}{2}})$ for a_{12} and a_{21}, respectively.

Summarizing the described properties of the cubic anharmonicities we can rewrite generalized Hamiltonian of the dipeptide

$$V\left(l_1, l_2\right) = \frac{1}{2}\left[k_1^f\left(\phi, \psi\right) + \frac{a}{3}l_1\right]l_1^2 + \frac{1}{2}\left[k_2^f\left(\phi, \psi\right) + \frac{a}{3}l_2\right]l_2^2 + k_{12}\left(\phi, \psi\right)l_1 l_2, \tag{2.72}$$

where k_j^f are the force constants calculated in the equilibrium. In the above equation we have redefined the local coordinates. They have been shifted (with respect to the original coordinates) such that zero values correspond to the minimum of the potential energy. In this way we can simplify Hamiltonian by the removal of the linear terms. If one wants to use the force constants of the restricted geometry one needs to add a linear term to the Hamiltonian, replace force constant, and leaves the second oder coupling and local cubic anharmonicities. It should be noticed that the above introduced generalization of the dipeptide Hamiltonian can be easily applied to the polypeptides Hamiltonian.

2.3.3 Explanation of the Disagreement Between FD and HMR Methods

In our previous paper [53] we have reported about qualitatively different conformational dependency of the normal mode frequencies calculated with the two different *ab initio* methods (Hessian matrix reconstruction and finite difference method) (see Fig. 2.16). Mentioned difference has been explained by the artificial geometry restrictions needed for the definition of the local modes used in the finite difference method. In other words the observed difference indicates that the force constants significantly depend not only on the dihedral angles ϕ and ψ but also on some other degrees of freedom which are involved in the geometry restriction. The more detailed analysis described in this paper has shown that to a large extent these degree of freedom are the local coordinates themself. For the sake of simplicity one can imagine the local modes as CO stretching. Geometry restrictions implies that for all configurations CO bond lengths will be the same. In the FD method the second derivatives are calculated for such restricted geometry for mesh of ϕ and ψ angles. However, in our present consideration before the calculation of the second derivatives we have variated local coordinate in such a way to make CO bond length equal to the CO bond lengths in the corresponding (the same ϕ and ψ) full optimized geometry. With such modified finite difference method we were able to rather accurately reproduce the conformational dependency of the normal mode frequencies of the full optimized geometry with the usage of the restricted (but "tuned") geometry. In such way we have shown that the disagreement between the two parameterization schemes can to a large extent be explained by the dependency of the normal mode frequencies on the local coordinates.

Withing the above considered cubic approximation the force and coupling constant are liner functions of the local coordinates. Let us find out whether the cubic approximation is sufficient to treat the difference between the normal mode frequencies of the two geometries. To do that we have used CO bond length as a measure of the local coordinate. From all available full optimized geometries we have extracted CO bond length and calculated values of the local coordinates corresponding to the extracted values. The obtained in such way set of the local coordinates values are shown in the Fig. 2.15. As we can see, almost all points are covered by the mesh with which the force and coupling constant has been calculated

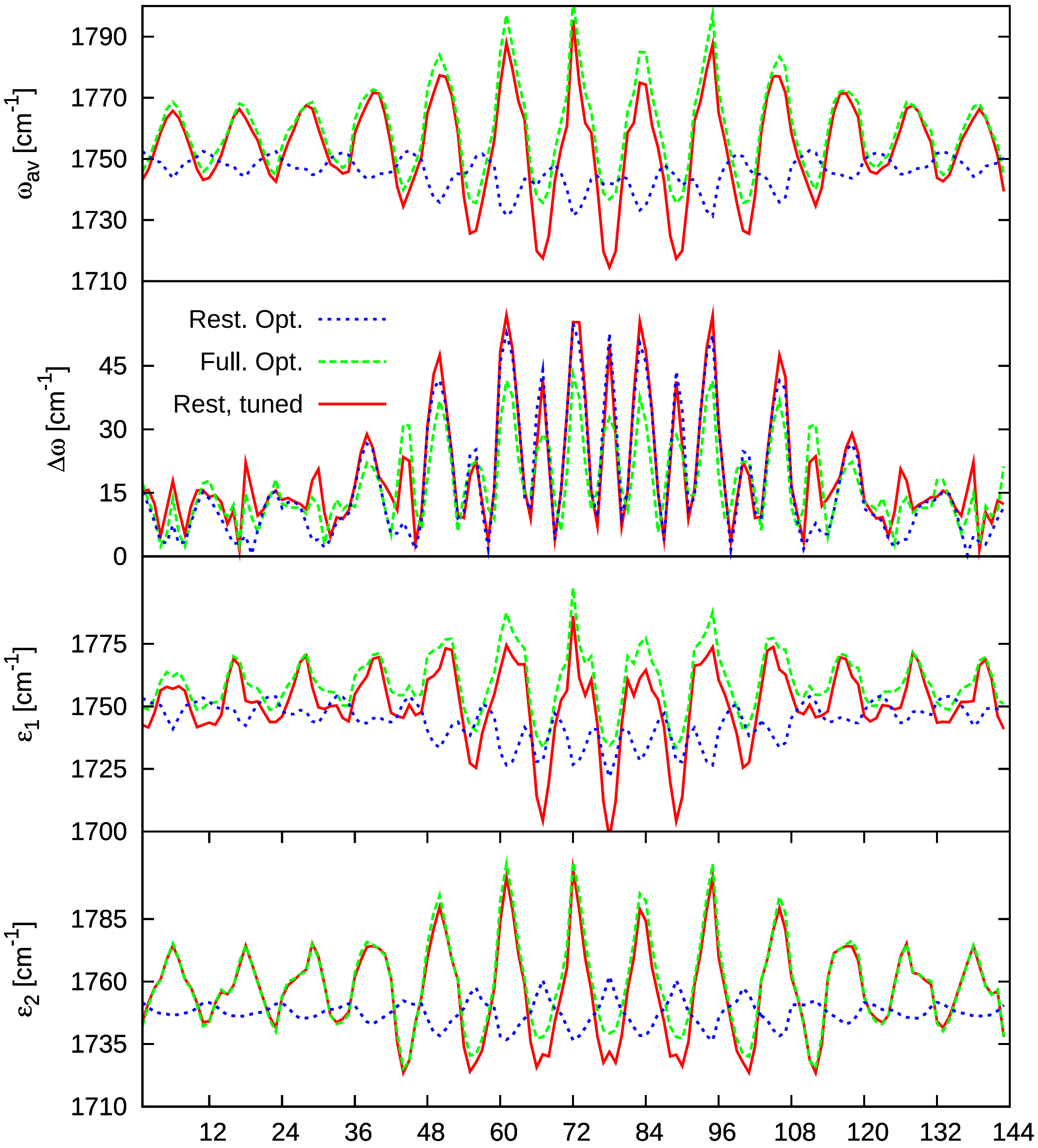

Figure 2.16: Average frequencies, frequency splitting and site energies obtained with the usage of full optimized, restricted and restricted tuned geometries.

and proven to be linear functions of the local coordinates. This fact indicates that it should be possible to calculate the force constants of the full optimized geometry with the usage of the force constants of the full optimized geometry (in the same conformation) and value of the local coordinate corresponding to the given full optimized geometry. This means that in oder to relate results of the two methods it is sufficient to consider potential energy as a cubic function of the local coordinates. Moreover, with above calculated anharmonicities we can explain why coupling being calculated with HMR and FD is almost the same. The reason is the smallness of the third mixed derivatives. I. e. the first derivative of the coupling over one of two local coordinates in nothing else as mixed third derivative which is small and, as a consequence, the coupling itself slightly depends on the corresponding local coordinate.

2.3.4 Effect of the Anharmonicities on the Transition Frequencies

Finally, we have studies effect of the anharmonicities on the transition frequencies. For different force constants, we have calculated the first three transition frequencies with and without the usage of the above described cubic anharmonicity. During calculation we have used basis set consisting of 20 functions. The transition frequencies obtained with the usage of the cubic term were plotted as functions of those calculated without anharmonicity (Fig. 2.17). As one can see in the figure, the cubic anharmonicity causes overall shift down as well as increase range of the variation of the transition frequencies. The latter is observed as increase of the incidence of the obtained dependencies with respect to the bisector. The both effect (shift and change of incidence) are the stronger the higher excitations are considered.

2.4 Building Block Model

In this chapter we are concerned with the *ab initio* modeling of the conformational dependence of ε_n and β_{nm} for the isolated molecule. Because the accurate calculation of vibrational properties in general requires a high theoretical level and large basis sets, direct *ab initio* calculations of vibrational frequencies for the entire molecule are restricted to small systems and/or only few conformations. As a remedy, various fragmentation schemes have been suggested [24, 68–72], whose idea is to perform an *ab initio* description of small peptide ("building block"), which subsequently can be put together in order to describe

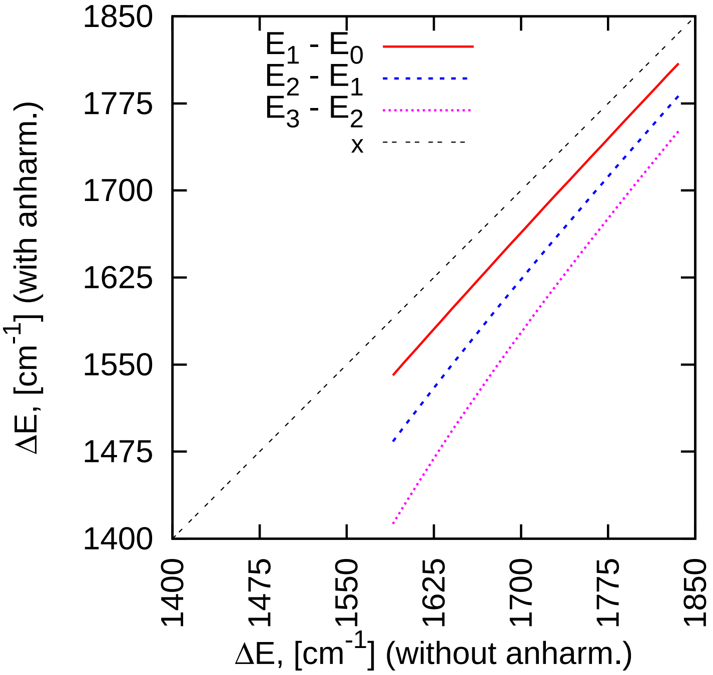

Figure 2.17: Transition frequencies shifted by the anharmonicities are shown as functions of the not shifted values.

larger systems. For example, Bour and Keiderling and their coworkers have developed a property tensor transfer method [24, 70], in which *ab initio*-based atomic force fields and polar and axial tensors are calculated for modest-size oligopeptide fragments and subsequently transferred to peptides of the same local secondary structure. Alternatively, it has been suggested to calculate transferable vibrational parameters of single [71] and two adjacent [72] peptide residues as a function of the (ϕ, ψ) dihedral angles of the peptide backbone. The idea is similar to the exciton model, which has been widely employed to describe electronic excitations of extended systems composed of many similar repeat units, such as aggregates and molecular crystals. [44, 73]

Here, we adopt blocked glycine peptides of the type H_3C-CONH-(CH_2-CONH$)_n$-CH_3 as a simple building-block model, see Fig. 2.18. The smallest system of this series ($n = 0$) is N-methylacetamide (NMA), which has been widely used as a minimal model of a peptide [28, 29, 35–38, 74]. For $n = 1$ we obtain “glycine dipeptide” (GD), which is the simplest model that describes the vibrational interaction between two peptide units. For this system, various groups have calculated the matrix elements ε_n and β_{nm} as functions of the (ϕ, ψ) backbone dihedral angles [28, 31–34, 53, 72]. Following the building block idea, we wish to study to what extent the parameters derived for NMA and GD can be used

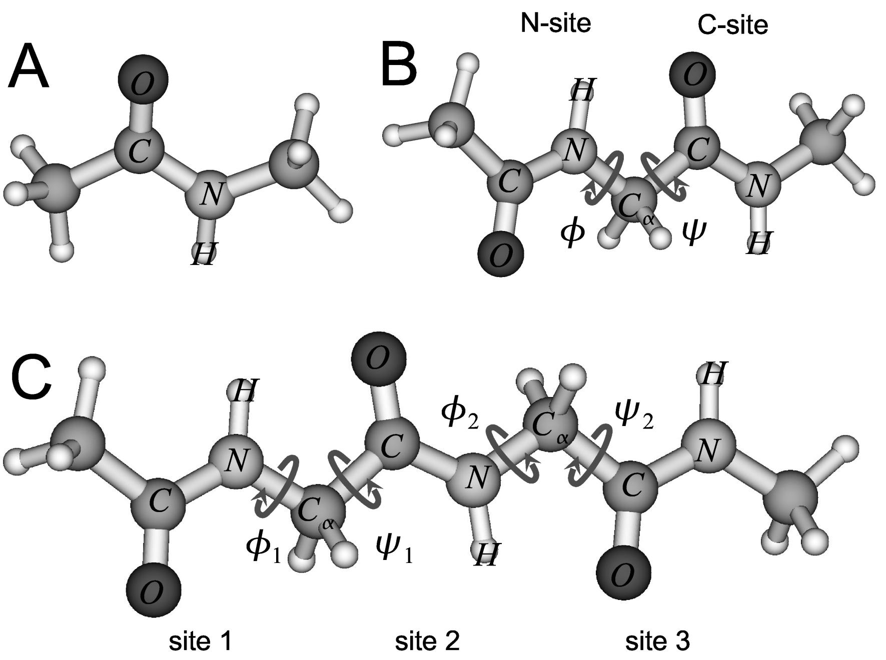

Figure 2.18: Blocked glycine peptides under consideration: **A** N-methylacetamide (NMA), **B** "glycine dipeptide" (GD), and **C** "glycine tripeptide" (GT).

to obtain accurate vibrational parameters for longer peptides ($n \geq 2$). To this end, we have performed extensive *ab initio* calculations of the case $n = 2$, henceforth referred to as "glycine tripeptide" (GT), which serve as reference data for testing various approximate schemes. We suggest and analyze various methods to calculate diagonal and off-diagonal vibrational matrix elements and compare the *ab initio* results to the prediction of popular electrostatic models [28, 37, 38, 41, 74].

In the building block model the values of the site energies are considered to be conditioned by an additive influence of the neighboring peptide unit(s) on the considered one. If the building block model is based on the usage of a dipeptide, only the first neighbor influence can be taken into account. One of the possible ways to improve the values of the site energies is to use the building block model in combination with electrostatic models for the calculation of the non-first neighbor influence. In such models the conformational changes of the local mode frequencies are related (in empirical way) with the electric field generated by other peptide units and located around the considered one. In the similar way, the usage of a dipeptide as the building block gives only the first neighbor coupling. Analogously to the case of the site energies, the non-first neighbor coupling can be calculated within different electrostatic models. Assuming that the vibrational mode coupling among localized vibrations originates mostly from electrostatic interaction, various workers have employed the transition dipole coupling model [4, 5] or multipole generalizations thereof. [31,32,75] In the present work we test popular electrostatic models [28,37,38,41,74] by comparison with the *ab initio* values. The same strategy is used to test the building block model as well as combination of the last one with the electrostatic models for the site energies. It has to be emphasized, that the mentioned electrostatic models, used for the calculation of the coupling and site energies, can also be applied for the calculation of the first neighbor coupling as well as the first neighbor effects on the site energies. However, the results obtained is such way are rather inaccurate. This makes the building block model indispensable for the construction of the polypeptides vibrational Hamiltonian.

2.4.1 Methods

Following previous work [53], we have performed DFT calculations of the isolated tripeptide GT at the B3LYP/6-31+G(d) theoretical level [64], using the GAUSSIAN suite of program [62]. Choosing a representative set of 360 conformations defined by the backbone dihedral angles ϕ_1, ψ_1, ϕ_2, and ψ_2 (see Fig. 2.18), we performed geometry optimizations of GT with fixed $\{\phi_i, \psi_j\}$ angles and subsequent normal mode analysis. A frequency scaling factor of 0.97 was used. Employing the C=O stretch modes of the peptide backbone as amide I local modes, next the transformation between normal and local amide I modes is constructed using the Hessian matrix reconstruction method [28, 29, 32]. When this transformation is applied to the diagonal normal mode Hamiltonian, we obtain the desired exciton model. In the same way, the (ϕ, ψ)-map of ε_n and β_{nm} of the dipeptide GD was calculated [53].

2.4.2 First-Neighbor Couplings

Let us first consider the description of the vibrational couplings β_{nm} of GT. An obvious approximation of the *next-neighbor* couplings $\beta_{n,n\pm1}$, depending in principle on all dihedral angles $\{\phi_i, \psi_j\}$ of the peptide, is to use the dipeptide couplings β^{GD} which depend only on a single pair of dihedral angles. In the case of the tripeptide we thus assume that

$$\beta_{n,n+1}(\phi_1, \psi_1, \phi_2, \psi_2) \approx \beta^{\mathrm{GD}}(\phi_n, \psi_n), \tag{2.73}$$

where $n = 1, 2$, see Fig. 2.18. The comparison of the resulting next-neighbor couplings β_{12} and β_{23} and their reference values from the direct calculations is shown in Fig. 2.19A. Approximation (2.73) is found to yield almost quantitative agreement with the reference calculations, although the additional residue of GT is expected to change the geometry

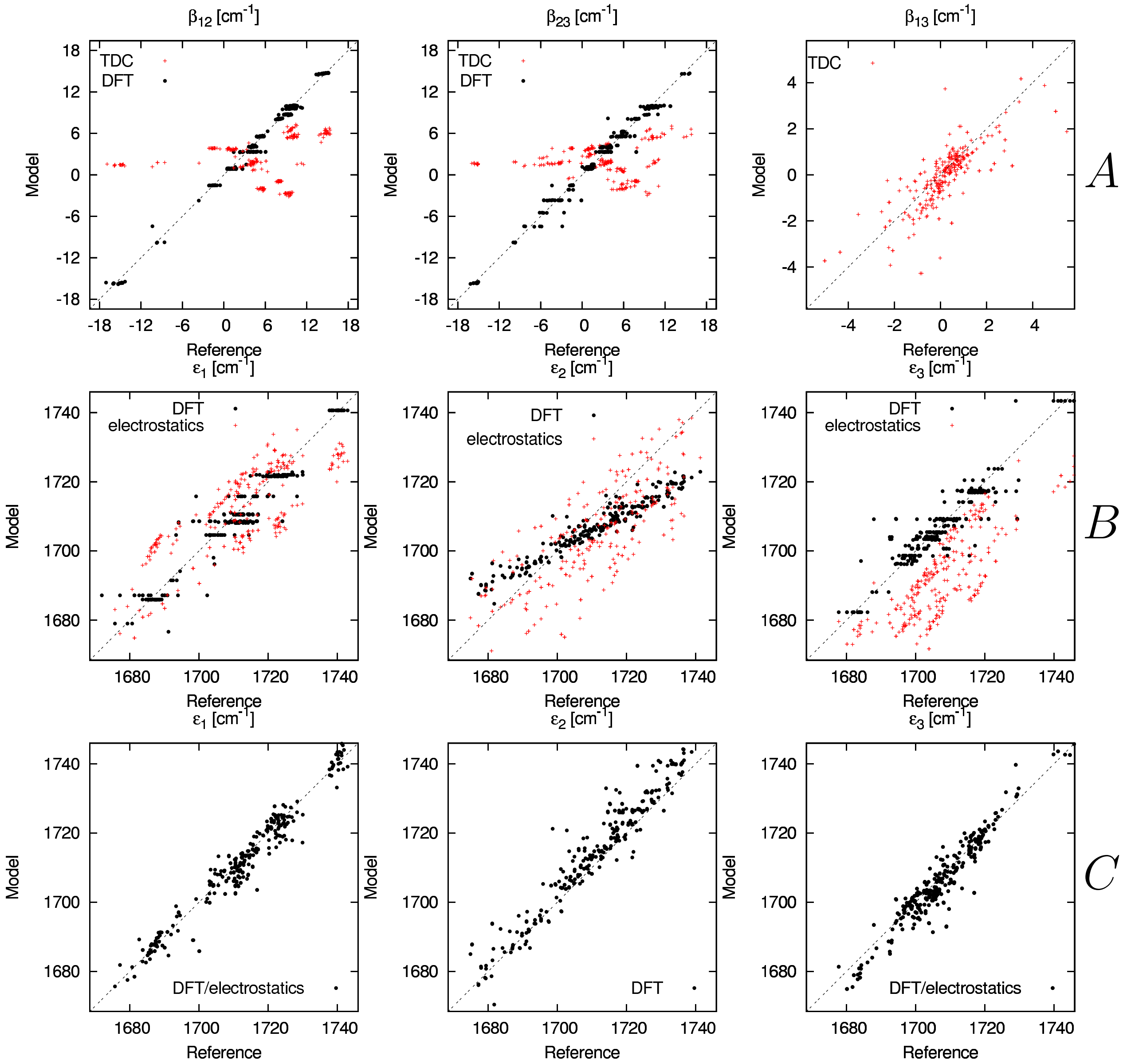

Figure 2.19: Amide I local-mode frequencies ε_n (panels B and C) and associated vibrational couplings β_{nm} (panels A) of glycine tripeptide. Compared are results obtained directly from DFT calculations ("Reference") and from various approximate schemes ("Model"), see text.

and the electronic structure of the system and consequently also the relative orientation of the two peptide units under consideration. The minor dependence of the next-neighbor couplings on the remaining dihedral angles [say, of β_{12} on the angles (ϕ_2, ψ_2)] is reflected in the "horizontal clustering" of some points, which refer to the same (ϕ_1, ψ_1) but different (ϕ_2, ψ_2) values, respectively. Even in these cases, though, the accuracy of approximation (2.73) is within a few wavenumbers, resulting in root mean squared deviations (RMSD) of 0.7 and 0.9 cm^{-1} for β_{12} and β_{23}, respectively.

It is instructive to compare the above study to the results of the popular transition dipole coupling (TDC) approximation [4, 76]

$$\beta_{nm}^{\mathrm{TDC}} = \frac{1}{4\pi\varepsilon_0}\left(\frac{\vec{\mu}_n \cdot \vec{\mu}_m}{r_{nm}^3} - 3\frac{(\vec{\mu}_n \cdot \vec{r}_{nm})(\vec{\mu}_m \cdot \vec{r}_{nm})}{r_{nm}^5}\right), \tag{2.74}$$

where the vector $\vec{r}_{nm}$ connects the two transition dipoles $\vec{\mu}_n$. The strength of the transition dipole is 0.37 D, it points towards the nitrogen atom with an angle of 20° with respect to the C=O bond, and its origin is located in between the carbon and oxygen atom with a distance of 0.868 Å from the carbon. In agreement with previous studies [31,32], Fig. 2.19A shows that the TDC approximation is not well suited to describe next-neighbor couplings. In particular, we find a significant (about a factor of five) underestimation of the amplitude of the coupling. This is because adjacent peptide units provide overlapping charge densities which cause through-bond interactions that are not accounted for by electrostatic models such as the TDC approximation or multipole generalization [32, 72] thereof.

2.4.3 Second-Neighbor Coupling

As an example of a "long-distance" (i.e., not next-neighbor) interaction, we next consider the calculation of the coupling β_{13}. Obviously, long-distance interactions cannot be derived from a dipeptide containing only a single next-neighbor coupling. In this case, however, it should be safe to assume that through-bond interactions are negligible, which justifies the use of the TDC approximation. Figure 2.19A reveals that the TDC model indeed reproduces the long-distance coupling β_{13} quite well, resulting in an RMSD of 0.7 cm^{-1}. Furthermore, we note that long-distance interactions are typically much (about a factor of five to ten) smaller than next-neighbor couplings and therefore provide only a minor correction to the vibrational exciton model.

2.4.4 Site Energies of the Terminal Residues

We turn to the calculation of the diagonal matrix elements of the model, that is, the local-mode frequencies ε_n. As indicated in Fig. 2.18, the dipeptide calculations provide an obvious approximation for the local-mode frequencies of the *terminal* residues, i.e.,

$$\begin{aligned} \varepsilon_1(\phi_1, \psi_1, \phi_2, \psi_2) &\approx \varepsilon_{\mathrm{N}}^{\mathrm{GD}}(\phi_1, \psi_1), \\ \varepsilon_3(\phi_1, \psi_1, \phi_2, \psi_2) &\approx \varepsilon_{\mathrm{C}}^{\mathrm{GD}}(\phi_2, \psi_2). \end{aligned} \tag{2.75}$$

Figure 2.19B reveals that this assumption results in a reasonable overall agreement of direct and model calculations, giving a RMSD of 4.3 and 4.6 cm^{-1} for ε_1 and ε_3, respectively. Similarly as found above for the vibrational couplings, however, the dependence of the terminal local-mode frequencies on the remaining dihedral angles [say, of ε_1 on the angles (ϕ_2, ψ_2)] show up as horizontal clustering of points with the same (ϕ_1, ψ_1) but different (ϕ_2, ψ_2) values, respectively.

Alternatively, the local-mode frequencies have been calculated from electrostatic models, which assume that the local amide I frequency shift $\delta\varepsilon_n = \varepsilon_n - \langle\varepsilon_n\rangle$ can be modeled via

$$\delta\varepsilon_n = \sum_i c_i \phi_{in}, \tag{2.76}$$

where $\phi_{in}(t)$ is the electrostatic potential at the position of the atom $i =$ C, O, N, and H of the nth residue, produced by the partial charges of all other peptide atoms [28, 37, 38]. As an representative example, we chose the popular four-site model of Cho and coworkers [28]. Figure 2.19B demonstrates that the electrostatic model only provides a qualitative description of the local-mode frequencies. Caused by the neglect of through-bond interactions, the RMSD is 11 and 18 cm^{-1} for ε_1 and ε_3, respectively.

The *ab-initio*-based approximation (2.75) takes into account the through-bond interaction between two local residues but neglects the effect of all other residues, while the electrostatic model (2.76) accounts for the electrostatic interactions of all residues but neglects through-bond effects. To combine the virtues of both approaches, one may use the *ab-initio* model to treat the local interactions and the electrostatic model to account for the effects of the remaining residues. As shown in Fig. 2.19C, this ansatz lifts the degeneracy of the *ab-initio* local-mode frequencies with different conformation of the additional residue, and leads to a improved agreement with an RMSD of 3.5 and 3.9 cm^{-1} for ε_1 and ε_3, respectively.

2.4.5 Site Energies of the Inner Peptide Unit

Finally, we consider the calculation of the local-mode frequency ε_2 of the *inner* peptide unit. Considering longer peptides, all but the two terminal residues are inner peptide units, meaning that the accurate modeling of these residues is essential. While it is straightforward to calculate ε_2 using the electrostatic model (2.76), Fig. 2.19B shows that this approach only yields a qualitative description with a relatively large RMSD of 14 cm^{-1}. On the other hand, when we try to achieve an *ab-initio* approximation similar to Eq. (2.75), we face the problem that ε_2 clearly depends on *both* pairs of dihedral angles, (ϕ_1, ψ_1) and (ϕ_2, ψ_2). A simple ansatz that accounts for both contributions in the same way is to consider the average

$$\varepsilon_2(\phi_1, \psi_1, \phi_2, \psi_2) = \tfrac{1}{2}\left(\varepsilon_C^{\mathrm{GD}}(\phi_1, \psi_1) + \varepsilon_N^{\mathrm{GD}}(\phi_2, \psi_2)\right). \tag{2.77}$$

However, Fig. 2.19B reveals that the resulting local-mode frequencies systematically increase too little compared to the reference calculations, giving a RMSD of 7.9 cm^{-1}.

To motivate a more physical approximation, we consider various limiting cases in which one of the two terminal residues is separated form the remaining dipeptide (and the missing end groups CH_3 or H are added). For example, by cutting off the left residue of GT, we obtain NMA on the left side and the dipeptide GD on the right side (see Fig. 2.18) and we find that $\varepsilon_2 = \varepsilon_N^{\mathrm{GD}}(\phi_2, \psi_2)$. In the same way, we obtain $\varepsilon_2 = \varepsilon_C^{\mathrm{GD}}(\phi_1, \psi_1)$ if the right-hand-side residue is cut off, and $\varepsilon_2 = \varepsilon^{\mathrm{NMA}}$ if both terminal residues are removed. The three cases can be cast in the expression

$$\begin{aligned}\varepsilon_2(\phi_1, \psi_1, \phi_2, \psi_2) = \varepsilon^{\mathrm{NMA}} \quad &+ \quad \gamma_1\left(\varepsilon_C^{\mathrm{GD}}(\phi_1, \psi_1) - \varepsilon^{\mathrm{NMA}}\right) \\ &+ \quad \gamma_2\left(\varepsilon_N^{\mathrm{GD}}(\phi_2, \psi_2) - \varepsilon^{\mathrm{NMA}}\right),\end{aligned} \tag{2.78}$$

where the value of γ_1 indicates if the left residue is removed ($\gamma_1 = 0$) or not ($\gamma_1 = 1$) and the value of γ_2 indicates if the right residue is removed. Interestingly, the interpolation scheme (2.78) also covers the case of the tripeptide (by setting $\gamma_1 = \gamma_2 = 1$). This way we obtain an approximation of the local-mode frequency ε_2 of the inner peptide unit, which exhibits the correct behavior in the limiting cases discussed above. Indeed, Fig. 2.19C shows that the interpolation scheme represents a clear improvement (5.9 cm^{-1} RMSD) over both the the electrostatic model (2.76) as well as the heuristic ansatz (2.77). Interpreting $(\varepsilon_C^{\mathrm{GD}}(\phi_1, \psi_1) - \varepsilon^{\mathrm{NMA}})$ and $(\varepsilon_N^{\mathrm{GD}}(\phi_2, \psi_2) - \varepsilon^{\mathrm{NMA}})$ as the frequency shifts of the local mode ε_2 due to the N and C sites, respectively, an equivalent expression has recently been suggested by Jansen *et al.* [72].

2.4.6 Concluding remarks

We have studied the virtues and limits of an *ab initio*-based building block model of amide I vibrations in peptides. An accurate *ab initio* description has been found essential to account for the effect of next-neighbor residues on the local-mode frequencies ε_n and the vibrational couplings β_{nm}, respectively, while the long-distance (i.e., not next-neighbor) interactions can be well approximated by electrostatic models. Employing this combination, we typically obtain an accuracy of a few wavenumbers for ε_n and less than wavenumber for β_{nm}. As the vibrational data are obtained for peptide conformations with arbitrary backbone dihedral angles $\{\phi_n, \psi_n\}$, the model is capable of describing large conformational rearrangements occurring, e.g., in protein binding or folding processes.

We note that approximations (2.75) and (2.78) are in the spirit of systematic fragmentation schemes recently suggested by Zhang [68] and Collins [69] and their coworkers. While we choose a dipeptide as smallest fragmentation unit, it is clear that the model in principle can be converged to the exact result if larger molecular fragments are considered. For example, one could use the above results for the tripeptide to calculate the vibrational response of polypeptides using an analogous decomposition into tripeptide units [69].

2.5 Transferability of Maps

It is interesting to investigate to what extent the (ϕ, ψ)-maps calculated above for GD (Ac-Gly-NHCH$_3$) are transferable to peptides with other side chains or end groups.

2.5.1 Effect of Side Chains

We have considered the systems Ac-Asp-NHCH$_3$ (as an example for a peptide with a hydrophilic side chain) and Ac-Phy-NHCH$_3$ (as an example for a peptide with a hydrophobic side chain). Plotted as a function of the dihedral angle ψ and $\phi = -60°$, Fig. 2.20 compares the vibrational constants obtained for the three dipeptide analogs.

The intersite coupling β is almost identical in all three cases, showing again that this quantity is quite insensitive. Moreover, the frequency splittings $\Delta\omega$ are found to coincide nicely for all three dipeptides. The site energies ε_1 and ε_2, on the other hand, are somewhat shifted depending on the side chain, but nevertheless show similar conformational dependence.

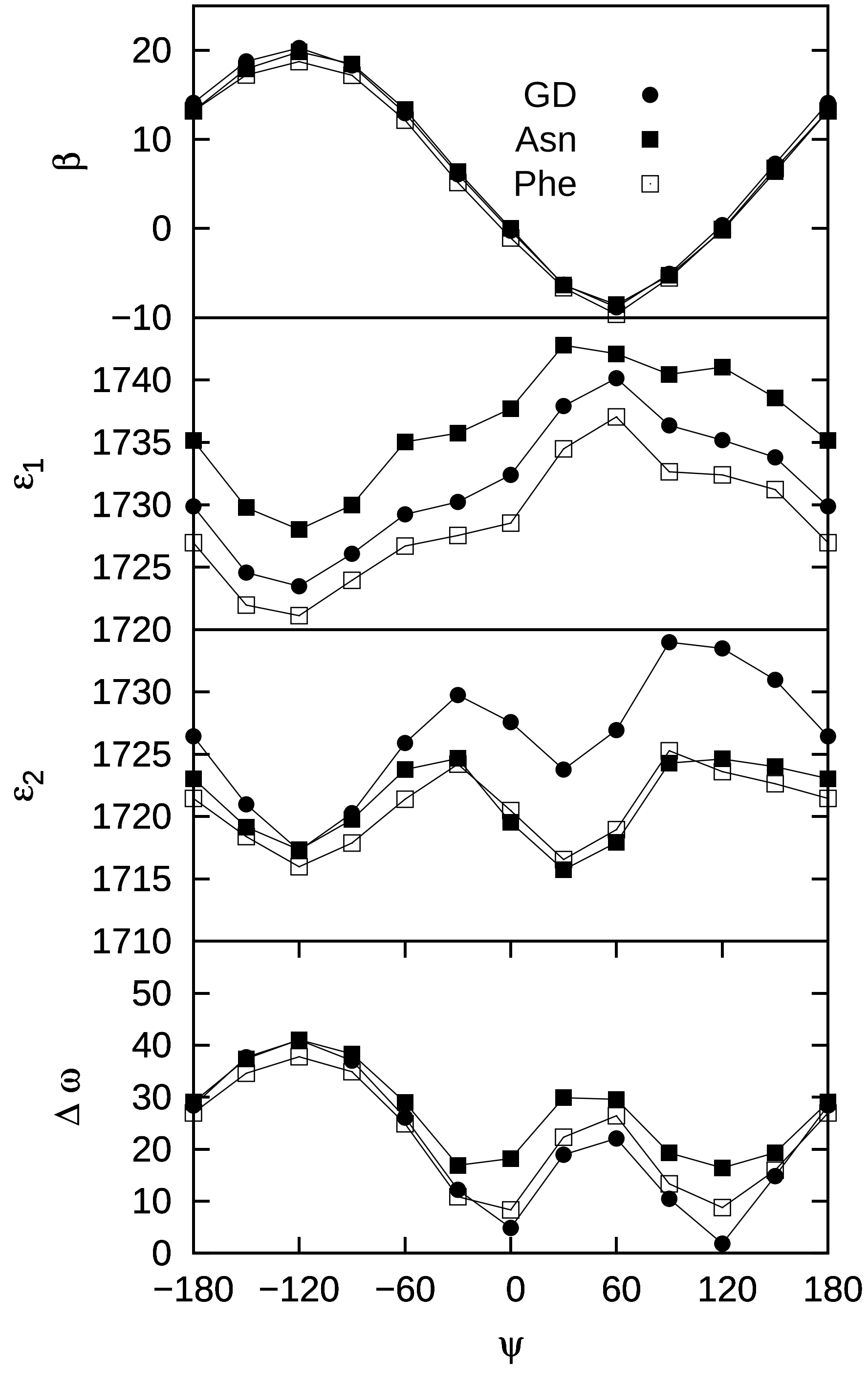

Figure 2.20: Transferability of the amide I vibrational constants for Ac-Gly-NHCH$_3$ (GD) to peptides with a hydrophilic side chain Ac-Asp-NHCH$_3$ (Asp) and a hydrophobic side chain Ac-Phy-NHCH$_3$ (Phy), respectively. Shown are (in cm^{-1}) the intersite coupling β, the site energies ε_1 and ε_2, as well as the frequency gap $\Delta\omega = \omega_+ - \omega_-$.

2.5.2 Effect of End Groups (Different Protonation States)

To study the effect of various end-groups, we have considered trialanine [which like GD is described by a single pair of flexible backbone dihedral angles (ϕ, ψ)] in its zwitterionic (A_3^{+-}), cationic (A_3^{+}), and anionic (A_3^{-}) state. As shown in Fig. 2.21, the intersite coupling β is hardly affected by the presence of the amino and carboxyl termini and by the different protonation states of trialanine. As may be expected, however, the site energies of trialanine only approximately match the results for GD. The frequency splitting $\Delta\omega$ again depends only weakly on the protonation state of trialanine. Assuming that trialanine occurs predominantly in the P_{II} conformation at $\psi = 140°$, [1] we obtain $\Delta\omega \approx 20$ cm^{-1} for A_3^{+} and A_3^{+-}, as well as 15 cm^{-1} for A_3^{-}. This is in qualitative agreement with the experimental results $\Delta\omega \approx 25$ cm^{-1} for A_3^{+} and A_3^{+-} and 11 cm^{-1} for A_3^{-}. [6, 7]

2.5.3 Comparison of GD and AAA Maps

In order to better understand the origin of the difference between the vibrational properties of different peptides we have compared the *ab initio* maps of GD and A_3^{+} in more details. Figure 2.22 shows the (ϕ, ψ)-maps of mean frequency $\bar{\omega} = (\omega_1 + \omega_2)/2 = (\varepsilon_1 + \varepsilon_2)/2$ and of the frequency splitting $\Delta\omega$ of glycine dipeptide and A_3^{+}. Also indicated are the three main conformational states α_R, β, and P_{II}. It is seen that both $\bar{\omega}$ and $\Delta\omega$ may significantly differ from glycine dipeptide to A_3^{+}. While these quantities look relatively similar in the populated regions α_R, β, and P_{II}, the maps deviate considerably for the corner regions around $(-180, -180)$, $(0, -180)$, and $(0, 180)$. The latter is mainly a consequence of intramolecular hydrogen bonds that are formed in A_3^{+} but not in glycine dipeptide. Using the atom labeling introduced in Fig. 1.2, these bonds are formed between the atoms (i) O2-H3 at $(-60 < \phi < 0, -180 < \psi < -150)$, (ii) O3-H3 at $(-75 < \phi < 0, -75 < \psi < -30)$ and $(-75 < \phi < -60, -15 < \psi < 45)$, (iii) O3-H1 at $(-180 < \phi < -135, -60 < \psi < -30)$, and (iv) O1-H2 at $(-60 < \phi < 0, 0 < \psi < 30)$, where only the latter hydrogen bond also occurs in glycine dipeptide. While these intramolecular hydrogen bonds hardly exist in aqueous solution, they only occur in the α_R conformational state and do not affect the description of the (mostly populated) states P_{II} and β.

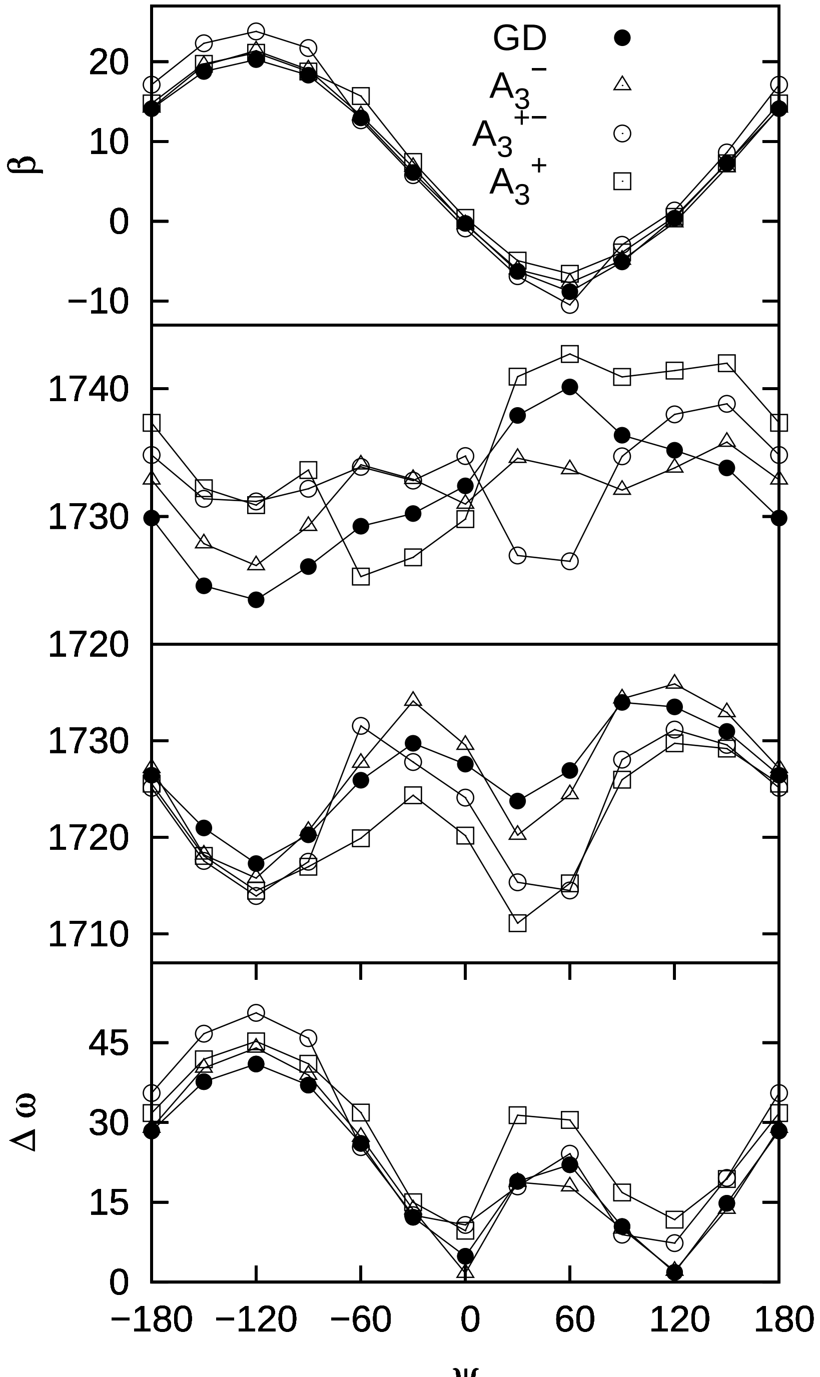

Figure 2.21: Transferability of the amide I vibrational constants for Ac-Gly-NHCH$_3$ (GD) to trialanine in its zwitterionic (A_3^{+-}), cationic (A_3^+), and anionic (A_3^-) state. Shown are (in cm^{-1}) the intersite coupling β, the site energies ε_1 and ε_2, as well as the frequency gap $\Delta\omega = \omega_+ - \omega_-$.

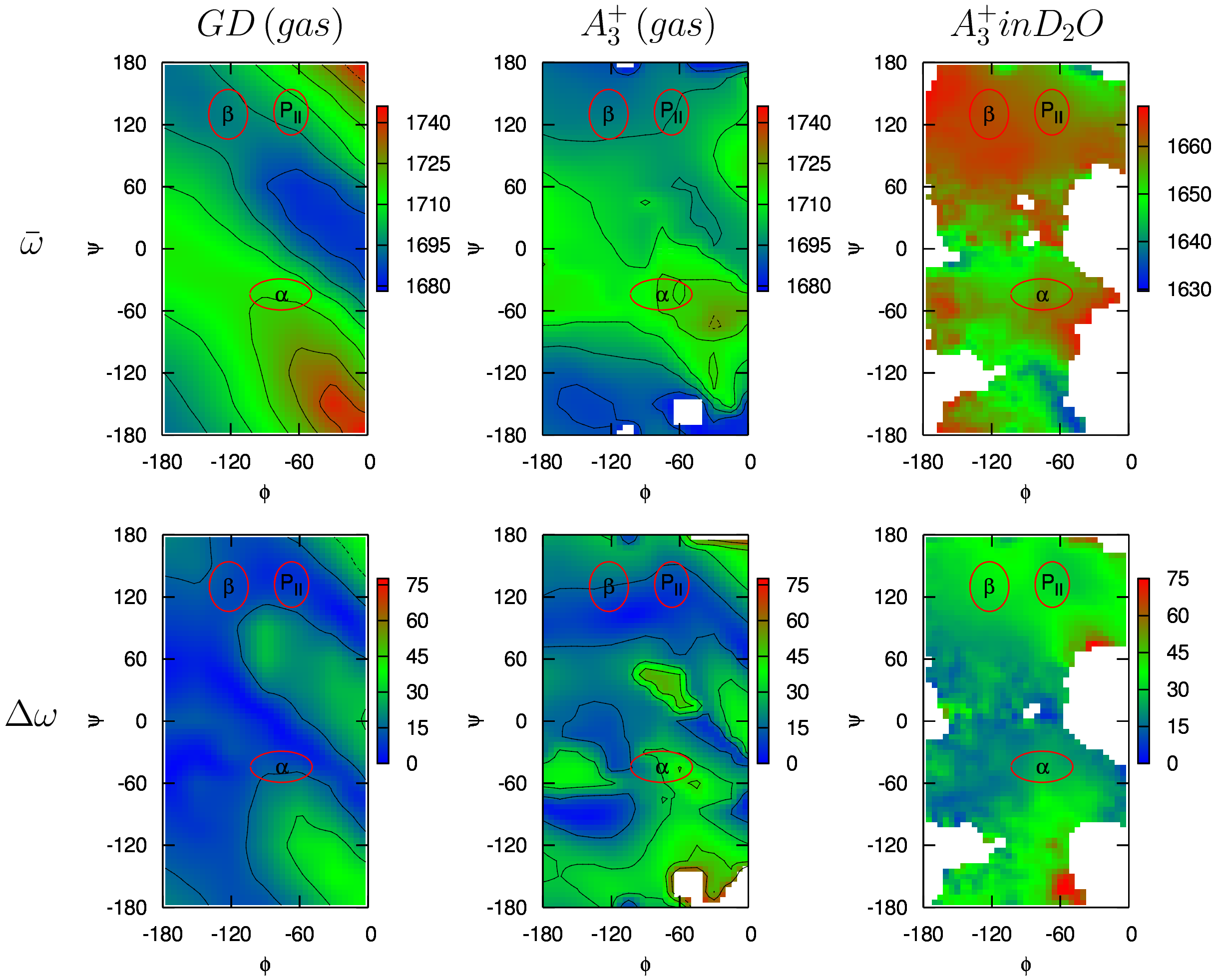

Figure 2.22: (ϕ, ψ)-maps of the mean $\bar{\omega}$ (top) and the splitting $\Delta\omega$ (bottom) of the two amide I frequencies, as obtained for isolated glycine dipeptide (left), isolated trialanine (middle), and trialanine in D$_2$O (right).

Chapter 3

Calculation of Infrared Spectra

3.1 Trialanine

In this section we apply the above developed vibrational Hamiltonian for calculation of the vibrational spectra of cationic trialanine (A_3^+). A_3^+ is a small peptide with two peptide bonds and one set of backbone dihedral angles (ϕ, ψ), see Fig. 1.2. Driven by a number of experimental [1,6,13–15,49,54] and theoretical [20,50–53] studies, trialanine has emerged as a paradigm to study conformational dynamics of a small peptide in aqueous solution. In a seminal paper, Woutersen and Hamm [1] have presented a two-dimensional IR study, which suggested that A_3^+ mostly adopts a conformation around $(\phi, \psi) \approx (-60°, 140°)$, known as poly-L-proline II (P_{II}) structure. A recent joint NMR/MD study [54] confirmed this finding, showing that A_3^+ is ≈ 90 % in P_{II}, ≈ 10 % in the extended conformation β located at $\approx (-120°, 130°)$, but hardly (< 5 %) in the right-handed helix conformation α_R located at $\approx (-80°, -50°)$. The availability of accurate experimental thermal population probabilities is important, because it has been found [51, 52, 77] that different MD force fields may yield quite different thermal populations for A_3^+. We adopt A_3^+ to test the above constructed vibrational Hamiltonian as well as a set of common spectroscopic approximations, which is the main goal of the given chapter. In particular, the following aspects were analyzed. First, we study a sensitivity of the amid I frequency distributions on the *ab initio* map of the vibrational properties. For that we compare the frequency distributions calculated with the maps of A_3^+ and "glycine dipeptide" (Ac-Gly-NHCH$_3$), which often is considered as generic example in amide I spectroscopy. [28, 31–34] The influence of the gas phase contribution was compared with the influence of the solvent. As there is no specific parametrization of the solvent-induced frequency shift available for A_3^+, we adopt and compare various *ab initio*-based models of solvated N-methylacetamide, which are found to yield different results for the solvent shift. [28, 29, 35–38] Finally, we present a detailed discussion of the validity of the various approximations usually employed in the calculation of condensed-phase vibrational spectra, including the adiabatic, the Franck-Condon, and the second-order cumulant approximations, respectively.

3.1.1 Semiclassical Line Shape Theory

Following semiclassical line shape theory, [43] in the following we briefly derive explicit expressions for the vibrational absorption spectrum. In particular, we introduce all approximations discussed below in the computational results. Adopting a local mode representation with basis states $|n\rangle$, the exciton model Hamiltonian of trialanine can be written as [1]

$$H = \sum_{n=0,N,C} \varepsilon_n |n\rangle\langle n| + \beta\{|N\rangle\langle C| + |C\rangle\langle N|\}, \tag{3.1}$$

comprising the ground state $|0\rangle$ as well as the two excited vibrational local-mode states $|N\rangle$ and $|C\rangle$ (see Fig. 1.2). Restricting ourselves to linear IR absorption, higher vibrational excitations (e.g., two-exciton states) can be neglected. The vibrational dipole operator of the system reads

$$\hat{\boldsymbol{\mu}} = \sum_{n=N,C} \boldsymbol{\mu}_n \{|0\rangle\langle n| + |n\rangle\langle 0|\}, \tag{3.2}$$

where $\boldsymbol{\mu}_n$ denote the transition dipole moment of the nth local mode.

Employing time-dependent perturbation theory, the linear IR absorption spectrum $\sigma(\omega)$ is proportional to the Fourier transform of the dipole autocorrelation function [44]

$$\begin{aligned} \sigma(\omega) &= \int_{-\infty}^{\infty} dt e^{-i\omega t} e^{-|t|/2T_1} C(t), && (3.3)\\ C(t) &= \langle 0|\hat{\boldsymbol{\mu}}(t)\hat{\boldsymbol{\mu}}(0)|0\rangle. && (3.4)\end{aligned}$$

Here the phenomenological decay term accounts for the finite life time $T_1 = 1$ ps of the excited amide I vibrations. [11] Furthermore we have introduced the Heisenberg dipole operator $\hat{\boldsymbol{\mu}}(t) = U^{\dagger}(t)\hat{\boldsymbol{\mu}}U(t)$ with time evolution operator $U(t)$, and we have assumed that the amide I system is initially in its ground state $|0\rangle$.

Within semiclassical line shape theory, [43] the fluctuations of the peptide and the surrounding solvent molecules result in a classical time-dependence of the Hamiltonian $H(t)$ via its matrix elements $\beta(t)$ and $\varepsilon_n(t)$ ($n = N, C$; we set $\varepsilon_0 = 0$ without loss of generality). We then obtain for the dipole autocorrelation function ($\hbar \equiv 1$)

$$C(t) = \sum_{n,m=N,C} \left\langle \boldsymbol{\mu}_n(t) \langle n| \exp_+ \left\{ -i \int_0^t H(\tau) d\tau \right\} |m\rangle \boldsymbol{\mu}_m(0) \right\rangle, \tag{3.5}$$

where $\langle \ldots \rangle$ denotes a statistical average over the classical fluctuations. Note that also the transition dipole moments $\boldsymbol{\mu}_n(t)$ may explicitly depend on time due to the fluctuations of the peptide. Solving numerically the time-dependent Schrödinger equation with respect to $H(t)$, the dipole autocorrelation function C(t) can be obtained from Eq. (3.5) without further approximations. Hereafter, this will therefore be referred to as "direct calculation."

As the averaging over the rapidly oscillating exponential function may be rather cumbersome, it is convenient invoke further simplifications. With this end in mind, we introduce the instantaneous eigenstates $|\psi_k(t)\rangle$ of the time-dependent Hamiltonian which satisfy

$$H(t)|\psi_k(t)\rangle = \omega_k(t)|\psi_k(t)\rangle \tag{3.6}$$

with $k = 1, 2$, where $\omega_k = \frac{1}{2}(\varepsilon_N + \varepsilon_C) + \frac{1}{2}(-1)^k \sqrt{(\varepsilon_N - \varepsilon_C)^2 + 4\beta^2}$. Inserting Eq. (3.6) into the propagator $\exp_+ \left\{ -i \int_0^t H(\tau) d\tau \right\}$, we obtain

$$\begin{aligned} & \exp_+ \left\{ -i \sum_k \int_0^t |\psi_k(\tau)\rangle\langle\psi_k(\tau)| \; \omega_k(\tau) d\tau \right\} \\ \approx \quad & \exp_+ \left\{ -i \sum_k |\psi_k(t)\rangle\langle\psi_k(t)| \int_0^t \omega_k(\tau) d\tau \right\} \\ = \quad & \sum_k |\psi_k(t)\rangle\langle\psi_k(t)| \; \exp\left\{ -i \int_0^t \omega_k(\tau) d\tau \right\}, \end{aligned} \tag{3.7}$$

where in the second line the explicit time dependence of the eigenstates was neglected, which corresponds to the adiabatic (or "vibrational Born-Oppenheimer") approximation. [20, 34, 48] The adiabatic approximation is usually justified if the eigenenergies $\omega_k(t)$ do not come close to each other during their time evolution. The approximation reduces the time-ordered exponential operator in Eq. (3.5) to a simple exponential function. Insertion of Eq. (3.7) in Eq. (3.5) then yields

$$C(t) = \sum_{k=1,2} e^{-i\langle\omega_k\rangle t} \, \langle M_k(t)\phi_k(t)\rangle \,, \tag{3.8}$$

$$M_k(t) = \sum_{n,m=N,C} \boldsymbol{\mu}_n(t)\boldsymbol{\mu}_m(0)\langle n|\psi_k(t)\rangle\langle\psi_k(t)|m\rangle, \tag{3.9}$$

$$\phi_k(t) = \exp\left\{ -i \int_0^t \delta\omega_k(\tau) d\tau \right\}, \tag{3.10}$$

where $\delta\omega_k(t) = \omega_k(t) - \langle\omega_k\rangle$, $M_k(t)$ describes the autocorrelation function of the transition dipole moments, and the function $\phi_k(t)$ accounts for the quantum-mechanical time evolution of the system. Eqs. (3.8)-(3.10) will be referred to as "adiabatic approximation."

As Eq. (3.10) still requires to average over a rapidly oscillating function, we next introduce the following two approximations:

$$C(t) = \sum_{k=1,2} e^{-i\langle\omega_k\rangle t} \langle M_k(t)\rangle \langle\phi_k(t)\rangle \, , \tag{3.11}$$

$$\langle\phi_k(t)\rangle = \exp\left\{ \int_0^t d\tau (t-\tau) \langle\delta\omega_k(\tau)\delta\omega_k(0)\rangle \right\} . \tag{3.12}$$

In Eq. (3.11), the classical average over $M_k(t)$ and $\phi_k(t)$ has been factorized in two separate averages. This approximation allows us to introduce in Eq. (3.12) a second-order cumulant expansion, [44] which advantageously shifts the average to the exponent, thus introducing the mean $\langle\omega_k\rangle$ and the autocorrelation function $\langle\delta\omega_k(\tau)\delta\omega_k(0)\rangle$ of the normal mode frequency ω_k. For brevity, we will refer to Eq. (3.12) as "cumulant approximation."

Furthermore, one may introduce various approximations to evaluate the function $\langle M_k(t)\rangle$ which contains the time-dependent overlap between local-mode and normal-mode basis functions as well as the autocorrelation function of the transition dipole moments. The latter contains (i) the absolute value of the transition dipole moments $|\boldsymbol{\mu}_n(t)|$ and (ii) the time-dependent relative orientation of the two transition dipoles [via $\boldsymbol{\mu}_1(t) \cdot \boldsymbol{\mu}_2(t)$]. As suggested by recent *ab initio* studies [23], we assume that $|\boldsymbol{\mu}_n(t)| = const.$, which appears to be a reasonable assumption for the local-mode representation. Neglecting furthermore the time dependence of overlap functions and of the dipoles' relative orientation, we obtain

$$\langle M_k(t)\rangle = \langle M_k(0)\rangle \tag{3.13}$$

which in the following will be referred to as Franck-Condon approximation.

For interpretative purposes, it is often instructive to consider all quantities for each conformational state α_{R}, β, and $\mathrm{P_{II}}$ of the peptide separately. For example, we may decompose the absorption spectrum as

$$\sigma(\omega) = \sum_{s=\alpha,\beta,P_{\mathrm{II}}} \sum_{k=1,2} \sigma_{ks}(\omega) \tag{3.14}$$

and calculate the state-specific absorption bands $\sigma_{ks}(\omega)$ via the Fourier transform in Eq. (3.3) of the dipole correlation functions

$$C_{ks}(t) = e^{-i\langle\omega_k\rangle_s t} \langle M_k(t)\rangle_s \langle\phi_k(t)\rangle_s \, , \tag{3.15}$$

where $\langle\ldots\rangle_s$ denotes the average over all molecular geometries pertaining to the conformational state s.

Finally, as the simplest means to estimate the IR absorption spectrum, we may assume that the fluctuations of the peptide and the solvent occur on a very slow time scale. In this inhomogeneous limit, the state-specific absorption bands defined in Eq. (3.14) are given by $\sigma_{sk}(\omega) \approx \langle M_k(0) \rangle_s \, \rho_{sk}(\omega)$, where

$$\rho_{sk}(\omega) = \frac{1/2T_1}{(\omega - \langle \omega_k \rangle_s)^2 + (1/2T_1)^2} \tag{3.16}$$

represents the (life-time broadened) frequency distribution of the kth normal mode.

To summarize, we have introduced a series of approximations to calculate the IR absorption spectrum of coupled fluctuating IR dipoles. Starting from semiclassical line shape theory [Eq. (3.5)], we have first invoked the adiabatic approximation [Eq. (3.8)]. Assuming that the classical average factorizes [Eq. (3.11)], we introduced a second-order cumulant expansion [Eq. (3.12)]. Furthermore, the Franck-Condon approximation [Eq. (3.13)] and the inhomogeneous limit [Eq. (3.14)] may be employed.

3.1.2 Solvent-Induced Frequency Shift

To estimate the amide I frequency shift due to aqueous solvation, various groups have performed extensive *ab initio* calculations of N-methylacetamide (NMA) and several surrounding D_2O molecules. [28,35–38] Generating an ensemble of representative structures of solvated NMA from MD calculations, geometry optimizations of NMA with fixed solvent and subsequent normal mode analysis were performed. It was found that the resulting solvent-induced frequency shift $\delta\varepsilon$ correlates well with the electrostatic potential (or the electrostatic field) of the solvent molecules. This finding can be cast into the expansion

$$\delta\varepsilon(t) = \sum_i c_i \phi_i(t), \tag{3.17}$$

where $\phi_i(t)$ is the electrostatic potential produced by the partial charges of all water molecules at the position of the ith atom of NMA, i.e., i = CH_3(C), C, O, N, H, and CH_3(N). Depending on the details of the *ab initio* calculations (e.g., number of MD snapshots and the number of water molecules considered in the fit), various parametrizations of the coefficients c_i in Eq. (3.17) have been suggested. [28, 35–38] In this work we have adopted the six-site model of Skinner and coworkers, [38] because its parametrization includes a large number of MD snapshots as well as a large number of water molecules. [Six-site means that all six atoms are considered in Eq. (3.17), whereas four-site models only account for the C, O, N, and H atoms.] The electrostatic potential was calculated by using either a simple cut-off scheme or the particle mesh Ewald method, [78] which virtually gave the same results.

First we study the effect of the solvent-induced frequency shift on the amide I response of trialanine. We begin by comparing (ϕ, ψ)-maps of the mean frequency $\bar{\omega}$ and the frequency splitting $\Delta\omega$ in the gas phase and in aqueous solution. Figure 2.22 shows these maps which were obtained by averaging all values of $\varepsilon_n = \varepsilon_n^{\mathrm{gas}} + \delta\varepsilon_n$ for a given (ϕ, ψ) along the 100 ns MD trajectory. First, it is noted that large regions of the Ramachandran plot (drawn in white) are not sampled at all by the MD trajectory. As a consequence, the above discussed intramolecular hydrogen bonds (which in the gas phase occur mostly in these regions) are not formed in aqueous solution. The mean amide I frequencies of trialanine undergo an overall redshift and the frequency splitting is increased compared to the gas phase. The latter effect is also seen from the distribution of the frequency splittings $\Delta\omega$ shown in Fig. 3.2, which exhibit a substantial broadening in solution.

Finally, we consider the effect of the solvent on the distributions $\rho_{ks}(\omega)$ ($s = \alpha_{\mathrm{R}}$, β, $\mathrm{P_{II}}$) of the two amide I normal-mode frequencies shown in Fig. 3.1. Again, we find a significant red-shift and a large broadening of the spectral densities. In fact, the resulting overall density of Eq. (3.14) is much broader than the experimental amide I spectrum (see Fig. 3.5), which indicates that motional narrowing effect are important for the amide I spectrum of trialanine.

3.1.3 Distributions of Vibrational Frequencies

The first and maybe most difficult task of the theoretical description is to obtain the correct distributions of vibrational frequencies. In the present work, this was done in three steps. In order to sample all relevant conformations of the peptide and the surrounding solvent molecules, first a classical all-atom MD simulation was performed (100 ns NTP, GROMOS96 [79] and SPC [80] force fields). In the spirit of the adiabatic approximation explained above and within the harmonic approximation, the vibrational frequencies of the isolated peptide are then obtained from a normal mode calculation following a geometry optimization at fixed (ϕ, ψ) dihedral angles (at B3LYP/6-31+G(d) level [64]). Third, the solvent-induced frequency shift was calculated using a DFT-based model [38] that correlates the frequency shift with the electrostatic potential of the solvent molecules. Put together, we obtain a trajectory of normal mode states $\{|\psi_{ks}(t)\rangle\}$ and frequencies $\{\omega_{ks}(t)\}$ for the conformations $s = \alpha_{\mathrm{R}}$, β, and $\mathrm{P_{II}}$. Finally the thermal weights of the peptide conformations were obtained from accurate NMR experiments. [54]

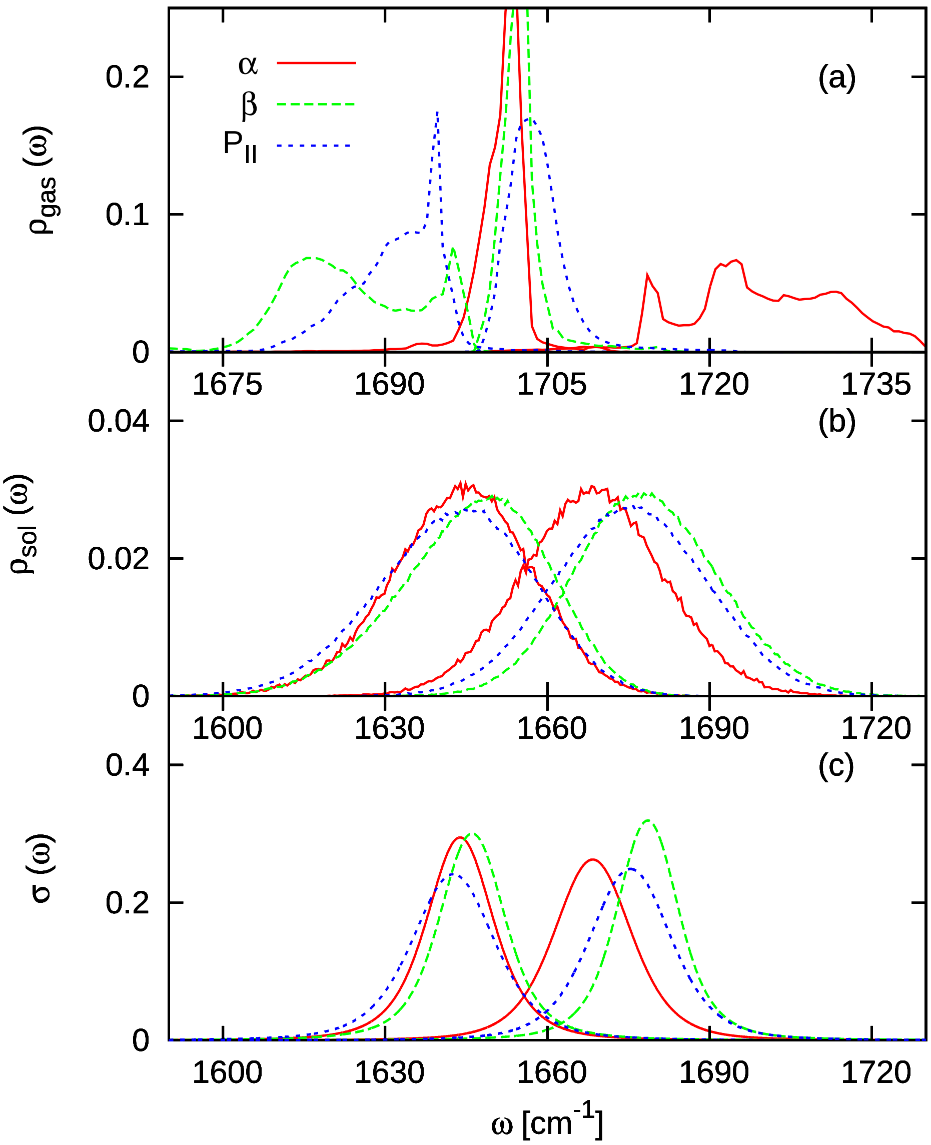

Figure 3.1: Distribution of the amide I normal-mode frequencies obtained for the three conformational states of trialanine (a) in the gas phase and (b) in solution. Panel (c) shows the corresponding absorption bands calculated within the cumulant approximation.

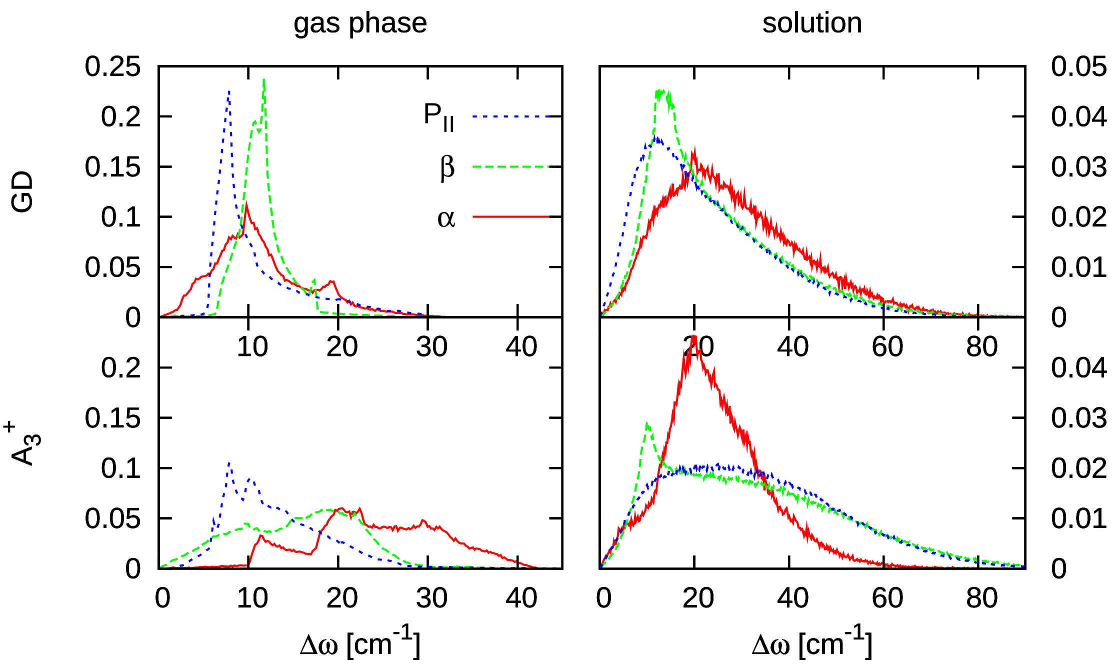

Figure 3.2: Distribution of the frequency splitting $\Delta\omega$ without (left) and with (right) the inclusion the solvent contribution, as obtained for the three conformational states of glycine dipeptide (top) and trialanine (bottom).

First, it is instructive to study how the different *ab initio* maps are reflected in the frequency distributions obtained by sampling the (ϕ, ψ)-maps along a MD trajectory in a specific conformational state (for more details about the MD simulations see [81]). As an example, Fig. 3.2 shows the gas-phase distributions of the frequency splitting $\Delta\omega$ corresponding to the states $\mathrm{P_{II}}$, β, and α_{R}. As may be expected from Fig. 2.22, the distributions for A_3^+ and glycine dipeptide are quite similar for the $\mathrm{P_{II}}$ state and differ mostly in the α_{R}-helical state. Furthermore, the $\Delta\omega$ distributions of A_3^+ suggest that the gas-phase frequency splittings in principle facilitate a straightforward discrimination of the three conformational states. The latter is confirmed by considering the corresponding distributions $\rho_{ks}(\omega)$ [$k = 1, 2$, $s = \alpha_{\mathrm{R}}$, β, $\mathrm{P_{II}}$, see Eq. (3.14)] of the amide I normal-mode frequencies shown in Fig. 3.1. Although the distributions of all three conformational states of trialanine overlap around $\omega \approx 1750$ cm^{-1}, they clearly differ at the red and blue ends of the spectrum.

Let us now consider the frequency distributions shown in Fig. 3.1. The obtained dependencies reveal that intramolecular and solvent-induced frequency shifts result in a conformation-dependent frequency splitting and broadening of the normal-mode lines. In both cases, the contribution of the solvent was more important than the intramolecular contribution. A comparison of various models for the solvent-induced frequency shift revealed significantly different results [81], which suggest that further improvements of the theoretical description of solvent-induced frequency shifts is desirable. Furthermore, we found various problems associated with the separate calculation of intramolecular and solvent contributions. First, the gas-phase calculations yielded intramolecular hydrogen bonds that would not occur in aqueous solvent (see Fig. 2.22). Moreover, we neglect fluctuations caused by the correlated motion of peptide and solvent, which presumably is at least partly responsible for the missing line width of the calculated spectrum (see Fig. 3.5).

3.1.4 Calculation of the Absorption Spectrum

In the following, we wish to go beyond the static picture of inhomogeneous averaging [Eq. (3.14)] and consider the effect of molecular dynamics on the IR spectrum. Hereby we proceed from simple to more exact formulations. At the next higher level of semiclassical line-shape theory, the line broadening is described in cumulant approximation [Eq. (3.12)]. To facilitate the direct comparison with the distributions $\rho_{ks}(\omega)$, we calculated the state-specific absorption bands $\sigma_{ks}(\omega)$ [Eq. (3.14)] by setting $\langle M_k(t) \rangle_s = 1$ and using the functions $\langle \phi_k(t) \rangle_s$ as given by Eq. (3.12). Comparing the results for $\sigma_{ks}(\omega)$ to the corresponding results for the frequency distributions $\rho_{ks}(\omega)$, Fig. 3.1 reveals that the frequency fluctuations affect a narrowing of the spectra by about a factor of two. The reason for this considerable effect is obtained from an analysis of the corresponding frequency autocorrelation functions $\langle \delta\omega_k(t)\delta\omega_k(0) \rangle_s$. A biexponential fit of these functions yields decay times of 70 fs (75 %) and 1.3 ps (25 %), thus conforming the existence of a significant sub-100 fs component, which causes the motional narrowing of the spectra.

The validity of the Franck-Condon approximation is studied in Fig. 3.3, which shows the time evolution of the correlation function $\langle M_k(t) \rangle_s$ of the transition dipole moment. According to its definition in Eq. (3.9), this function contains the time-dependent autocorrelation function of the transition dipole moments as well as the overlap matrix elements between local-mode and normal-mode basis functions. The latter define the initial value $\langle M_k(0) \rangle_s$, which clearly depends on the conformational state ($s = \alpha_\mathrm{R}$, β, $\mathrm{P_{II}}$) and on the vibrational eigenstate ($k = 1, 2$). The time evolution of the local-mode dipole-moment autocorrelation functions $\langle \boldsymbol{\mu}_n(t)\boldsymbol{\mu}_m(0) \rangle$, on the other hand, has only a minor effect on $\langle M_k(t) \rangle_s$. Following a weak initial decay, the functions are almost constant on the time scale of interest. Combining $\langle M_k(t) \rangle_s$ and the cumulant approximation of the propaga-

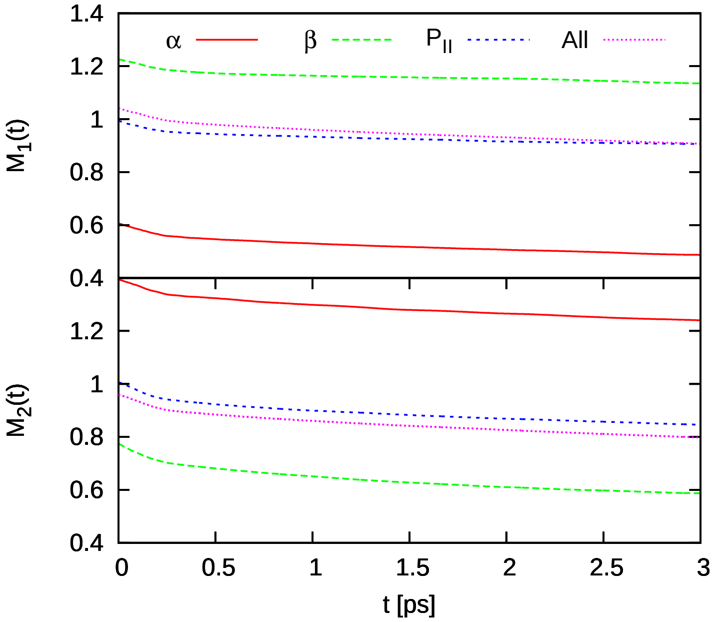

Figure 3.3: Correlation functions $\langle M_k(t) \rangle_s$ of the total transition dipole moment, shown for both amide I normal modes ($k = 1, 2$) and the conformations $s = \alpha_{\rm R}$, β, and $\rm P_{II}$.

tor, Fig. 3.4 shows the resulting eigenstate-averaged absorption bands $\sigma_s(\omega) = \sum_k \sigma_{sk}(\omega)$ for the three conformational states $s = \alpha_{\rm R}$, β, and $\rm P_{II}$. Compared to Fig. 3.1(c) where $\langle M_k(t) \rangle_s = 1$, it is found that the function $\langle M_k(t) \rangle_s$ essentially causes a relative weighting of the eigenstate peaks (e.g., an enhancement of the low-frequency peak of $\sigma_\beta(\omega)$). Hence, although the values of the transition dipole moments are roughly constant in local-mode representation (i.e., the Franck-Condon approximation holds there), this not the case in the normal-mode representation.

To go beyond the cumulant approximation, the IR absorption spectrum was also calculated via Eq. (3.5), i.e., directly from semiclassical line shape theory, and via Eq. (3.8), i.e., by invoking the adiabatic approximation only. Figure 3.4 compares the resulting absorption bands $\sigma_s(\omega)$ obtained for the three levels of line shape theory. The spectra calculated with and without the cumulant approximation are found to be quite similar. This finding reflects the fact that the underlying frequency distributions $\rho_{sk}(\omega)$ shown in Fig. 3.1 resemble Gaussian functions. Going beyond the adiabatic approximation, on the other hand, is found to have a significant effect on the overall line shape of the spectra. In all three cases the overall line width of the absorption band becomes smaller. In the case of the $\alpha_{\rm R}$ and

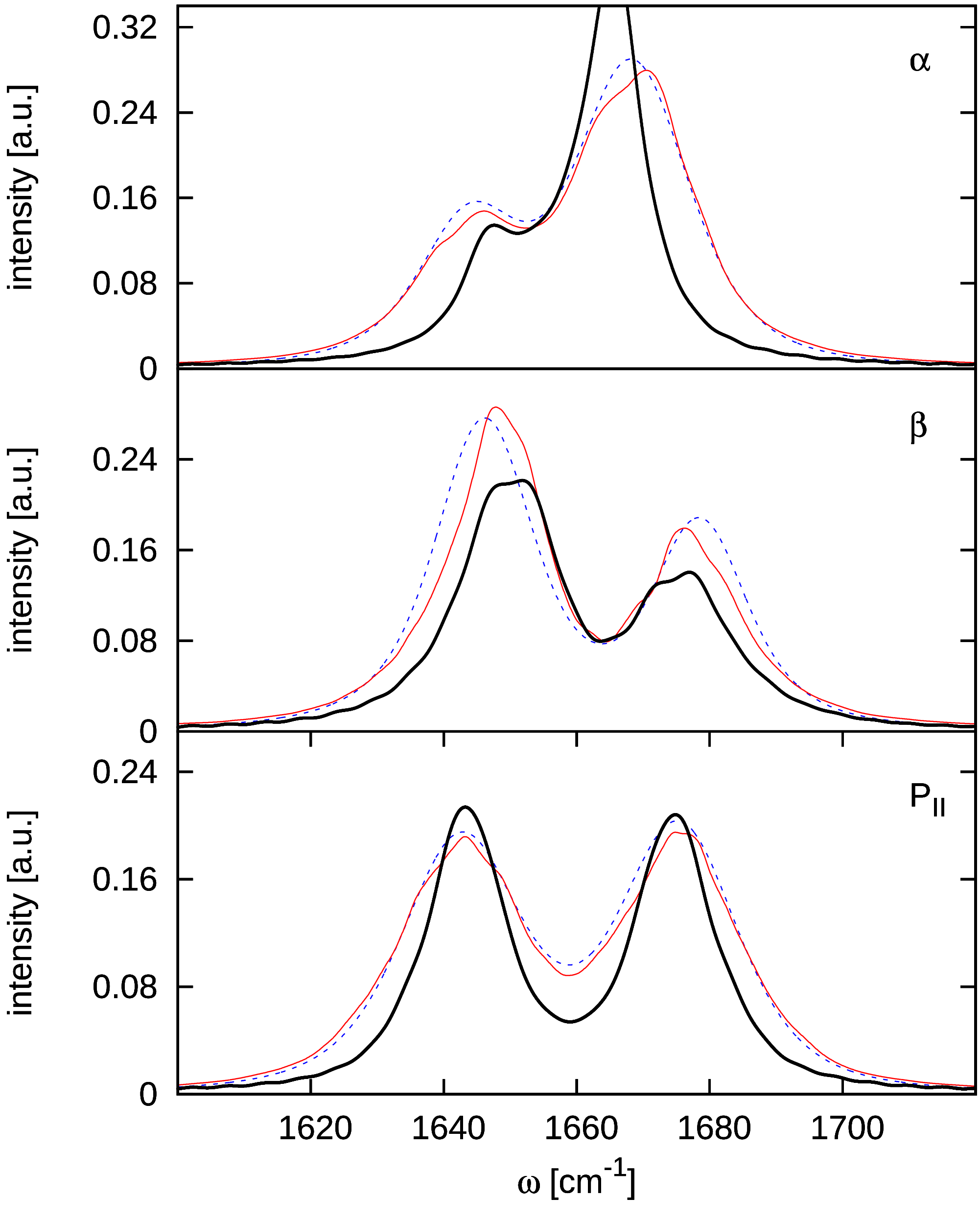

Figure 3.4: Amide I absorption bands $\sigma_s(\omega)$ of trialanine obtained for the conformations $s = \alpha_R$, β, and P$_{II}$. Compared are results calculated directly from semiclassical line shape theory (via Eq. (3.5), thick black lines), by invoking only the adiabatic approximation (via Eq. (3.8, thin red lines), and by invoking adiabatic and cumulant approximations (via Eq. (3.8), blue dashed lines).

the β spectrum, moreover, the position of the peaks is shifted. The largest change is found for the absorption band of the α_{R} conformation, whose low-frequent peak is suppressed almost completely. The findings are in line with an approximate treatment of the nonadiabatic vibrational dynamics via a stochastic Liouville equation. [20] Significant changes of spectral features due to nonadiabatic couplings are well also known from vibronic coupling theory, see, e.g., Ref. [82].

While the direct and adiabatic results are quite similar for the $\mathrm{P_{II}}$ state, they differ considerably for the α_{R} and the β conformations. To explain this finding, we reconsider Fig. 3.2 which shows the frequency splitting $\Delta\omega$ of the two amide I normal modes. The adiabatic approximation is expected to break down when the two amide I modes become (nearly) degenerate, thus giving rise to nonadiabatic curve crossings. Indeed, we find that a considerable part of the α_{R} and the β trajectories samples frequency splittings close to zero, while this is less the case in the $\mathrm{P_{II}}$ conformation.

We are finally in a position to compare our calculations to the experimental results of Ref. [1]. Apart from the two amide I peaks of interest, the experimental spectrum of trialanine exhibits a blue-shifted third peak, which corresponds to the terminal CO group of the molecule. To facilitate the interpretation, we have fitted the experimental spectrum by a sum of three peaks, each represented by sum of a Gaussian and a Lorentzian function. By subtracting the blue-shifted peak, we then obtain an estimate of the absorption band of only the two amide I modes. Figure 3.5 compares the resulting experimental spectrum to our simulated spectrum obtained from the direct calculation of the absorption bands $\sigma_s(\omega)$, weighted by the correct thermal populations of the conformational states $\mathrm{P_{II}}$ ($\approx$ 90 %) and β ($\approx$ 10 %). [54] Also shown is a calculation which uses the adiabatic and the cumulant approximations which, by chance, appears to agree better with experiment. Apparently, the approximate calculation benefits from a fortunate cancellation of the errors stemming from the adiabatic approximation and (most likely) of the solvent-induced frequency calculations.

To explain the above described break-down of the adiabatic approximation, it is helpful to recall the Born-Oppenheimer approach to solve the molecular Schrödinger equation comprising electronic (r) and nuclear (R) coordinates. Exploiting the time scale separation between the electronic and nuclear motions, one first solves the electronic Schrödinger equation for fixed nuclei, thus obtaining the adiabatic potential-energy surfaces $W_n(R)$ (the eigenvalues) and the adiabatic electronic wave functions $\psi_n(r;R)$ (the eigenstates) of the problem. In a second step, the total wave function is expanded in the $\{\psi_n\}$ basis

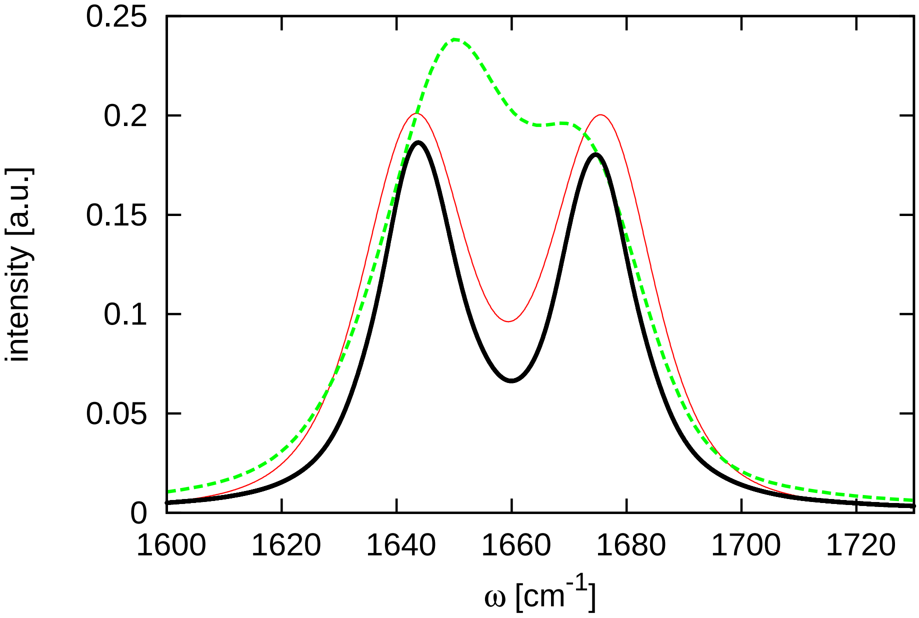

Figure 3.5: Comparison of experimental (Ref. [1], green dashed line) and calculated amide I absorption spectra of trialanine. The latter were obtained directly from semiclassical line shape theory (via Eq. (3.5), thick black line) and by invoking adiabatic and cumulant approximations (via Eq. (3.8), thin red line).

and inserted in the molecular Schrödinger equation, which yields coupled equations for the vibrational wave functions of the system. Assuming that the adiabatic electronic states are well separated in energy, these couplings can be neglected (the Born-Oppenheimer approximation), and the vibrational dynamics of the system can be considered in a single electronic state.

Essentially the same strategy has been employed above to calculate vibrational line shapes of a flexible molecule in solution (and is commonplace in various areas of chemical physics, see e.g., Refs. [83–85]). Assuming that the period of the amide I mode is short compared to the motion of the solvent and the conformational rearrangement of the peptide, we have calculated the vibrational frequencies of the molecule for fixed solvent coordinates s and fixed dihedral angles (ϕ, ψ). In harmonic approximation, this is done through a geometry optimization at fixed $x = \{s, \phi, \psi\}$, followed by a normal mode calculation. The resulting normal-mode states $|\psi_k(x)\rangle$ with vibrational frequencies $\omega_k(x)$ correspond to excited vibrational states of the molecule, which (after a transformation to a local-mode basis) build up the vibrational Hamiltonian (3.1) used above. Similar as in the vibronic problem, the resulting Schrödinger equation represents a nonadiabatically coupled problem. Unlike to the former case, however, the time scale separation between the period of the amide I mode ($\approx$ 20 fs) and the solvent and conformational degrees of freedom is only weakly satisfied, since the latter also show a subpicosecond component in their dynamics. As a consequence, it should not come as a surprise that the "vibrational Born-Oppenheimer" or adiabatic approximation (that is, the neglect of the vibrational nonadiabatic coupling) is in general not applicable to the calculation of the amide I spectrum of peptides. As a qualitative indication of the validity of the adiabatic approximation, one may consider the distribution of the frequency splittings of the normal modes, since nonadiabatic curve crossings occur mostly for (nearly) degenerated vibrational modes.

In this section we have introduced a sequence of approximations that is usually employed to calculate condensed-phase IR spectra. That is, (i) the semiclassical line shape theory [Eq. (3.5)], (ii) the adiabatic approximation [Eq. (3.8)], the cumulant approximation [Eq. (3.11)], and the inhomogeneous limit [Eq. (3.14)]. Sequence means that the order of the approximation is important, e.g., it is difficult to perform the cumulant expansion without previously making the adiabatic approximation. At any point of the theory one may furthermore employ the Franck-Condon approximation [Eq. (3.13)].

As starting point, we have chosen semiclassical line shape theory, [43] which assumes that the fluctuations of the peptide and the surrounding solvent molecules result in a classical time-dependence of the vibrational Hamiltonian. While the formulation describes well the dephasing of the vibrational transitions, it is not capable to account for the finite life time (1 ps) of the excited amide I states, which therefore had to be included in a phenomenological manner. Reduced density matrix formulations that include MD-based spectral densities appear to be a promising approach to go beyond the semiclassical level of theory. [44, 86]

Proceeding from simple to more exact formulation, it has been shown in Fig. 3.1 that the sub-100 fs component of the frequency autocorrelation functions $\langle \delta\omega_k(t)\delta\omega_k(0)\rangle_s$ affects a significant (a factor of two) motional narrowing of the spectra. The effect was found to be well described in cumulant approximation, because the distributions of the frequencies shown in Fig. 3.1 are well approximated by Gaussians. We note that the latter assumption may deteriorate, [46, 47] e.g., if the classical average is performed over several conformational states instead of separately for each state as in Eq. (3.14). The Franck-Condon approximation was found to hold well in the local-mode representation but not so in the normal-mode representation. This is reassuring, considering the dramatic breakdown of the approximation found in the description of the OH stretch spectrum of water. [45] However, as explained above, the adiabatic approximation cannot be expected valid in the case of closely lying amide I normal modes and may therefore lead to significant deviations.

3.1.5 Summary

As a first attempt to establish a model system for the simulation of IR spectra of flexible peptides in aqueous solution, we have outlined a quantum-classical description of the amide I vibrational spectrum of trialanine cation in D_2O. Heading towards a quantitative theoretical prediction of time- and frequency-resolved IR spectra of biomolecules, the following main challenges are to be met:

- Sampling of the true conformational distribution of the system by a state-of-the-art MD simulation. If several conformational states coexist, their thermal population probability need to be determined, usually via comparison to experiment.
- Modeling of the solvent-induced frequency fluctuations by exhaustive quantum-chemical calculations. In particular, the coupling of solvent and conformational motions needs to be accounted for.
- Description of nonadiabatic transitions between vibrational eigenstates. Although the direct propagation of the time-dependent Schrödinger equation in principle is a straightforward matter, it requires the averaging over rapidly oscillating functions, which results in an considerably larger sampling effort as it is case for the standard cumulant approximation.

Considering the achieved agreement between simulated and experimental IR absorption spectra, we clearly did not yet succeed in a quantitative first principle prediction. Furthermore, it is clear that only the simulation of the much more detailed multidimensional IR spectra will truly asses the quality of the theoretical model. However, we think that —for the system considered— two out of the three challenges have been met. That is, the appropriate description of the conformational distribution and the correct calculation of the dynamic absorption spectrum. What is missing is an accurate enough quantum-chemical modeling of the vibrational frequencies of a solvated peptide — a topic that represents a quite active field of research. Nevertheless, our study underlines that the amide I response of peptides depends on a number of aspects. While time-resolved IR spectroscopy hold the promise to resolve conformational dynamics in real time, [87] it certainly requires substantial theoretical support.

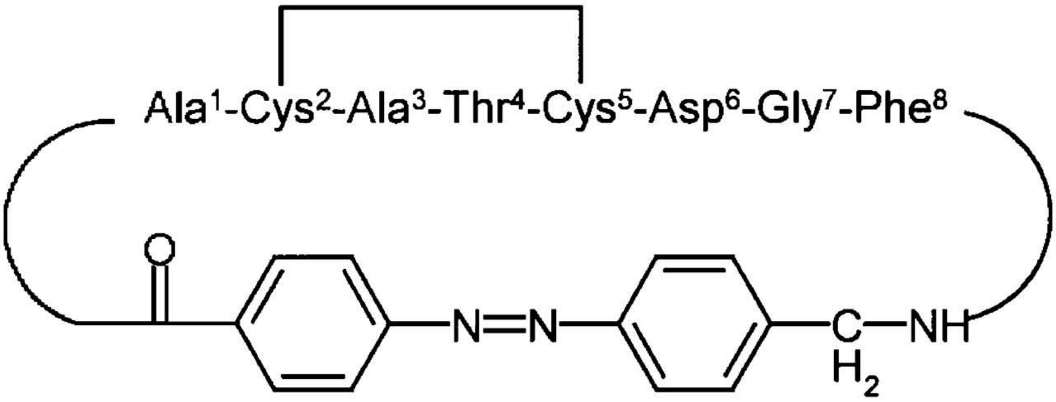

Figure 3.6: Structure and amino acid labeling of the bicyclic azobenzene peptide bcAMPB.

3.2 Photoswitchable Bicyclic Azobenzene Octapeptide

In this section we apply the above constructed vibrational Hamiltonian of solvated polypeptides to explain the experimentally observed time dependency of vibrational spectra of photoswitchable bicyclic azobenzene octapeptide. The considered system is octapeptide fragment H-Ala-Cys-Ala-Thr-Cys-Asp-Gly-Phe-OH connected head to tail via (4-aminomethyl)-phenylazobenzoic acid as well as a disulfide bridge, see Fig. 3.6. A small blue-shift of the vibrational spectra happening during transition from *cis* to *trans* conformation is observed experimentally [58]. The considered system is interesting from different aspects. First, its description gives as an example of a practical implementation of the building block model. Second, the transition from the *cis* to *trans* conformation is a non-equilibrium process, so we can can develop and test a strategy for its description. Third, the mentioned changes of the vibrational properties are rather small and it is a challenge to reproduce them within our vibrational model. And finally, we can demonstrate a concrete example of how the time-dependent changes of the vibrational spectra can be treated in terms of the conformational changes.

3.2.1 Methods

The vibrational Hamiltonian of the polypeptide was constructed with the usage of the building block model based on the *ab initio* maps of glycine dipeptide. The non-first neighbor effects, usually treated within electrostatic models, were neglected. This simplifies the relation between the vibrational properties and conformational structure of the polypeptide. Since the non-first neighbor effects play only a minor role and we are aimed to a qualitative description of the effect, the simplification of the Hamiltonian seams to be reasonable.

Because of the same reason we did not include the solvent contribution into account. As we have shown in the consideration of A_3^+, the solvent contribution significantly changes the shape of the total spectrum (with respect to its gas phase shape). However, in the considered case, the number of the contributions to the total spectra is larger (9 instead of 2) and, as a consequence, the shape of the total spectra is mostly conditioned by the position and intensity of these contributions (and not by their form). Therefore, the solvent correction, changing shape of the contributions to the total spectra, should influence on the shape of the total spectra not so significantly like in the case of A_3^+. Moreover, since the considered molecule is not solvable in watter we have used DMSO as solution. In this case the influence of solvent is expected to be smaller. First, the overlap of the amide I normal mode frequencies of the peptide with the normal mode frequencies of DMSO is smaller and, as a consequence, the interaction of amide I modes with the solvent vibrations is expected to be smaller. Moreover, the number of hydrogen bonding in the case of DMSO is smaller than in the case of watter. The latter also makes the influence of solvent smaller. Moreover, the solvent correction is an additive quantity. This allows for an independent consideration of the *ab initio* and solvent contribution to the considered effect. However, we have to emphasize that the solvent effects are expected to be much more important than the not next neighbor influence on the site energies and they have to by included in a model dedicated to a qualitative treatment of the system.

Like in the case of A_3^+, we start from a MD simulation to get the dihedral angles ϕ_j and ψ_j as functions of time. The angles are then substituted into the vibrational Hamiltonian of the system which explicitly depends on the dihedral angles. The obtained in such way time-dependent Hamiltonian is diagonalized, which gives the normal mode frequencies ω_j as functions of time. In turn, the normal mode frequencies $\omega_j(t)$ are used as the input for the spectra calculation.

About 300 MD trajectories of the considered system have been generated. Each trajectory is 1050 *ps* long. The first 50 *ps*, the MD simulations were performed in microcanonical ensemble. To keep the energy constant the time step was set to be equal to 0.2 *ps*. After 50 *ps*, the MD simulations were performed in canonical ensemble with the time step of 2 *ps*. For more technical details of the MD simulations see Ref. [88].

3.2.2 frequencies Distributions

We start our consideration from the distributions of the vibrational frequencies, which can be considered as a rough approximation to the vibrational spectra. In order to calculate the time dependent frequency distribution we count for the number of vibrational modes which in a fixed time interval have frequencies from a given frequency range. In this section we discuss selection of appropriate steps for frequency and time. On one hand, the frequency step has to be sufficiently large to get enough point for the averaging and, as a consequence, to get rid of fluctuations and obtain smooth dependencies. On the other hand, the step has to be sufficiently small to save small features in the dependency on frequency. At fixed time we have about 300*9 values of the normal mode frequencies, whre 300 comes from the number of the MD trajectories and 9 is the number of the normal modes. This number of points is not sufficient to get a reasonable statistical averaging. The possible solution is an additional averaging over the time. However, we consider a nonequilibrium process and, as a consequence, points corresponding to different times are not equivalent. It means that the time step should be sufficiently small to neglect systematic time dependency within the selected interval. Keeping this in mind we divided the whole frequency range (from 1704 cm^{-1} to 1837 cm^{-1}) into 150 segments. That corresponds to the frequency step of about 0.89 cm^{-1}. The time step was taken to be equal to 30 ps (the whole range was from 0 to 1050 ps). In such way we got 35 points on the time scale. The above described way to calculate the frequency distribution can be summarized by the following mathematical expression

$$\rho\left(\omega,t\right) = \lim_{\substack{\Delta\omega\to 0\\ \Delta\tau\to 0\\ N\to\infty}} \frac{1}{\Delta\omega\Delta\tau N}\sum_{n=1}^{N}\sum_{k=1}^{9}\int_{\omega-\frac{\Delta\omega}{2}}^{\omega+\frac{\Delta\omega}{2}} d\omega^{'} \int_{t-\frac{\Delta t}{2}}^{t+\frac{\Delta t}{2}} dt^{'}\delta\left(\omega^{'}-\omega_{k}^{(n)}\left(t^{'}\right)\right), \qquad (3.18)$$

where $\Delta\omega$ and Δt are the frequency and time steps, respectively. Indexes k and n are used to numerate normal modes and MD trajectories, respectively and $\omega_k^{(n)}\left(t\right)$ are normal mode frequencies.

In the Fig. 3.7 we have plotted the frequency distributions calculated for the beginning and end of the time evolution of the system (the first and last time intervals). As we can see in the figure, the above selected time and frequency steps give rather smooth frequency distributions. At the same time the small features remain. Another important observation is that we have a clear and qualitatively correct dependency of the frequency distributions on time.

Let us fix frequency and consider the frequency distribution as function of time. In the

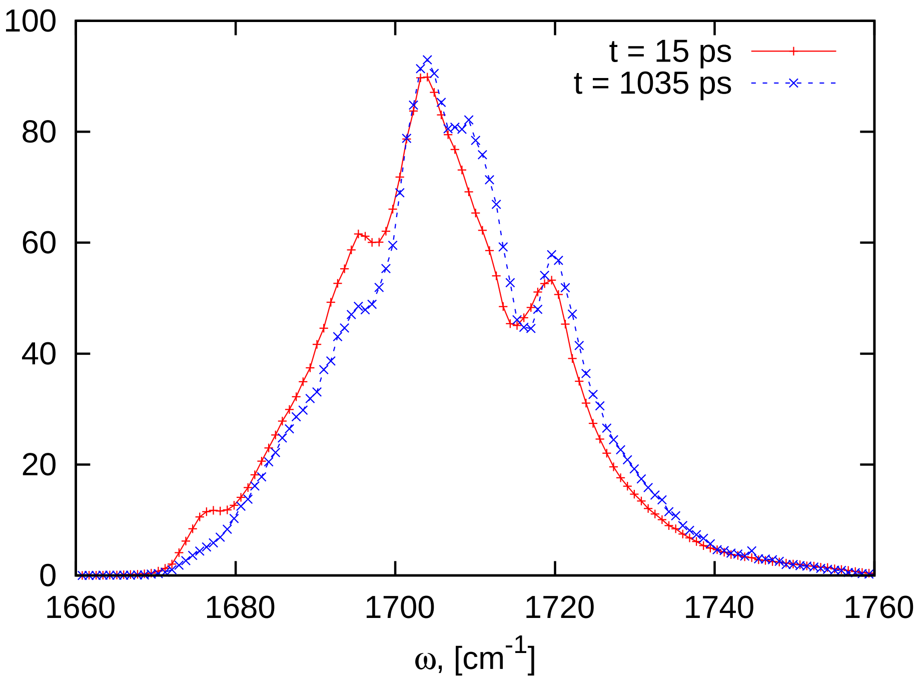

Figure 3.7: Frequency distributions corresponding to the beginning and end of the time evolution of the system.

Fig. 3.8 we have plotted such dependency for the frequency equal to about 1695.26 cm^{-1} as well as for the two neighboring frequencies. As we can see in the above figure, the time dependency is much weaker than dependence on frequency and, as a consequence, the fluctuations become more significant. However, the selected time and frequency steps allow us to observe the clear systematic dependency on time.

In order to get rid of the fluctuations and describe the time dependency in a quantitative way, it has been fitted by cubic polynomial for all frequencies on the considered frequency grid. In other words the frequency distribution has been fitted by the following function.

$$\rho\left(\omega,t\right)=c_0\left(\omega\right)+c_1\left(\omega\right)t+c_2\left(\omega\right)t^2+c_3\left(\omega\right)t^3. \tag{3.19}$$

The examples of the fitting functions are also shown in the Fig. 3.8.

In the Fig. 3.9 we have plotted the obtained fitting coefficient c_j as functions of frequency. In the figure we see the obvious correlation between the parameters of the fitting. This effect is shown in other representation in the Fig. 3.10, where c_2 and c_3 are plotted as functions of c_1. The shown dependencies can be accurately fitted by the linear function of the form $f\left(x\right)=k\ x$. The fitting functions are shown on the same figure. Since the fitting parameters are strongly related with each other, we can conclude that time dependence

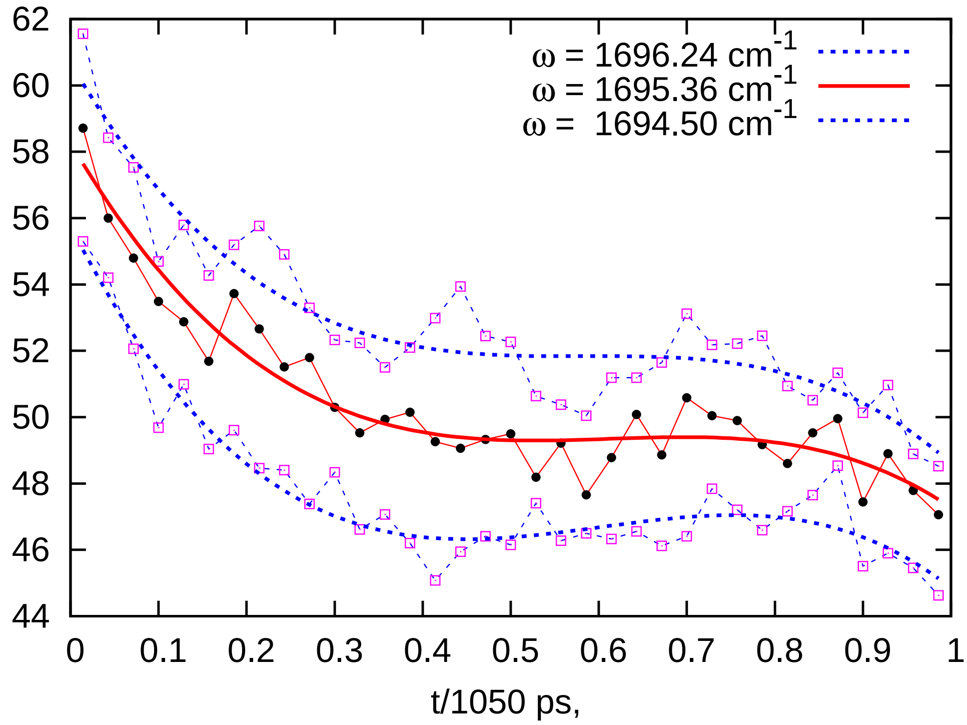

Figure 3.8: Dependency of the frequency distribution on time.

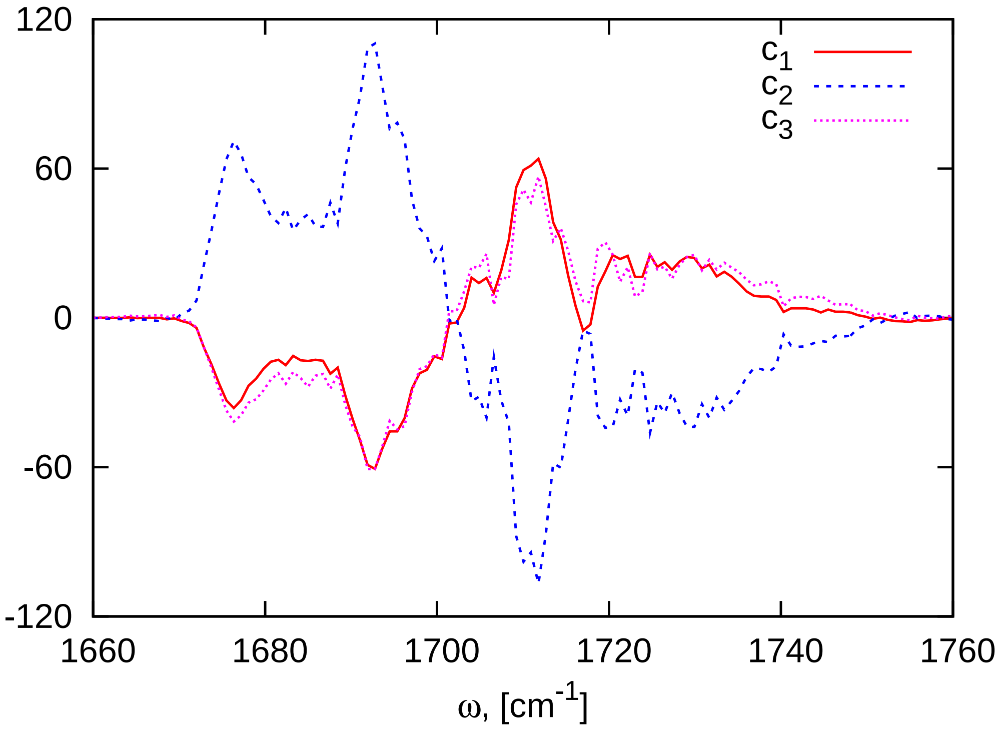

Figure 3.9: Dependencies of the fitting coefficients on frequency.

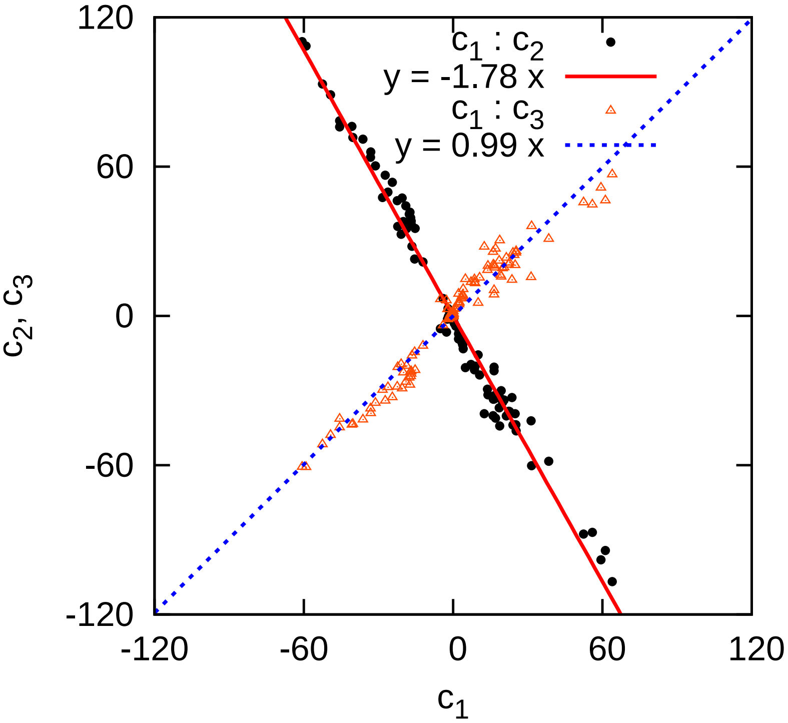

Figure 3.10: Correlation between coefficients of the fitting.

has lees degree of freedom than the fitting function do and is, to large extent, determined by initial value of the frequency distribution and one additional parameter. Moreover, in spite on the fact that for different frequencies the time dependencies of the distribution is qualitatively different, we can say that it is described by the same function and, as a consequence, has the same physical origin.

In the Fig. 3.11 the frequency distribution is shown as function of two variables (frequency and time). In order to see the time dependency on the background of much stronger frequency dependency we have extracted the initial frequency distribution

$$\rho_{pl}\left(\omega,t\right)=\rho\left(\omega,t\right)-\rho\left(\omega,0\right), \tag{3.20}$$

where ρ_{pl} is the shown in the figure function and $\rho\left(\omega,t\right)$ is the considered frequency distribution. As we can see in the figure, the main changes occur at first 50 ps and then the distribution remains qualitatively unchanged. In the Fig. 3.12 we show the same function obtained with the above considered cubic fitting polynomial. As we can see the fitting reproduces the general behavior quit accurately.

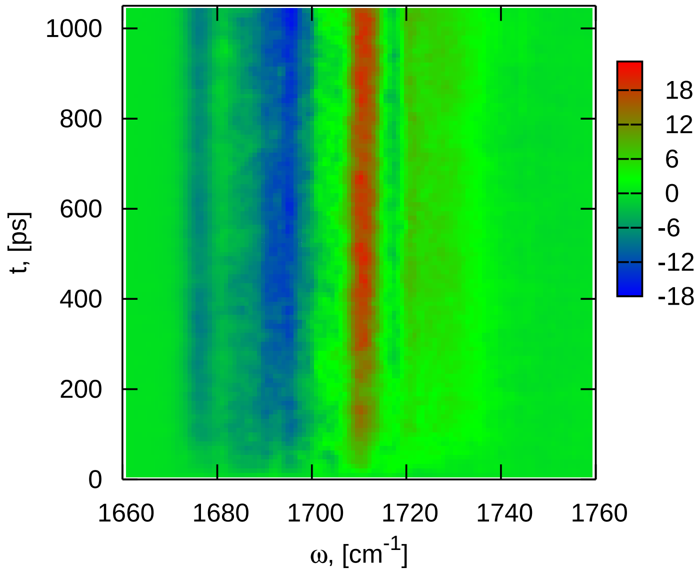

Figure 3.11: Time dependent changes of the frequency distribution.

3.2.3 Frequencies Correlation Functions

The above considered frequency distribution is only a rough approximation to the vibrational spectra. More realistic representation can be obtained within the above considered cumulant expansion. The second order cumulant expansion has been chosen since it significantly improves convergence. The latter is critical because of the lack of points for the statistical averaging [since the process is non-equilibrium, the averaging over time is restricted by small time intervals]. Moreover, the shape of the total spectra is only slightly depends on the shape of the contributions and mostly conditioned by their width and positions, which can be rather accurately reproduced within the cumulant approximation. Another factor, conditioning the shape of the total spectra, is intensity of the contributions, which is not affected by the cumulant expansion.

The above consideration of the distributions has shown that the system shows a fast processes in the beginning of the time evolution. To get a more detailed description at the short-times we decreased the time step from 30 to 10 *ps*. For each time interval and each normal mode the corresponding correlation function is calculated. In more details, in each time interval the average frequency has been calculated and extracted from the initial

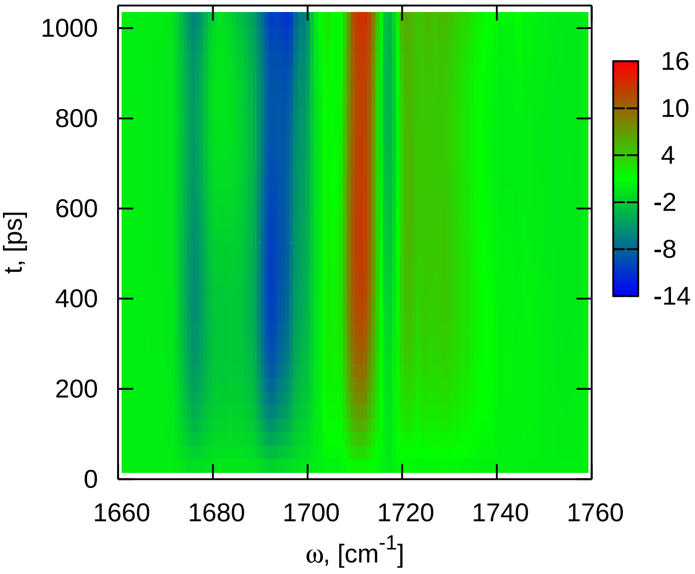

Figure 3.12: Time dependent changes of the frequency distribution. The time dependency is calculated with the usage of the fitting functions.

frequency

$$\delta\omega_j\left(t,\tau\right)=\omega_j\left(t,\tau\right)-\overline{\omega}_j\left(\tau\right), \tag{3.21}$$

where index j indicates normal mode number, τ is time corresponding to the middle of the segment, over which the averaging has been performed, and t is time parameterizing points within the segment. The obtained in such way frequencies $\delta\omega\left(t,\tau\right)$ were used in the calculation of the correlation functions

$$C_j\left(\Delta t,\tau\right)=\frac{1}{\Delta\tau}\int_{\tau-\frac{\Delta\tau}{2}}^{\tau+\frac{\Delta\tau}{2}}\langle\delta\omega_j\left(t+\Delta t,\tau\right)\delta\omega_j\left(t,\tau\right)\rangle dt. \tag{3.22}$$

As it is indicated in the above expression, for different time intervals (indicated by τ) we perform the averaging over time t within the interval (which corresponds to the integration). As a result, for different time segments, we obtain different correlation functions. The latter is explicitly indicated by the τ in the list of argument of the correlation function. Additionally to this averaging we perform also average over different MD trajectories. The average frequencies and correlation functions give position and shape of the peaks, respectively.

$$\sigma(\omega,\tau)=\sum_{k=1}^{9}\int_{-\infty}^{\infty}dte^{-i\omega t}\int_{0}^{t}dt^{'}\left(t-t^{'}\right)C_j(t^{'},\tau). \tag{3.23}$$

Let us first focus on the dependency of the peaks position $\overline{\omega}_j\left(\tau\right)$ on time τ. The averaged normal mode frequencies as functions of time are shown in the Fig. 3.13. On the upper panel we have plotted the averaged time dependent normal mode frequencies with subtracted initial value. This facilitates comparison of different normal modes as well as makes changes within a single normal mode more pronounced. The following features can be noticed in the considered figure. All normal mode frequencies quickly increase the first 200 ps. The 1, 2, 3, and 4 modes as well as 6 and 7 ones show very similar time evolution. Moreover, we can even see correlation between the normal modes frequencies from different groups mention before. The higher is number of the normal mode the smaller is difference between the initial and final value. And the last observation is that the distance between the normal mode frequencies is larger for the edge normal modes and smaller for the central ones.

Among all mentioned properties of the time dependent normal mode frequencies we will focus on the systematic increase of their frequencies. This effect is one of the contributions to the blue shift of the vibrational spectra which is a sum of peaks corresponding to all normal modes. Another possible origin of the total blue shift is time dependent changes of

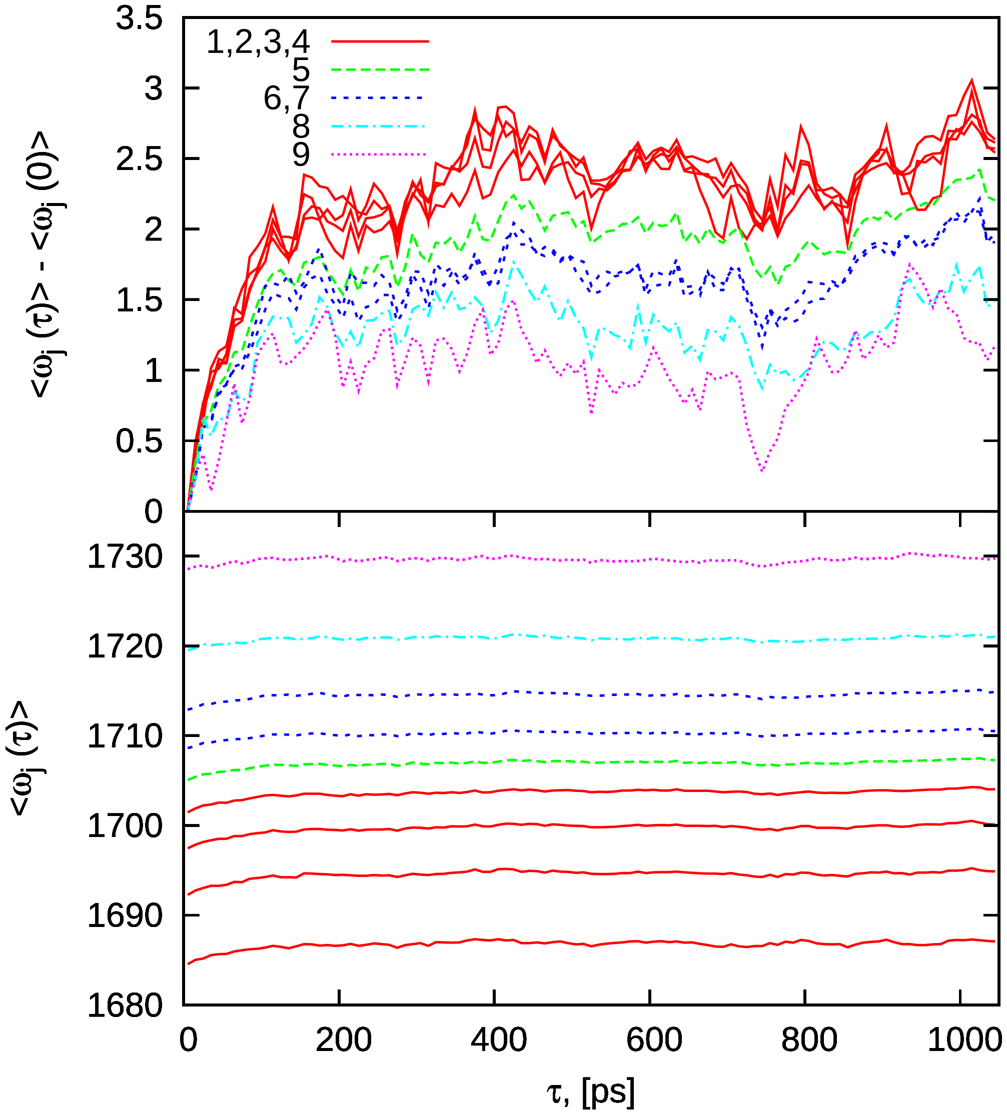

Figure 3.13: Averaged [over time-segments and MD trajectories] normal mode frequencies as functions of time.

the peaks intensities. This effect, conditioned by the time dependency of the transitional dipole moments, is not considered in the present work. The found blue shift of the normal modes frequencies has a universal character (it is observed for all normal modes and has similar form). In order to simplify interpretation of this effect in terms of the conformational changes we will consider the sum of all normal modes frequencies. In this case we can make use of the fact that trace of any matrix is invariant with respect to arbitrary unitary transformation. In the considered case this property means that sum of all normal mode frequencies is equal to the sum of all local modes frequencies (site energies). In this way we can replace the consideration of the normal mode frequencies by the consideration of the local frequencies. This facilitate the considered problem since the site energies depend in the simple way on the local conformational structure.

In the Fig. 3.14 we show the site energies as functions of time. As before, in order to facilitate the comparison of the site energies we have also shown time dependencies from which the initial value is subtracted. In the figure we can see that first three site energies give the main contribution to the blue shift.

Let us relate the observed behavior of the three site energies with the conformational changes. According to the used building block model, the considered site energies depend on the dihedral anglers in the following way.

$$\begin{aligned}
&\varepsilon_1\left(\phi_1,\psi_1\right)=\varepsilon_N^{GD}\left(\phi_1,\psi_1\right), \\
&\varepsilon_2\left(\phi_1,\psi_1,\phi_2,\psi_2\right)=\varepsilon_C^{GD}\left(\phi_1,\psi_1\right)+\varepsilon_N^{GD}\left(\phi_2,\psi_2\right)-\varepsilon^{NMA}, \\
&\varepsilon_3\left(\phi_2,\psi_2.\phi_3,\psi_3\right)=\varepsilon_C^{GD}\left(\phi_2,\psi_2\right)+\varepsilon_N^{GD}\left(\phi_3,\psi_3\right)-\varepsilon^{NMA},
\end{aligned} \tag{3.24}$$

where ε_N^{GD} and ε_C^{GD} are site energies of the GD molecule on the N and C sites, respectively. ε^{NMA} is amide I normal mode frequency of the NMA molecule. In the Fig. 3.15 we have shown the time dependency of the five contributions to the considered site energies. We can see that the changes of the first and second pare of the dihedral angles is the main reason of the blue shift of the three site energies. In order to relate the changes of the ε_N^{NMA} and ε_C^{NMA} with the time dependency of the dihedral angles ϕ_j and ψ_j we have combined the time dependent distributions over dihedral angles with the maps of the site energies (see Figs.3.16 and 3.17). In the given figures, we can see as peaks of the distributions shift to the region where the site energies have large values.

Above we have considered the shifts of peaks corresponding to 9 normal modes and making contribution to the total spectra. Let us now consider the shape of the peaks, which is given by the shape of corresponding frequency autocorrelation function (3.22).

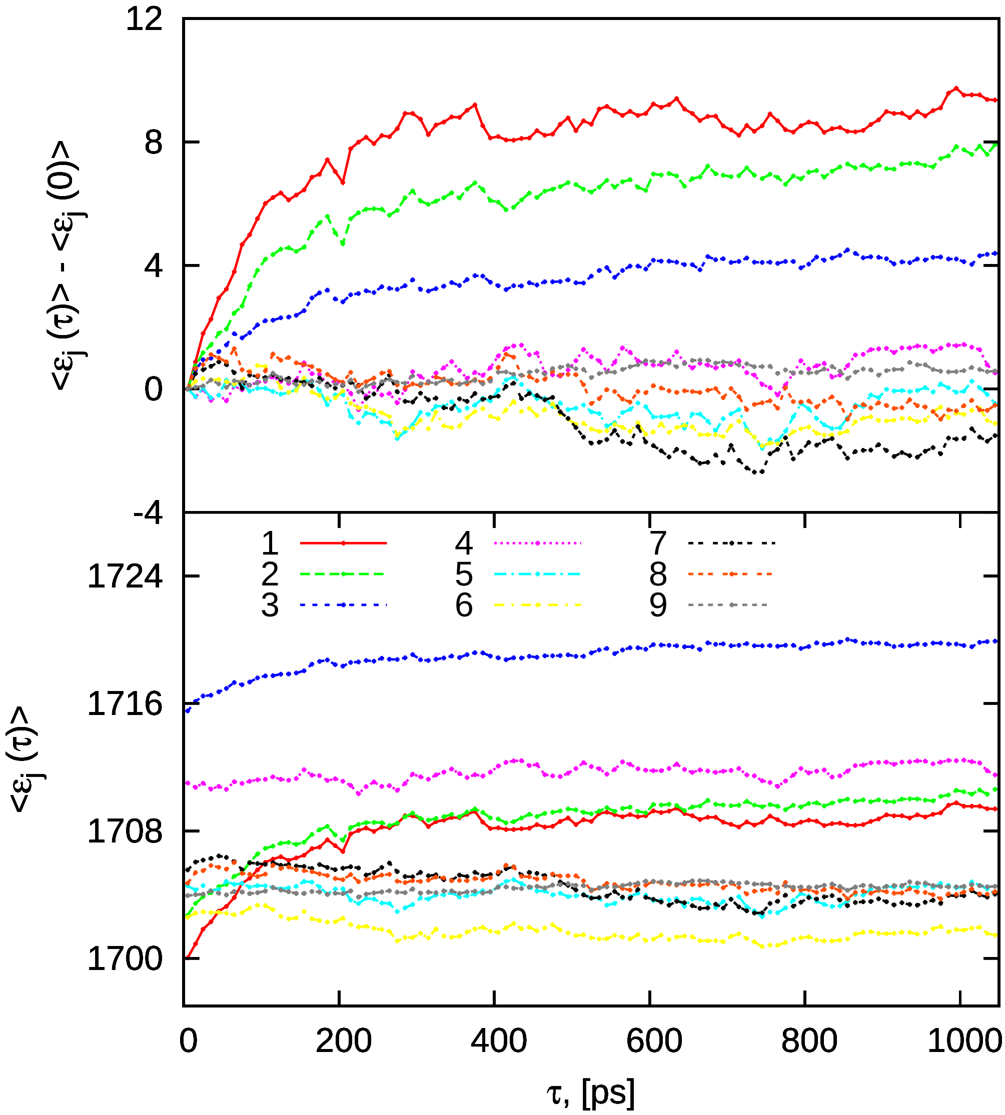

Figure 3.14: Averaged [over time-segments and MD trajectories] site energies as functions of time.

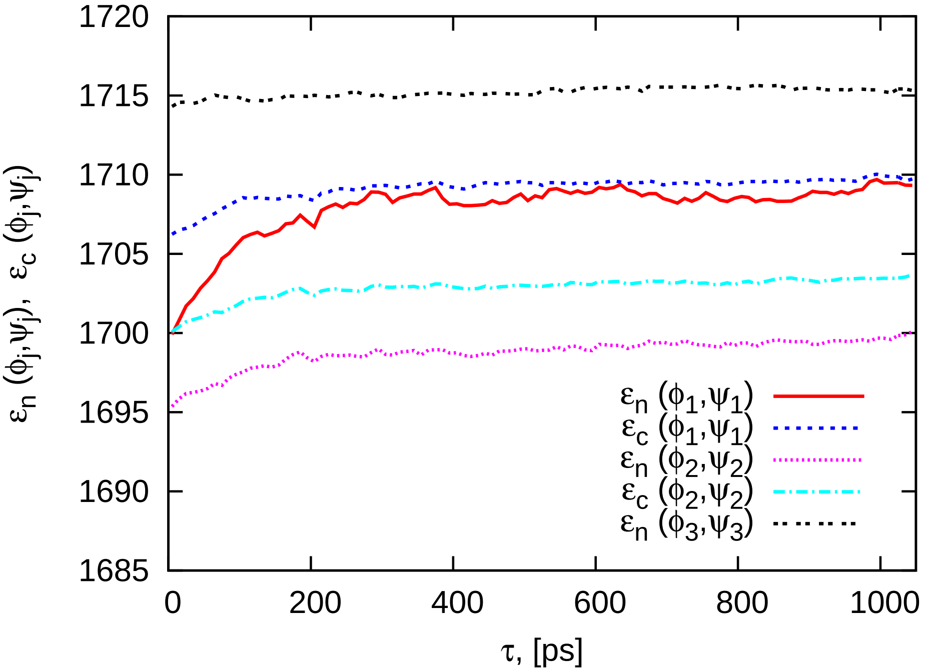

Figure 3.15: Contributions to the first tree site energies as functions of time.

In order to get rid of the fluctuations present in the correlation functions (conditioned by insufficient convergence) as well as to quantitatively describe shape of the correlation functions we have performed their fitting by the biexponential function of the following form:

$$f_j\left(t,\tau\right) = \delta\omega_j^2\left(\tau\right)\left[\nu_{1j}\left(\tau\right)e^{-\alpha_{1j}(\tau)t} + \left(1-\nu_{1j}\left(\tau\right)\right)e^{-\alpha_{2j}(\tau)t}\right], \qquad (3.25)$$

where the index j is used to indicate of the normal mode number. The parameters of the fitting are constant within a fixed time-segment (they do not depend on t) and depend on time τ which is used for the indication of the time-segments. It has to be mentioned that fitting parameters are time dependent in the sense that they are different for different time segments but still constant within fixed segment. As we have already mentioned, the length of the time segments, for which the correlation functions were calculated, was taken to be equal to 10 ps. The averaging has been performed over all 300 MD trajectories as well as over time within the segments. The calculated in such way correlation functions are defined on the time range from 0 to 10 ps. During the fitting we have used the correlation functions in the time range from 0 ps to 4 ps. Restricting the time range we have increased portion of the fast component of the correlation function and truncated the part (at large times) with worser convergence.

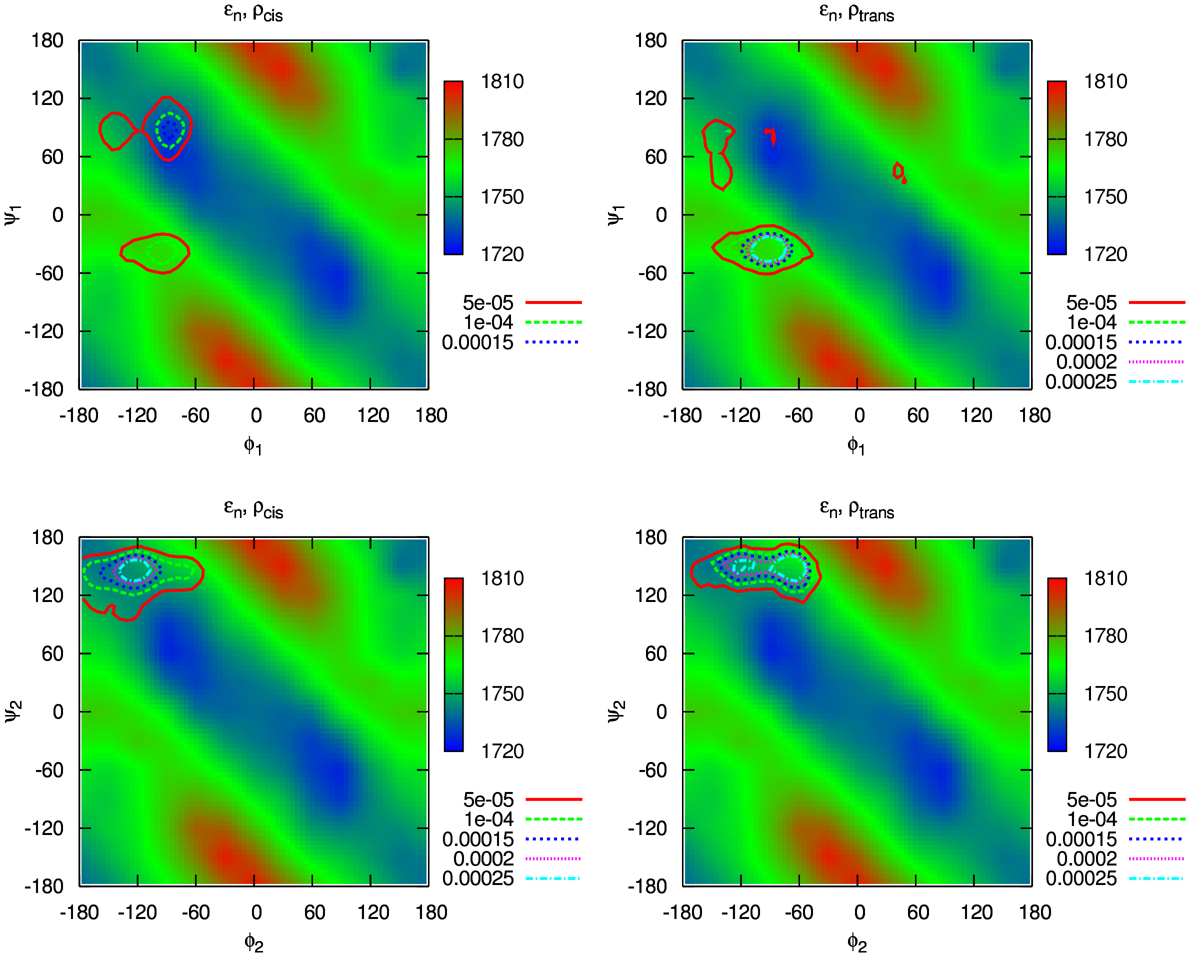

Figure 3.16: *Ab initio* maps of the ε_n combined with the distributions of the dihedral angles ϕ and ψ for the beginning and end of the time evolution.

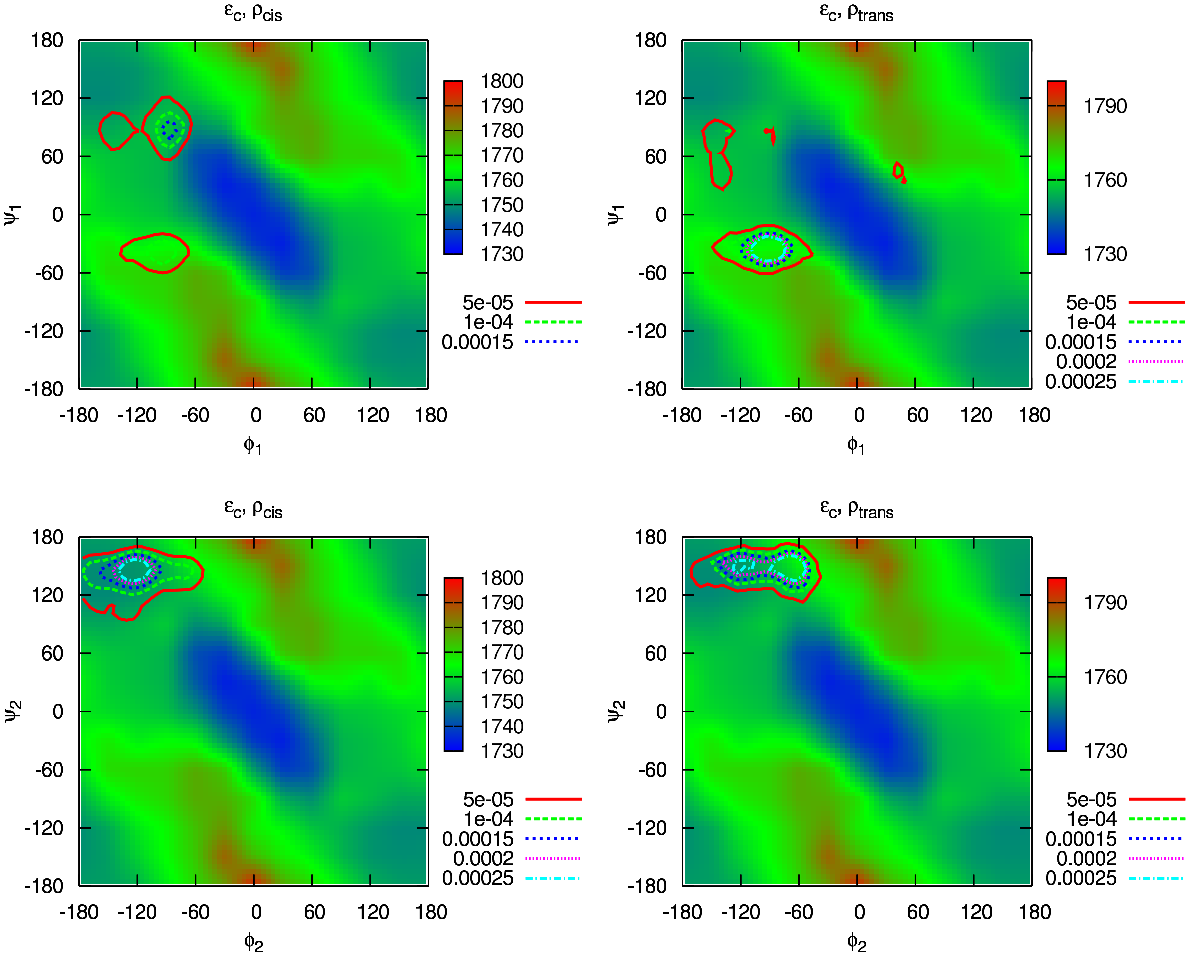

Figure 3.17: *Ab initio* maps of the ε_c combined with the distributions of the dihedral angles ϕ and ψ for the beginning and end of the time evolution.

First, we have studied dependence of the fitting parameters (which define shape of the corresponding vibrational spectra) on time. In the Fig. 3.18 the fitting parameters of the 8 normal modes are shown as functions of time. We can see that systematic time dependence either does not exist or essentially small in comparison with the amplitude of the fluctuations. Moreover, the possible systematic time dependencies of the fitting parameters are rather small in comparison with their absolute values. We have also studied how the fitting parameters of the correlation functions depend on the normal mode number. In the same figure we can see that dependence of the fitting parameters $\delta\omega_j$ on the normal mode number is much more significant than their dependence on time. We can also easily notice dependency of the parameter ν_j on normal mode number j. For the fitting parameters τ_{1j} and τ_{2j} there is no obvious dependency on the normal mode number.

However, dependence of τ_{1j} and τ_{2j} on the normal mode number can be found by calculation of the average over time of the considered parameters. The last ones (together with averaged over time $\delta\omega_j$ and ν_j) are shown in the Fig. 3.19 (red curve). The dependence of the averaged α_{1j} and α_{2j} on the normal mode number is rather regular what likely shows that it should not be treated as a coincidence. In order to extract a possible time dependency of the fitting parameters of the correlation functions we fitted them by linear over time function ($f\left(t\right) = c + kt$). The results of this fitting is also shown in the considered figure (blue curves). The right panel corresponds to the parameter c and the left one to the k. As we can see, the the parameter k is also rather regularly depends on the normal mode number. The last indicates that its values are conditioned by some regular effects (not by random fluctuations) and, as a consequence, we found an evidence of a slight time dependency of the coefficients of the correlation functions.

3.2.4 Spectra

In order to calculate the resulting spectra as well as difference between the initial and final spectra we assumed that shape of the contribution is constant in time. For better agreement with the experimental spectra we have used scaling factor for frequencies equal to 0.948. The spectra obtained in such way is shown on the Fig. 3.20. On the Fig. 3.21 we show time dependent changes of the vibrational spectra calculated as difference between the spectra corresponding to the considered time and the initial spectrum.

These results together with the earlier considered dependency of the average frequency on time and on normal mode number can be formulated in terms of the vibrational spectra in the following way. Shapes and positions of peaks are different for different normal

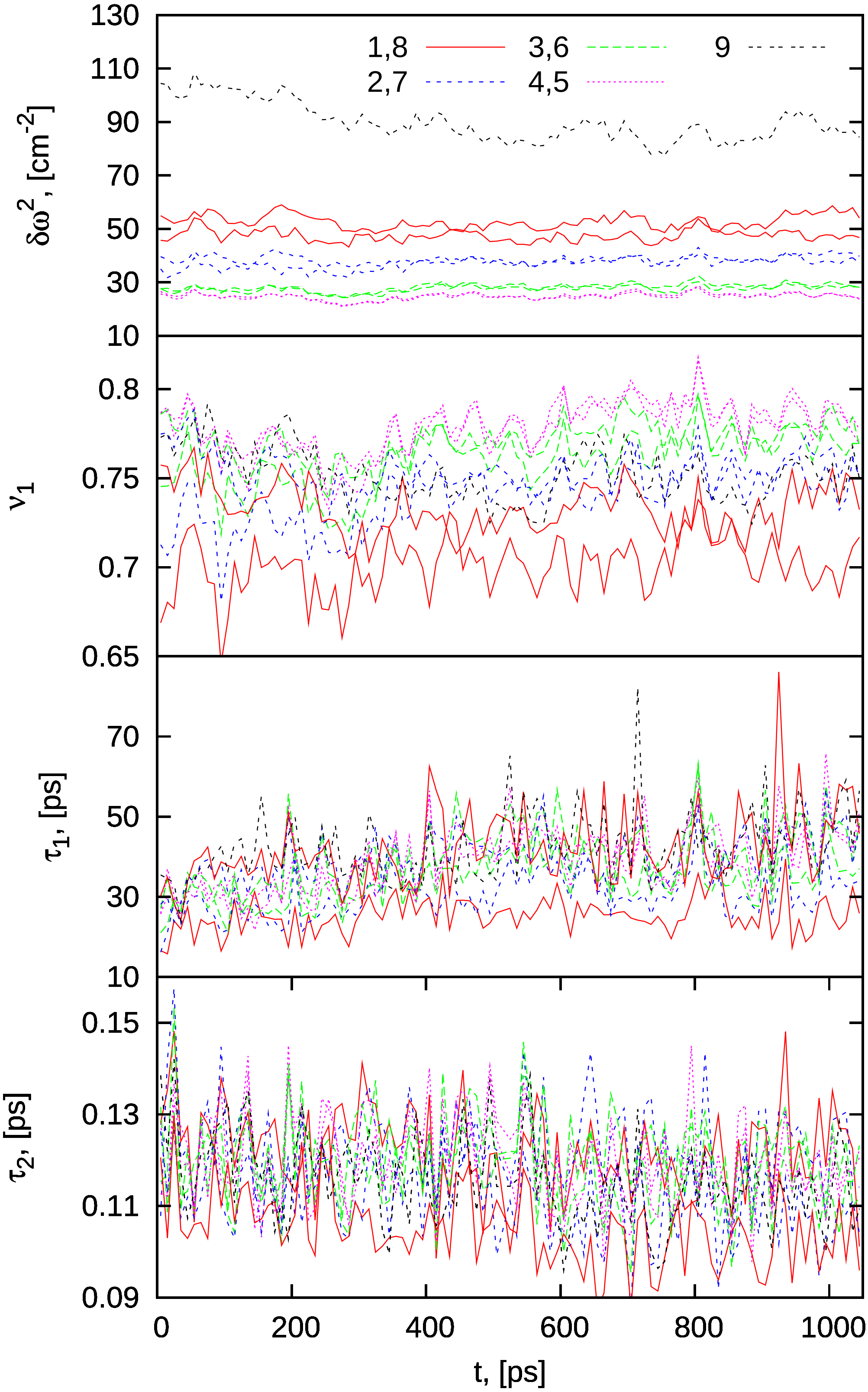

Figure 3.18: Fitting parameter of the correlation functions of the 8 normal mode frequencies as functions of time.

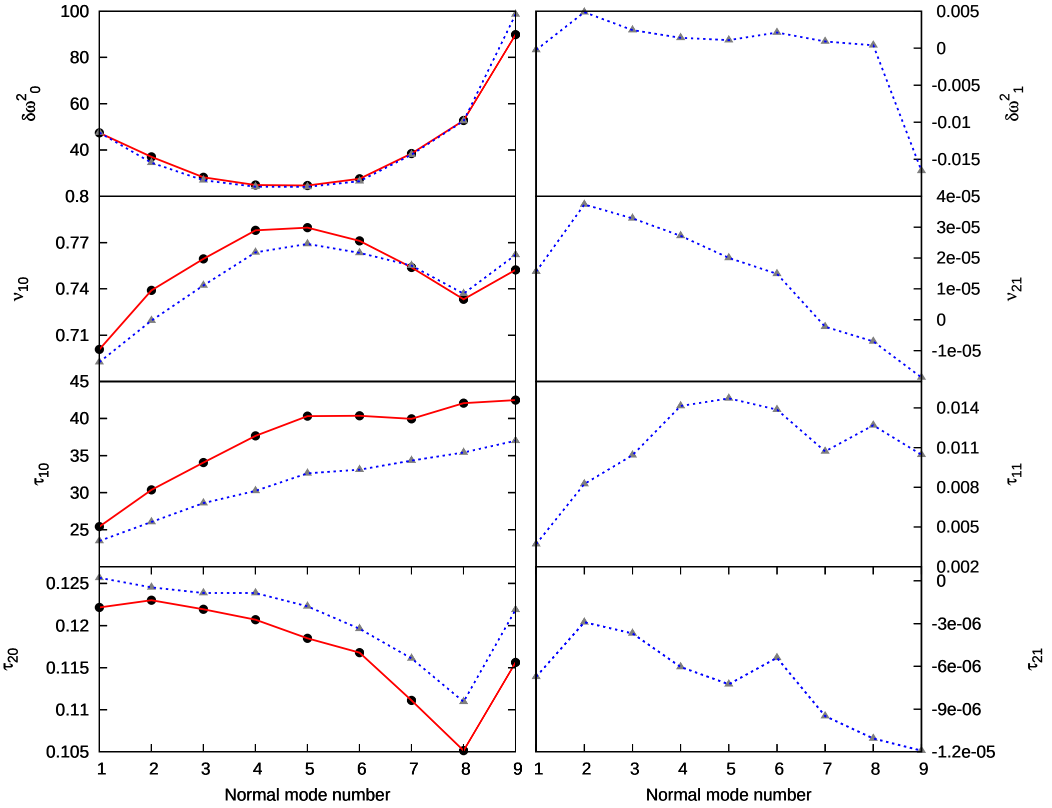

Figure 3.19: Fitting parameter of the correlation functions of the normal modes frequencies obtained by the averaging over time (red curves) as well as by linear fit of the time dependencies (blue curves).

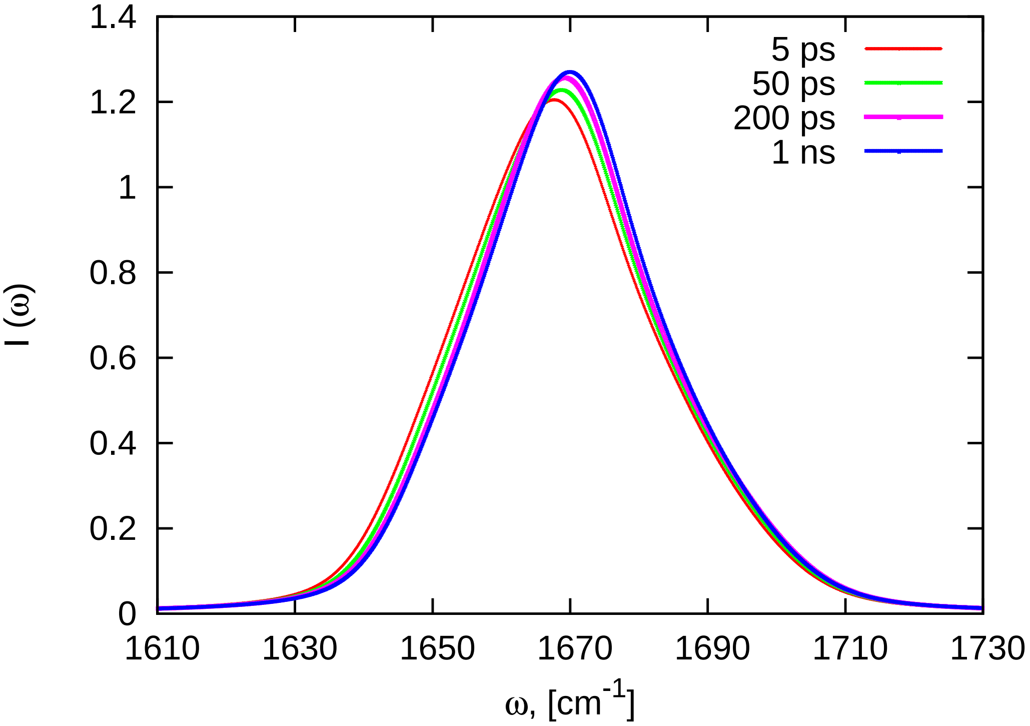

Figure 3.20: Time dependent spectra of photoswitchable peptide.

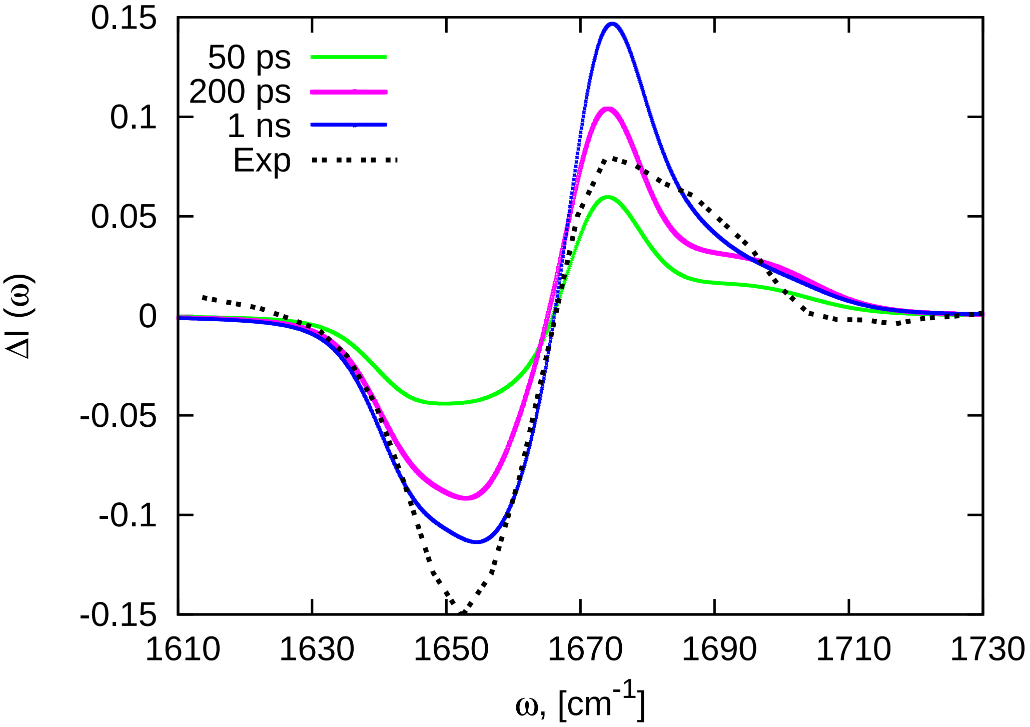

Figure 3.21: Changes of the vibrational spectrum of photoswitchable peptide.

modes. However, for the fixed peak the form is almost constant in time. Moreover, the most significant changes in the positions of peaks occur at first 200 ps. All peaks shift to the higher frequency region. In the central region difference between positions and forms of the neighboring peaks are smaller.

In such way we have demonstrated how our vibrational Hamiltonian can be used to interpreter small time dependent changes of the vibrational spectra in terms of conformational changes.

Conclusions

The presented work can be divided into two main parts. In the first one we have discussed various aspects of the *ab initio*-based parameterization of an exciton model of amide I vibrations in peptides. In the second part the constructed vibrational Hamiltonian has been used for the calculation of peptide vibrational spectra. As a first attempt to establish a model system for the simulation of IR spectra of flexible peptides in aqueous solution, we have performed a quantum-classical description of the amide I vibrational spectrum of trialanine cation in D_2O. The vibrational Hamiltonian has also been applied to the modeling of time dependent spectrum of a photoswitchable peptide.

The first part of the *ab initio*-based parameterization of the Hamiltonian was dedicated to the detailed consideration of a small peptide which can be used as the building block for the construction of polypeptide Hamiltonians. Adopting glycine dipeptide as a simple building-block model that describes the vibrational interaction between two peptide units, we have performed comprehensive DFT calculations to investigate the effect and importance of the level of theory, choice of local modes, and parameterization schemes. The main results of this study can be summarized as follows:

- DFT calculations using a 6-31G+(d) basis set and the B3LYP functional are a good compromise between high accuracy and low computational effort. This level of theory allows for a qualitatively correct characterization of the amide I vibrations in peptides.

- Different aspects of the *ab initio*-based parameterization have been considered. In particular we have compared different parameterizations schemes, local modes and types of geometry optimization. The Hessian matrix reconstruction in combination with the full geometry optimization and C=O based local modes was found to be the best choice.

- The main effects of a solvent continuum model on the amide I vibrations are an overall redshift of ≈ 80 cm^{-1} of the frequencies as well as an overall reduction of the conformational fluctuations of the vibrational constants. Furthermore, several details of the (ϕ, ψ)-maps change upon solvation.

- Performing conformational averages of the (ϕ, ψ)-maps with respect to the most important conformational states in solution (that is, α, β, $\mathrm{P_{II}}$, and C_5), the site energies and the intersite coupling do reflect the conformation of the peptide, although there is a large overlap of the state-specific distributions.
- Generally speaking, the vibrational coupling is found to be quite robust with respect to most modifications under consideration. In contrast, the vibrational frequencies depend sensitively on many details of the calculation.
- We have considered higher oder anharmonicities of the amide I potential energy in mono- and di-peptides. In particular we have found that cubic anharmonicities in dipeptides do not depend on conformation within a good approximation.
- The (ϕ, ψ)-maps of the vibrational coupling β as well as of the frequency splitting $\Delta\omega$ obtained for GD were found to be transferable to dipeptides with hydrophilic and hydrophobic side chains as well as to tripeptides with charged end-groups. However, the corresponding maps for the site energies match the results for GD only approximately.

The second part of the *ab initio*-based parameterization is the the usage of the building block model in combination with different electrostatic models. In this work we have also studied the virtues and limits of an *ab initio*-based building block model of amide I vibrations in peptides. It has been tested on the tripeptide for which the vibrational properties have been obtained directly. An accurate *ab initio* description has been found essential to account for the effect of next-neighbor residues on the local-mode frequencies ε_n and the vibrational couplings β_{nm}, respectively, while the long-distance (i.e., not next-neighbor) interactions can be well approximated by electrostatic models. Employing this combination, we typically obtain an accuracy of a few wavenumbers for ε_n and less than wavenumber for β_{nm}. As the vibrational data are obtained for peptide conformations with arbitrary backbone dihedral angles $\{\phi_n, \psi_n\}$, the model is capable of describing large conformational rearrangements occurring, e.g., in protein binding or folding processes.

We note that the considered building block model is in the spirit of systematic fragmentation schemes recently suggested by Zhang [68] and Collins [69] and their coworkers. While we choose a dipeptide as smallest fragmentation unit, it is clear that the model in principle can be converged to the exact result if larger molecular fragments are considered. For example, one could use the above results for the tripeptide to calculate the vibrational response of polypeptides using an analogous decomposition into tripeptide units [69].

In the second part of the work we apply the obtained vibrational Hamiltonians to calculate vibrational spectra of different solvated peptides. In such way we test accuracy of different spectroscopic approximations. The first considered system is trialanine molecule. Driven by a number of theoretical and experimental studies this molecule has emerged as a paradigm to study conformational dynamics of a small peptide in aqueous solution. The consideration of trialanine was a first attempt to establish simple yet nontrivial model system, for which most parts of the modeling can be achieved accurately enough to be trusted. On the basis of the obtained results we can conclude that the appropriate description of the conformational distribution and the correct calculation of the dynamic absorption spectrum has been reached. What is missing is an accurate enough quantum-chemical modeling of the vibrational frequencies of a solvated peptide — a topic that represents a quite active field of research.

Concerning the accuracy of spectroscopic approximations the following conclusions have been drawn. The cumulant expansion is accurate enough if the state-specific spectra are calculated. In the case of the total spectra, consisting of peaks coming from different conformations and shifted with respect to each other, we have clearly asymmetric spectra and, as a consequence, the second-order cumulant expansion, giving symmetric spectra, cannot be applied. Similarly to the second-order cumulant expansion the Condon approximations seams to perform well if the state-specific spectra are calculated. The latter means that within a specific conformation the transition dipole moment of the system can be considered as a constant within a good approximation. We have also found that the total spectra can, within a good approximation, be presented as a weighted sum of the state-specific spectra. This gives us a possibility to treat the total spectra in terms of the contributions coming from different conformations. And finally, nonadiabatic transitions were found to give a significant effect on the vibrational spectra.

The second system, for which the vibrational spectra have been calculated, was a photoswitchable peptide. With this consideration we have demonstrated a practical procedure of how the time dependent spectra can be treated in terms of conformational changes. Moreover, we have shown that even small spectroscopic effects can, in principle, be obtained within the considered model. And finally, we have developed a strategy of consideration of nonequilibrium processes.

Our study underlines that the amide I response of peptides depends on a number of aspects. While time-resolved IR spectroscopy hold the promise to resolve conformational dynamics in real time, [87] it certainly requires substantial theoretical support.

Bibliography

[1] Woutersen, S., Hamm, P. Structure determination of trialanine in water using polaization sensitive two-dimensional vibrational spectroscopy. J. Phys. Chem. B 104:11316, 2000.

[2] Mukamel, S., Hochstrasser, R. M., eds. Special Issue on Multidimensional Spectroscopies. .: Chem. Phys. **266**, No.2-3. 2001.

[3] Woutersen, S., Hamm, P. Nonlinear two-dimensional vibrational spectroscopy of peptides. J. Phys. Cond. Mat. 14:1035, 2002.

[4] Krimm, S., Bandekar, J. Vibrational spectroscopy and conformations of peptides, polypeptides, and proteins. Adv. Prot. Chem. 38:181, 1986.

[5] Torii, H., Tatsumi, M. In: Infrared spectroscopy of Biomolecules. Mantsch, H. H., Chapman, D. eds. . Wiley-Liss New York 1996 1.

[6] Schweitzer-Stenner, R., Eker, F., Huang, Q., Griebenow, K. Dihedral angles of trialanine in D_2O determined by combining FTIR and polarized visible Raman spectroscopy. J. Am. Chem. Soc. 123:9628, 2001.

[7] Eker, F., Cao, X., Nafie, L. A., Schweitzer-Stenner, R. Tripeptides adopt stable structures in water: A combined polarized visible raman, ftir, and vcd spectroscopy study. J. Am. Chem. Soc. 124:14330, 2002.

[8] Schweitzer-Stenner, R., Eker, F., Griebenow, K., Cao, X., Nafie, L. A. The conformation of tetraalanine in water determined by polarized raman, ft-ir, and vcd spectroscopy. J. Am. Chem. Soc. 126:2768, 2004.

[9] Eker, F., Griebenow, K., Cao, X., Nafie, L. A., Schweitzer-Stenner, R. Preferred peptide backbone conformations in the unfolded state revealed by the structure analysis

of alanine-based (axa) tripeptides in aqueous solution. Proc. Natl. Acad. Sci. USA 101:10054, 2004.

[10] Chung, H., Khalil, M., Smith, A. W., Ganim, Z., Tokmakoff, A. Conformational changes during the nanosecond-to-millisecond unfolding of ubiquitin. Proc. Natl. Acad. Sci. USA 102:612–617, 2005.

[11] Hamm, P., Lim, M., Hochstrasser, R. M. Struture of the amide I band of peptides measured by femtosecond nonlinear infrared spectroscopy. J. Phys. Chem. B 102:6123, 1998.

[12] Zanni, M. T., Ge, N.-H., Kim, Y. S., Hochstrasser, R. M. Proc. Natl. Acad. Sci. USA 98:11265, 2001.

[13] Woutersen, S., Hamm, P. Isotope-edited two-dimensional vibrational spectroscopy of trialanine in aqueous solution. J. Chem. Phys. 114:2727–2737, 2001.

[14] Woutersen, S., Mu, Y., Stock, G., Hamm, P. Subpicosecond Conformational Dynamics of Small Peptides Probed by Two-dimensional Vibrational Spectroscopy. Proc. Natl. Acad. Sci. USA 98:11254, 2001.

[15] Woutersen, S., Pfister, R., Hamm, P., Mu, Y., Kosov, D., Stock, G. Peptide conformational heterogeneity revealed from nonlinear vibrational spectroscopy and molecular dynamics simulations. J. Chem. Phys. 117:6833, 2002.

[16] Marx, D., Hutter, J. In: Ab Initio Molecular Dynamics: Theory and Implementation. Grotendorst, J. ed. . NIC FZ Jülich 2000 301. www.fz-juelich.de/nic-series/Volume1/.

[17] Carloni, P., Röthlisberger, U., Parrinello, M. The role and perspective of ab initio molecular dynamics in the study of biological systems. Acc. Chem. Res. 35:455, 2002.

[18] Rousseau, R., Kleinschmidt, V., Schmidt, U. W., Marx, D. Assigning protonation patterns in water networks in bacteriorhodopsin based on computed ir spectra. Angew. Chem. Int. Ed. 43:4804, 2004.

[19] Corcelli, S. A., Lawrence, C. P., Skinner, J. L. Combined electronic structure/molecular dynamics approach for ultrafast infrared spectroscopy of dilute HOD in liquid H2O and D2O. J. Chem. Phys. 120:8107, 2004.

[20] la Cour Jansen, T., Zhuang, W., Mukamel, S. Stochastic Liouville equation simulation of multidimensional vibrational line shapes of trialanine. J. Chem. Phys. 121:10577, 2004.

[21] Gnanakaran, S., Hochstrasser, R. M., Garcia, A. E. Nature of structural inhomogeneities on folding a helix and their influence on spectral measurements. Proc. Natl. Acad. Sci. USA 101:9229–9234, 2004.

[22] Schmitz, M., Tavan, P. Vibrational spectra from atomic fluctuations in dynamics solutions. i. theory, limitations, and a sample application. J. Chem. Phys. 121:12233–12246, 2004.

[23] Hahn, S., Ham, S., Cho, M. Simulation studies of amide I IR absorption and two-dimensional IR spectra of β-hairpins in liquid water. J. Phys. Chem. B 109:11789–11901, 2005.

[24] Bour, P., Keiderling, T. A. Vibrational spectral simulation for peptides of mixed secondary structure: Method comparisons with the trpzip model hairpin. J. Phys. Chem. B 109:23687–23697, 2005.

[25] Zhuang, W., Abramavicius, D., Hayashi, T., Mukamel, S. Simulation protocols for coherent femtosecond vibrational spectra of peptides. J. Phys. Chem. B 110:3362–3374, 2003.

[26] Ganim, Z., Tokmakoff, A. Spectral signatures of heterogeneous protein ensembles revealed by md simulations of 2dir spectra. Biophys. J. 91:2636–2646, 2006.

[27] Moran, A. M., Park, S.-M., Dreyer, J., Mukamel, S. Linear and nonlinear infrared signatures of local alpha- and 3_{10}-helical structures in alanine polypeptides. J. Chem. Phys. 118:3651, 2003.

[28] Ham, S., Cho, M. Amide I modes in the n-methylacetamide dimer and glycine dipeptide analog: Diagonal force constants. J. Chem. Phys. 118:6915, 2003.

[29] Torii, H. Vibrational interactions in the amide i subspace of the oligomers and hydration clusters of n-methylacetamide. J. Phys. Chem. A 108:7272, 2004.

[30] Gregurick, S. K., Chaban, G. M., Gerber, R. B. J. Phys. Chem. A 106:8696, 2002.

[31] Torii, H., Tasumi, M. Ab initio molecular orbital study of the amide I vibrational interactions between the peptide groups in di- and tripeptides and considerations on the conformation of the extended helix. J. Raman Spectrosc. 29:81–86, 1998.

[32] Hamm, P., Woutersen, S. Coupling of the amide I modes of the glycine dipeptide. Bull. Chem. Soc. Jpn. 75:985, 2002.

[33] Cha, S., Ham, S., Cho, M. Amide I vibrational modes in glycine dipeptide analog: Ab initio calculation studies. J. Chem. Phys. 117:740, 2002.

[34] Antony, J., Schmidt, B., Schütte, C. Nonadiabatic effects on peptide vibrational dynamics induced by conformational changes. J. Chem. Phys. 122:014309, 2005.

[35] Kwac, K., Cho, M. Molecular dynamics simulation study of n-methylacetamide in water. i. amide i mode frequency fluctuation. J. Chem. Phys. 119:2247, 2003.

[36] Choi, J.-H., Ham, J., Cho, M. Local amide I mode frequencies and coupling constants in polypeptides. J. Phys. Chem. B 107:9132–9138, 2003.

[37] Bour, P., Keiderling, T. A. Empirical modeling of the peptide amide I band IR intensity in water solution. J. Chem. Phys. 119:11253, 2003.

[38] Schmidt, J. R., Corcelli, S. A., Skinner, J. L. Ultrafast vibrational spectroscopy of water and aqueous N-methylacetamide: Comparison of different electronic structure/molecular dynamics approaches. J. Chem. Phys. 121:8887, 2004.

[39] Besley, N. A. Ab initio modeling of amide vibrational bands in solution. J. Phys. Chem. A 109:10794, 2004.

[40] Hayashi, T., la Cour Jansen, T., Zhuang, W., Mukamel, S. Collective sovent coordinates for the infrared spectrum of hod in d_2o based on an ab initio electrostatic map. J. Phys. Chem. A 109:64–82, 2005.

[41] la Cour Jansen, T., Knoester, J. A transferable electrostatic map for solvation effects on amide i vibrations and its application to linear and two-dimensional spectroscopy. J. Chem. Phys. 124:044502, 2006.

[42] Li, S., Schmidt, J. R., Corcelli, S. A., Lawrence, C. P., Skinner, J. L. Approaches for the calculation of vibrational frequencies in liquids: Comparison to benchmarks for azide/water clusters. J. Chem. Phys. 124:204110, 2006.

[43] Kubo, R. J. J. Phys. Soc. Japan 12:570, 1957.

[44] Mukamel, S. Principles of Nonlinear Optical Spectroscopy. Oxford: University Press. 1995.

[45] Schmidt, J. R., Corcelli, S. A., Skinner, J. L. Pronounced non-condon effects in the ultrafast infrared spectroscopy of water. J. Chem. Phys. 123:044513, 2005.

[46] Kwac, K., Lee, H., Cho, M. Non-gaussian statistics of amide I frequency fluctuation of *n*-methylacetamide in methanol solution: Linear and nonlinear vibrational spectra. J. Chem. Phys. 120:1477–1490, 2004.

[47] Piryatinski, A., Lawrence, C. P., Skinner, J. L. Vibrational spectroscopy of hod in liquid d2o iv. infrared two-pulse photon echoes. J. Chem. Phys. 118:9664, 2003.

[48] Giese, K., Petkovic, M., Naundorf, H., Kühn, O. Multidimensiona quantum dynamics and infrared spectroscopy of hydrogen bonds. Phys. Rep. 430:211–276, 2006.

[49] Schweitzer-Stenner, R. Biophys. J. 83:523, 2002.

[50] Mu, Y., Stock, G. Conformational dynamics of trialanine in water: A molecular dynamics study. J. Phys. Chem. B 106:5294, 2002.

[51] Mu, Y., Kosov, D. S., Stock, G. Conformational dynamics of trialanine in water II: Comparison of AMBER, CHARMM, GROMOS, and OPLS force fields to NMR and infrared experiments. J. Phys. Chem. B 107:5064, 2003.

[52] Gnanakaran, S., Garcia, A. E. Validation of an all-atom protein force field: from dipeptides to larger peptides. J. Phys. Chem. B 107:12555–12557, 2003.

[53] Gorbunov, R. D., Kosov, D. S., Stock, G. Ab initio-based exciton model of amide I vibrations in peptides: Definition, conformational dependence, and transferability. J. Chem. Phys. 122:224904–224915, 2005.

[54] Graf, J., Nguyen, P. H., Stock, G., Schwalbe, H. Structure and dynamics of the homologues series of alanine peptides: A joint molecular-dynamics/nmr study. J. Am. Chem. Soc. in press, 2006.

[55] Renner, C., Cramer, J., Behrendt, R., Moroder, L. Photomodulation of conformational states. ii mono- and bicyclic peptides with (4-aminomethyl)-phenylazobenzoic acid as backbone constituent. Biopolymers 54:501, 2000.

[56] Spoerlein, S., Carstens, H., Renner, C., Behrendt, R., Moroder, L., Tavan, P., Zinth, W., Wachtveitl, J. Ultrafast spectroscopy reveals subnanosecond peptide conformational dynamics and validates molecular dynamics simulation. Proc. Natl. Acad. Sci. USA 99:7998, 2002.

[57] Wachtveitl, J., Sporlein, S., Saltger, H., Fonrobert, B., Renner, C., Behrendt, R., Oesterhelt, D., Moroder, L., Zinth, W. Ultrafast conformational dynamics in cyclic azobenzene peptides of increased flexibility. Biophys. J. 86:2350, 2004.

[58] Bredenbeck, J., Helbing, J., Sieg, A., Schrader, T., Zinth, W., Renner, C., Behrendt, R., Moroder, L., Wachtveitel, J., Hamm, P. Picosecond conformational transition and equilibration of a cyclic peptide. Proc. Natl. Acad. Sci. 100:6452, 2003.

[59] Bredenbeck, J., Helbing, J., Renner, C., Moroder, L., Wachtveitl, J., Hamm, P. Transient 2d-ir spectroscopy: Snapshots of the nonequilibrium ensemble during the picosecond conformational transition of a small peptide. J. Phys. Chem. B 107:8654, 2003.

[60] Mukamel, S., Hochstrasser, R. M. Editors special issue on multidimensional spectroscopies. Chem. Phys. 266:135, 2001.

[61] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Zakrzewski, V. G., J. A. Montgomeryr, J., Stratmann, R. E., Burant, J. C., Dapprich, S., Millam, J. M., Daniels, A. D., Kudin, K. N., Strain, M. C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G. A., Ayala, P. Y., Cui, Q., Morokuma, K., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Cioslowski, J., Ortiz, J. V., Baboul, A. G., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Andres, J. L., Gonzalez, C., Head-Gordon, M., Replogle, E. S., Pople, J. A. Gaussian 98, Revision A.9. Pittsburgh PA: Gaussian, Inc.. 1998.

[62] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Montgomery, J. A., Jr., Vreven, T., Kudin, K. N., Burant, J. C., Millam, J. M., Iyengar, S. S., Tomasi, J., Barone, V., Mennucci, B., Cossi, M., Scalmani, G., Rega, N., Petersson, G. A., Nakatsuji, H., Hada, M., Ehara, M., Toyota, K., Fukuda, R.,

Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Klene, M., Li, X., Knox, J. E., Hratchian, H. P., Cross, J. B., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Ayala, P. Y., Morokuma, K., Voth, G. A., Salvador, P., Dannenberg, J. J., Zakrzewski, V. G., Dapprich, S., Daniels, A. D., Strain, M. C., Farkas, O., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Ortiz, J. V., Cui, Q., Baboul, A. G., Clifford, S., Cioslowski, J., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Challacombe, M., Gill, P. M. W., Johnson, B., Chen, W., Wong, M. W., Gonzalez, C., Pople, J. A. Gaussian 03, Revision B.03. Pittsburgh PA: Gaussian, Inc.. 2003.

[63] Becke, A. D. Phys. Rev. A 38:3098, 1988.

[64] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. J. Chem. Phys. 98:5648–5652, 1993.

[65] Lee, C., Yang, W., Parr, R. G. Phys. Rev. B 37:785, 1988.

[66] Mirkin, N. G., Krimm, S. Amide iii mode , dependence in peptides: A vibrational frequency map. J. Phys. Chem. A 106:3391, 2002.

[67] Ham, S., Cha, S., Choi, J., Cho, M. Amide i modes of tripeptides: Hessian matrix reconstruction and isotope effects. J. Chem. Phys. 119:1451, 2003.

[68] Zhang, D. W., Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein-molecule interaction energy. J. Chem. Phys. 119:3599–3605, 2003.

[69] Deev, V., Collins, M. A. Approximate *ab initio* energies by systematic molecular fragmentation. J. Chem. Phys. 122:154102, 2005.

[70] Bour, P., Sopkova, J., Bednarova, L., Maloncaron, P., Keiderling, T. A. Transfer of molecular property tensors in cartesian coordinates: A new algorithm for simulation of vibrational spectra. J. Comput. Chem. 18:646–659, 1997.

[71] Choi, J.-H., Kim, J.-S., Cho, M. Amide I vibrtional circular dichroism of polypeptides: Generalized fragmentation approximation method. J. Chem. Phys. 122:174903, 2005.

[72] la Cour Jansen, T., Dijkstra, A. G., Watson, T. M., Hirst, J. D., Knoester, J. Modeling the amide i bands of small peptides. J. Chem. Phys. 125:044312, 2006.

[73] Davidov, A. S. Theory of molecular excitons. New York: Plenum. 1971.

[74] Hayashi, T., Zhuang, W., Mukamel, S. Electrostatic DFT map for the complete vibrational amide band of NMA. J. Phys. Chem. A 109:9747–9759, 2005.

[75] Moran, A. M., Mukamel, S. The origin of vibrational mode couplings in various secondary structural motifs of polypeptides. Proc. Natl. Acad. Sci. USA 101:506, 2004.

[76] Torii, H., Tasumi, M. Model calculations on the amide-i infrared bands of globular proteins. J. Chem. Phys. 96:3379–3387, 1992.

[77] Hu, H., Elstner, M., Hermans, J. Comparison of a qm/mm force field and molecular mechanics force fields in simulations of alanine and glycine "dipeptide" (ace-ala-nme and ace-gly-nme) in water in relation to the problem of modeling the unfolding peptide backbone in solution. Proteins 50:451–463, 2003.

[78] Adam, D. J., Dubey, G. S. J. Comput. Phys. 72:156, 1987.

[79] van Gunsteren, W. F., Billeter, S. R., Eising, A. A., Hünenberger, P. H., Krüger, P., Mark, A. E., Scott, W. R. P., Tironi, I. G. Biomolecular Simulation: The GROMOS96 Manual and User Guide. Zürich: Vdf Hochschulverlag AG an der ETH Zürich. 1996.

[80] Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., Hermans, J. Interaction models for water in relation to protein hydration. In: Intermolecular Forces. Pullman, B. ed. . D. Reidel Publishing Company Dordrecht 1981 331–342.

[81] Gorbunov, R. D., Nguyen, P. H., Kobus, M., Stock, G. Quantum-classical description of the amide i vibrational spectrum of trialanine. J. Chem. Phys. submitted.

[82] Köppel, H., Domcke, W., Cederbaum, L. S. Multimode molecular dynamics beyond the Born-Oppenheimer approximation. Adv. Chem. Phys. 57:59, 1984.

[83] Light, J. C., Bacic, Z. Adiabatic approximation and nonadiabatic corrections in the discrete variable representation - highly excited vibrational-states of triatomic-molecules. J. Chem. Phys. 87:4008–4019, 1987.

[84] Makri, N. Improved feynman propagators on a grid and nonadiabatic corrections within the path integral framework. Chem. Phys. Lett. 193:435, 1992.

[85] Hammes-Schiffer, S. Multiconfigurational molecular dynamics with quantum transitions: Multiple proton transfer reactions. J. Chem. Phys. 105:2236–2246, 1996.

[86] May, V., Kühn, O. Charge and Energy Transfer Dynamics in Molecular Systems. Berlin: Wiley-VCH. 2004.

[87] Kolano, C., J.Helbig, Kozinski, M., Sander, W., Hamm, P. Ultrafast β-turn opening observed by transient 2D-IR spectroscopy. Nature (London) 444:469–472, 2006.

[88] Nguyen, P. H., Stock, G. Nonequilibrium molecular dynamics simulation of a photoswitchable peptide. Chem. Phys. 323:36, 2006.

Acknowledgments

I would like to express my extreme gratitude to Professor Gerhard Stock who gave me the opportunity, recourses and support to complete the work for this theses. The discussions with him were very helpful and motivating. He always entered into details of the problems, gave a lot of valuable ideas and recommendations. I also highly appreciate knowledge on spectroscopy which he imparted to me.

I am vary thankful to Professor Daniel Kosov for his support in the first part of my PhD. In particular I highly appreciate his strong involvement into my PhD project as well as his readiness to help at any time. He shared with me his excellent experience in *ab initio* calculations and second quantization formalism.

I would like to thank Doctor Nguyen Phuong for close and nice collaboration. Without his data, large part of my calculations would be impossible. I vary appreciate his efforts taken to generate input for my calculations.

I would like to thank Doctor Rainer Hegger for taking care of all technical problems related with the computers and application softwares. His professional support creates a convenient environment for all types of calculations.

I am thankful to the other members of the Gerhard group past and present: Elisabeth Widjajakusuma, Alexandros Altis, Igor Uspenskiy, Sang Min Park, Alessandra Villa, Maja Kobus, Birgit Strodel and Jessica Koplin for positive and lively working atmosphere. I also would like to thank all members of groups of Professor Josef Wachtveitl and Doctor Andreas Dreuw for nice and friendly company.

I am also grateful to my friends outside the institute. In particular to Olexander Tsymbalyuk who was very friendly and great fun to be with. Special though goes to Alena Siarheyeva. Thanks to her my stay in Frankfurt remains one of the brightest moments of my life to remember.

I would especially like to thank my parents and brother for all their love and support, it really helps me.

Zusammenfassung

Die vorliegende Arbeit behandelt die Entwicklung und Überprüfung von Modellen zur Berechnung von Schwingungspektren von Peptiden und Proteinen. Solche Modelle verbinden die Konformationsstruktur eines Moleküls mit seinen Schwingungseigenschaften und sind demzufolge wichtig für die Interpretation der Schwingungspektren. Die im Rahmen dieser Arbeit durchgeführte theoretische Erforschung dieses Gebietes beschränkt sich auf die Betrachtung der Amide-I-Moden, welche aufgrund ihrer physikalischen Eigenschaften sich zur Untersuchung der Peptidkonformationen eignen. Die Arbeit kann prinzipiell in zwei Teile separiert werden. In dem ersten Teil werden Fragen betrachtet, die mit der Entwicklung des Schwingungshamiltonian verbunden sind. Im zweiten Teil wurden die erhaltenen Hamiltonian für die Berechnung der Schwingungspektren verwendet. Bei der Berechnung der Schwingungspektren wurden verschiedene spektroskopische Näherungen verwendet und erforscht.

Die Entwicklung des Schwingungshamiltonian beinhaltet zwei Aufgaben. Die *ab initio* Parametrisierung des Schwingungshamiltonian von Dipeptiden, sowie die Analyse der Entwicklungsmethoden für Schwingungshamiltonian von Polypeptiden. Die Entwicklungsmethoden stützen sich auf *ab initio* berecheten Schwingungseigenschaften von Dipeptiden und/oder elektrostatische Modelle. Die *ab initio* Parametrisierung basiert auf einer Geometrieoptimierung und anschließender Berechnung von Normalmoden. Hierbei wurde die Abhängigkeit der Ergebnisse vom theoretischen Niveau und dem verwendeten Basissatz untersucht. Die Transformation der errechneten Normalmoden lieferte die Schwingungseigenschaften der lokale Amide-I-Mode. Die Lokalisierung der Normalmode folgt diversen Kriterien. Sie ist von der Wahl der Lokalmoden und somit implizit auch von der Art der Geometrieoptimierung abhängig. Mit dieser Arbeit konnte die Abhängigkeit der Ergebnisse von der Parameterwahl weitgehend aufgeklärt und eine für das Amide-I-System geeignet Parametrisierung gefunden werden.

Im nächsten Arbeitsschritt wurde die Abhängigkeit der Amide-I-Schwingungseigenschaften von den Peptidseitenketten und terminalen Gruppen untersucht. Desweiteren wurden

Methoden zur Formulierung der Hamiltonian für Polypeptide konzeptionell entwickelt. Diese Untersuchung ist außerordentlich wichtig, da direkte quantenmechanische Berechnungen von Polypeptiden zu zeitaufwendig sind. Solche Methoden beruhen auf dem sogenannten "Building-Block"-Ansatz und verschiedenen elektrostatischen Modellen. In dieser Arbeit wurden sowohl die einzelnen Methoden als auch ihre Kombination für die Entwicklung des Hamiltonians verwendet. Zur Abschätzung der Genauigkeit der verwendeten Methoden wurden Vergleichsrechnungen durchgeführt. Weiterhin wurden nichtlineare Terme der Amide-I-Potenzialenergie von Mono- und Dipeptiden analysiert. Es stellte sich heraus, dass in guter Näherung die ungemischten kubischen Terme konformationsunabhängig sind und die gemischten Terme Null sind. Ausschließend wurde die Konvergenz der ersten zwei Energieniveaus der Amide-I-Schwingung in dem Monopeptid mit größer werdender Gradzahl der nichtlinearen Terme in der Potenzialenergie untersucht.

Im zweiten Teil dieser Arbeit wurden die erhaltenen Schwingungshamiltonian zur Berechnung von Schwingungsspektren diverser gelöster Peptide angewandt. In diesem Zusammenhang konnte die Genauigkeit unterschiedlicher spektroskopischer Approximationen überprüft werden. Das erste behandelte System ist das Trialanin Molekül. Aufgrund der Vielzahl der theoretischen und experimentellen Untersuchungen ist dieses Molekül eine gute Referenz, um konformationelle Dynamik eines kleinen Peptids in wässriger Lösung zu studieren. Die Betrachtung von Trialanin war ein erster Versuch um ein einfaches aber nicht triviales Modellsystem zu erhalten, bei dem die Modellierung größtenteils genau genug ist. Auf Grundlage der erhaltenen Ergebnisse können wir sagen, dass eine angemessene Beschreibung der konformationellen Verteilung und eine korrekte Berechnung des dynamischen Absorptionsspektrum gewährleistet ist. Was noch fehlt, ist ein hinreichend genaues quantenchemisches Modell für die Schwingungsfrequenzen eines gelösten Peptids. Diese Aufgabe stellt zur Zeit ein aktives Forschungsgebiet dar.

Bezüglich der untersuchten spektroskopischen Näherungen wurden die folgenden Schlüsse gezogen: Die Kumulantentwicklung ist hinreichend genau, um das Spektrum eines spezifischen Zustandes wiederzugeben. Hinsichtlich des gesamten Spektrums kann die Kumulantenäherung zweiter Ordnung nicht angewandt werden, da diese nur für symmetrische Spektren geeignet ist. Aufgrund der verschiedenen Peptid Konformationen und der gegenseitigen Verschiebung der Peaks, ist jedoch ein asymmetrisches Spektrum zu erwarten. Entsprechend zuverlässig erweist sich die Condon Näherung im Falle eines spezifischen Zustands. Bezüglich einer spezifische Konformation können die Übergangsdipolmomente als konstant angenommen werden. Ferner kann das gesamte Spektrum in guter Näherung als gewichtete Summe der zustandsspezifischen Spektren dargestellt werden. Darüberhinaus zeigte sich, dass nichtadiabatische Übergänge sich signifikant auf das Schwingungsspektrum auswirken.

Zuletzt wurde das Schwingungsspektrum eines sogenanten "Photoschaltbaren"-Peptids simuliert. Mit Hilfe des dafür aufgestellten Hamiltonians ist man in der Lage spektroskopische Beobachtungen auf Konformationsänderungen direkt zu übertragen. Im Rahmen dieser Arbeit wurde eine Strategie für die Beschreibung von Nichtgleichgewichtsprozessen entwickelt.

Zusammenfassend kann man sagen, dass die spektroskopische amide-I-Antwort der Peptide von zahlreichen Aspekten abhängt. Auf experimentieller Seite bietet die zeitaufgelöste IR-Spektroskopie die Möglichkeit Konformationsdynamiken von Peptiden zu betrachten. Umso mehr müssen neue theoretische Modellierungen gefunden werden, um die Beobachtungen richtig interpretieren zu könen.

Lebenslauf

Roman D. Gorbunov

Personliche Angaben

Geburtsdatum und -ort	07.09.1978 in Dniprodzerzhinsk, Ukraine
Familienstand	ledig
Staatsangehorigkeit	ukrainisch

Schulausbildung:

- Sep. 1985 - Juni 1995: Allgemeinbildende Schule Nr.1, Stadt Auli, Dnipropetrovsk Gebiet, Ukraine

Studium:

- Sep. 1995- Juni. 2000: Studium an der Physik-Fakultät mit Schwerpunkt theoretische Physik der Dnipropetrovsk Staatlichen Universität
- Juni. 2000 Diplom mit Auszeichnung im Fach "Physik" mit Schwerpunkt "Theoretische Physik"

Promotion:

seit 01.02.2003	Doktorand in der Arbeitsgruppe von Prof. Dr. Gerhard Stock am Institut fur Physikalische und Theoretische Chemie der Johann Wolfgang Goethe - Universität in Frankfurt am Main Thema: " Models for the Calculation of Peptide Vibrational Spectra "

Publikationen:

- R. D. Gorbunov, D. S. Kosov, and G. Stock, "Ab initio-based exciton model of amide I vibrations in peptides: Definition, conformational dependence, and transferability", J. Chem. Phys., **122**, 224904 (2005).
- R. D. Gorbunov, P. H. Nguyen, M. Kobus, and G. Stock, "Quantum-classical description of the amide I vibrational spectrum of trialanine", J. Chem. Phys., (Accepted) (2007).
- R. D. Gorbunov and G. Stock, "Ab initio based building block model of amide I vibrations in peptides", CPL, (Accepted) (2007).
- P. H. Nguyen, R. D. Gorbunov, and G. Stock. "Photoinduced conformational dynamics of a photoswitchable peptide: A nonequilibrium molecular dynamics simulation study", BJ, **91**, 1224 (2006).