Benchmarking of Vibrational Exciton Models Against Quantum-Chemical Localized-Mode Calculations

Anna M. van Bodegraven[1], Kevin Focke[2], Mario Wolter[3], Christoph R. Jacob*,[4]

Technische Universität Braunschweig, Institute of Physical and Theoretical Chemistry, Gaußstraße 17, 38106 Braunschweig, Germany

Date: November 20, 2024

[1]ORCID: 0009-0005-1559-7422
[2]ORCID: 0009-0009-9079-1769
[3]ORCID: 0000-0001-6948-6801
[4]ORCID: 0000-0002-6227-8476, E-Mail: c.jacob@tu-braunschweig.de

Abstract

Vibrational exciton models are widely used for the simulation of biomolecular vibrational spectra, in particular of two-dimensional infrared (2D-IR) spectra. The parameters entering such models, specifically local-mode frequencies and coupling constants, are provided by vibrational maps, which have been parametrized agains computational data for small molecules as well as experimental data. Here, we put forward a novel approach for assessing the quality of these vibrational maps against quantum-chemical reference data. For a test set consisting of molecular dynamics snapshots of polypeptides and small proteins, covering different secondary structure motifs, we we performed full quantum-chemical calculations of harmonic vibrational frequencies and normal modes, and applied a localization of normal modes to obtain localized-mode frequencies and coupling constants. These can be directly compared to those predicted by vibrational maps. We find that while there is a good correlation for the coupling constants and for local-mode frequencies of isolated polypeptides, there is hardly any correlation for the local-mode frequencies of solvated polypeptides. This striking finding calls into question the accuracy of the electrostatic maps that are used to model the effect of the solvent molecules on local-mode frequencies.

1 Introduction

Unraveling the structure and dynamics of biomolecular systems requires spectroscopic methods that are able to probe fast processes. Vibrational spectroscopy is an attractive tool for such purposes, as — in contrast to X-ray diffraction or NMR spectroscopy — it potentially offers femtosecond time resolution and is directly applicable to biomolecules in aqueous solution [1–4]. In particular two-dimensional infrared (2D-IR) spectroscopy [5–8] is a versatile spectroscopic method that can probe the secondary and tertiary structure of biomolecules [9–16] and that can be used to obtain a detailed picture of the ultrafast dynamics in proteins [17–24].

However, vibrational spectroscopy does not provide direct structural information. Therefore, the ability to extract structural and dynamical information from vibrational spectra, in particular from 2D-IR spectra, crucially depends on the availability of computational tools [25–27]. Only by predicting vibrational spectra for (models of) biomolecular structures and by carefully comparing predicted and experimental spectra it becomes possible to establish relationships between the underlying molecular structure and spectral features (for examples, see Refs. 28–32), and to extract structural and dynamical information from experimental 2D-IR spectra [27, 33, 34].

While quantum-chemical calculations of vibrational spectra are well established [25], their applicability for large molecular systems is limited by the required computational effort. This particularly holds for biomolecular systems, as the need to account for solvent effects as well as conformational flexibility and dynamical effects further increases the computational burden. Moreover, for 2D-IR spectroscopy a treatment within the harmonic approximation is not sufficient, as 2D-IR spectra directly probe anharmonicities as well as anharmonic couplings. Even though there has been considerable progress in the development of efficient methods for anharmonic theoretical vibrational spec-

troscopy [26, 35–38], quantum-chemical calculations of biomolecular 2D-IR spectra are only just emerging [39, 40].

Instead, the simulation of 2D-IR spectroscopy relies on vibrational exciton models [41], which consider a simplified vibrational Hamiltonian of the form,

$$\hat{H}^{\text{exciton}} = \sum_i \left(\hbar\tilde{\omega}_i - \Delta_i\right) \hat{b}_i^\dagger \hat{b}_i + \sum_{i>j} J_{ij} \left(\hat{b}_i^\dagger \hat{b}_j + \hat{b}_j^\dagger \hat{b}_i\right) - \sum_i \frac{\Delta_i}{2} \hat{b}_i^\dagger \hat{b}_i^\dagger \hat{b}_i \hat{b}_i, \tag{1}$$

to model a set of relevant vibrations. Here, $\hat{b}_i^\dagger$ and $\hat{b}_i$ are bosonic creation and annihilation operators with respect to a basis of local oscillators (i.e., $\hat{b}_i^\dagger$ creates a vibrational quantum for the i-th local oscillator). The above vibrational exciton Hamiltonian contains as parameters the harmonic frequencies of the local oscillators, $\tilde{\omega}_i$, the anharmonic shifts Δ_i of these local oscillators, as well as the couplings J_{ij} between local oscillators. For these parameters, one applies an empirical model (vibrational frequency map), in which the local harmonic frequencies as well as coupling constants are parametrized in terms of the molecular structure and the electric field at the local oscillators. The anharmonic shift is commonly modeled using a fixed value.

For the amide I vibrations in polypeptides and proteins, which have been extensively studied in 2D-IR spectroscopy, such vibrational exciton models as well as the associated vibrational maps for modeling the required parameters, are well established and widely used for modeling 2D-IR spectra [42–47]. The vibrational maps used in such simulations have been carefully constructed using quantum-chemical calculations for small model systems, such as N-methylacetamide (NMA) as well as dipeptides [42, 43] or by fitting to experimental data [44, 45].

Vibrational frequency maps for the amide I region are commonly validated by comparison to experimental infrared and 2D-IR spectra, either for small model systems such as NMA [45, 48–50], or by applying them in the calculation of 2D-IR spectra of polypeptides of proteins [44, 46, 51–54]. While they generally do show a good accuracy in such a com-

parison, such an approach is rather indirect and prone to error compensation. Spectra simulated using vibrational exciton models depend on the interplay of all different parts of the vibrational frequency map (such as the different contributions to the local frequencies as well as the coupling constants) simultaneously, and it is not possible to trace back differences between simulated and experimental spectra to specific components of the model. Moreover, the simulation of vibrational spectra and their comparison with experiments in solution additionally depends on molecular dynamics simulations, which require an empirical force field, and errors due to these simulations as well as a mismatch between the conditions in the simulation and in experiment might further mask (or amplify) errors of the vibrational frequency map.

Quantum-chemical calculations for polypeptides and small proteins provide a way of accessing (harmonic) vibrational frequencies and normal modes for specific molecular structures. By performing a localization of the relevant normal modes (i.e., the amide I vibrations) [55], one can reconstruct a picture in terms of the local oscillators that form the basis of vibrational exciton models. This local-mode picture can be directly related to vibrational exciton models [39,56,57] and allows one to extract local-mode frequencies as well as coupling constants. These, in turn, can be directly compared to those predicted by vibrational frequency maps.

Here, we apply such a strategy to asses to accuracy of commonly used vibrational frequency maps for the amide I vibrations by comparing their local-mode harmonic frequencies as well as coupling constants to those extracted from quantum-chemical calculations for a test set of polypeptides and small proteins.

2 Theoretical Background

Starting point for the quantum-chemical calculation of vibrational spectra is usually the harmonic approximation, in which the normal modes and vibrational frequencies can be obtained by diagonalizing the mass-weighted molecular Hessian matrix,

$$H^{(m)}_{I\alpha,J\beta} = \frac{1}{\sqrt{m_I m_J}} \frac{\partial^2 E}{\partial R_{I\alpha} \partial R_{J\beta}}. \tag{2}$$

Here, $R_{I\alpha}$ is the $\alpha = x, y, z$ Cartesian component of the position of the I-th nucleus and m_I is its atomic mass. By diagonalizing the mass-weighted Hessian, one obtains the eigenvalues ω_p^2, which correspond to the squared angular frequencies of the harmonic vibrations, as well as the corresponding normal modes $\boldsymbol{Q}_p$.

In extended systems such as polypeptides, the normal modes are in general delocalized over the whole system. For the amide I band, the normal modes consist of combinations of the C=O stretching vibrations of several peptide groups. To reconstruct a simpler picture in terms of localized vibrations, one collects the subsets of modes that contribute to the vibrational band of interest (e.g., the amide I band), $\boldsymbol{Q}^{\text{sub}}$, and performs a unitary transformation [55],

$$\tilde{\boldsymbol{Q}}^{\text{sub}} = \boldsymbol{Q}^{\text{sub}} \boldsymbol{U}, \tag{3}$$

such that the resulting modes $\tilde{\boldsymbol{Q}}^{\text{sub}}$ are "as localized as possible". To measure the locality of the resulting modes, we apply the atomic-contribution criterion introduced in Ref. 55. Localized vibrational modes on the one hand provide a valuable tool for the analysis of calculated vibrational spectra of polypeptides [29, 30, 32, 58–60] and polymers [61], and can serve as a starting point for the efficient calculation of anharmonic vibrational spectra [35, 62–65], including the quantum-chemical calculation of 2D-IR spectra [39, 40, 56].

In fact, vibrational exciton models are based on a similar picture: They assume that the delocalized vibrations in extended systems emerge from coupled local oscillators. In the

case of the harmonic approximation, the exciton Hamiltonian of Eq. (1) reduces to,

$$\hat{H}^{\text{exciton}} = \sum_i \hbar\tilde{\omega}_i \, \hat{b}_i^\dagger \hat{b}_i + \sum_{i>j} J_{ij} \left(\hat{b}_i^\dagger \hat{b}_j + \hat{b}_j^\dagger \hat{b}_i \right), \tag{4}$$

and its eigenvalues are obtained by diagonalizing the matrix $\mathbf{H}^{\text{exiton}}$ that contains the local-mode energies $\hbar\omega_i$ on its diagonal and the coupling constants J_{ij} as off-diagonal elements.

The transformation matrix from normal modes to local modes, $\boldsymbol{U}$ can be used to construct a coupling matrix,

$$\tilde{\boldsymbol{\Omega}} = \boldsymbol{U}^T \boldsymbol{\Omega} \boldsymbol{U}, \tag{5}$$

where $\boldsymbol{\Omega}$ is a diagonal matrix with the normal-mode vibrational energies $\hbar\omega_p$ of the considered normal modes on its diagonal. By construction, the diagonalization of this coupling matrix recovers the original normal modes as well as the corresponding vibrational energies and frequencies. Here, we postulate that the matrix $\mathbf{H}^{\text{exciton}}$ underlying vibrational exciton models and the coupling matrix $\tilde{\boldsymbol{\Omega}}$ that is accessible from quantum-chemical calculations via a localization of normal modes, are identical, i.e.,

$$\mathbf{H}^{\text{exciton}} \equiv \tilde{\boldsymbol{\Omega}}. \tag{6}$$

This opens up the possibility to obtain quantum-chemical reference values for the local-mode frequencies as well as their coupling constants to those predicted by vibrational frequency maps. For a more detailed discussion of the connection between a localized-mode picture and vibrational exciton models that also goes beyond the harmonic approximation, we refer to Ref. 57.

Vibrational frequency maps [41, 47] employ an empirical parametrization of the local-mode frequencies $\tilde{\omega}_i$ and the coupling constants J_{ij} that enter the vibrational exciton Hamiltonian given in Eq. (1). For the local-mode frequencies, one commonly uses a parametrization of the form

$$\tilde{\omega}_i = \omega_{\text{ref}} + \Delta\omega_{\text{NN}}(\phi_{i,i-1}, \psi_{i,i-1}, \phi_{i,i+1}, \psi_{i,i+1}) + \Delta\omega_{\text{env}}\big(\boldsymbol{E}'_i(\boldsymbol{r}), \nabla \boldsymbol{E}'_i(\boldsymbol{r})\big), \tag{7}$$

where the gas-phase frequency of NMA ω_{ref} is used as reference point. The shift due to the nearest neighbors $\Delta\omega_{\mathrm{NN}}$ is expressed as a function of the backbone dihedral angles ϕ_i and ψ_i as derived from quantum-chemical calculations on N-acetyl-glycine-N'-methylamide (glycine dipeptide, GLDP) [42]. All three vibrational frequency maps considered here use the GLDP parametrization for the nearest-neighbor shifts, and express these as

$$\Delta\omega_{\mathrm{NN}}(\phi_{i,i-1},\psi_{i,i-1},\phi_{i,i+1},\psi_{i,i+1}) = \Delta\omega_{\mathrm{N}}(\phi_{i,i-1},\psi_{i,i-1}) + \Delta\omega_{\mathrm{C}}(\phi_{i,i+1},\psi_{i,i+1}), \tag{8}$$

where $\Delta\omega_{\mathrm{N}}(\phi_{i,i-1},\psi_{i,i-1})$ and $\Delta\omega_{\mathrm{C}}(\phi_{i,i+1},\psi_{i,i+1})$ are obtained by interpolation between the frequency shifts calculated in a two-dimensional scan of ϕ and ψ for GLDP in Ref. 42.

The shift due to the electrostatic environment $\Delta\omega_{\mathrm{env}}$ is parametrized in terms of the electric field (and possibly its gradient) at the atoms of the peptide group. Here, we consider three widely used electrostatic maps, namely those by Jansen [43], Skinner [44], and Tokmakoff [45].

The Jansen map employs a linear model in terms of the electric field $\boldsymbol{E}$ and the electric field gradient $\nabla\boldsymbol{E}$ at the positions of the C, O, N, and H atoms of the peptide group, i.e.,

$$\Delta\omega_{\mathrm{env}}^{\mathrm{Jansen}} = \sum_{A=\mathrm{C,O,N,H}} \left[\sum_{\alpha=x,y} c_{A,\alpha} E_\alpha(\boldsymbol{r}_A) + c_{A,xx}\nabla_z E_z(\boldsymbol{r}_A) + c_{A,xy}\nabla_z\nabla_x E_y(\boldsymbol{r}_A) \right], \tag{9}$$

where the local coordinate system is chosen such that the x-axis is aligned with the C=O bond and that the peptide group lies in the xy-plane. The coefficients are determined by a fit to calculations for NMA, and and can be found in Ref. 43.

The Skinner and Tokmakoff maps have been determined by fitting to selected experimental data, and include only the x-component of the electric field at the position of the carbon and nitrogen atoms and at the oxygen atom, respectively, i.e.,

$$\Delta\omega_{\mathrm{env}}^{\mathrm{Skinner}} = c_{\mathrm{C,xx}}^{\mathrm{Skinner}} E_x(\boldsymbol{r}_C) + c_{\mathrm{N,xx}}^{\mathrm{Skinner}} E_x(\boldsymbol{r}_N) \tag{10}$$

and

$$\Delta\omega_{\mathrm{env}}^{\mathrm{Tokmakoff}} = c_{\mathrm{O,xx}}^{\mathrm{Skinner}} E_x(\boldsymbol{r}_O). \tag{11}$$

For all three maps, the electric field of the environment is modeled via the charged WHAT? of a classical force field (see also Section "Computational Details"). Finally, we note that each of the three maps also uses a different reference frequency ω_{ref} as its starting point.

For the coupling constants, the nearest-neighbor coupling constants are modeled explicitly using the GLDP map as function of the backbone dihedral angles between the two peptide groups, i.e.,

$$J_{i,i+i} = J^{\mathrm{GLDP}}(\phi_{i,i+1}, \psi_{i,i+1}), \tag{12}$$

where J^{GLDP} is calculated by interpolating between the values calculated for GLDP in Ref. 42. For the remaining couplings, one can can employ a transition dipole coupling (TDC) model, in which these coupling constants are calculated as

$$J_{ij} = \frac{\boldsymbol{\mu}_i \cdot \boldsymbol{\mu}_j}{|\boldsymbol{r}_{ij}|^3} - \frac{3(\boldsymbol{\mu}_i \cdot \boldsymbol{r}_{ij})(\boldsymbol{r}_{ij} \cdot \boldsymbol{\mu}_j)}{|\boldsymbol{r}_{ij}|^5}. \tag{13}$$

Here, we only consider the model by Torii and Tasumi [66], in which $\boldsymbol{r}_{ij}$ is a vector pointing from one peptide unit to the other (using a suitable definition of the center of each peptide group in terms of the atomic positions of the C, O, N, and H atoms), and where $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_j$ are local transition dipole moments, for which another parametrization in terms of the geometry of the peptide group is used.

3 Results and Discussion

3.1 Test set and reference data

To assess the accuracy of vibrational exciton models, we constructed a test set of six polypeptides that cover a range of different secondary structure elements (see Fig. 1).

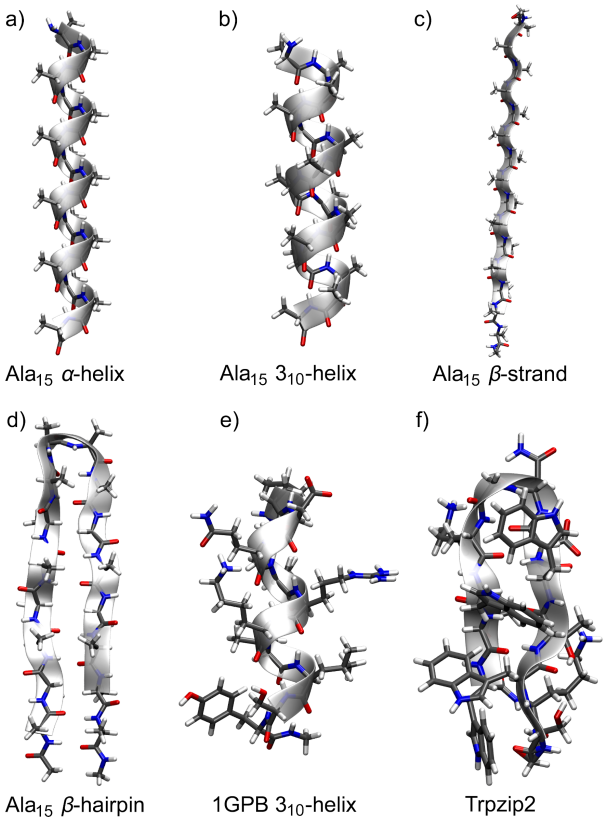

Figure 1: Molecular structures of the six considered polypeptides and proteins. (a) $(\mathrm{Ala})_{15}$ α-helix, (b) $(\mathrm{Ala})_{15}$ 3_{10}-helix, (c) $(\mathrm{Ala})_{15}$ β-strand, (d) $(\mathrm{Ala})_{16}$ β-hairpin, (e) 3_{10}-helix from 1GPB, and (f) Trpzip2. Shown are the initial structures that were the starting point of the MD simulations used to generate structural snapshots.

Our test set contains three polypeptides consisting of 15 alanine units, creating an α-helix, a 3_{10}-helix, and a β-strand as typical secondary structural motifs. In addition, a model of a β-hairpin consisting of 16 alanine units is considered. Finally, we included a 3_{10}-helix cut-out from glycogen phosphorylase B (1GPB) and the tryptophan zipper 2 (Trpzip2) protein, which forms a β-hairpin. The latter two test cases consist of 10 and 12 amino acids, respectively, and both contain two positively and one negatively charged amino acid. With a larger variety of amino acid side chains and their positive total charge, they make the test set significantly more diverse. All structures in our test set have been capped at their C- and N-termini, where the N-terminus is bound to an acetyl group (ACE) and a N-methyl amid group (NME) is added to the C-terminus.

For each of these six structures, classical molecular dynamic (MD) simulations were performed for the polypeptides solvated in water, and from each of these MD trajectories, eleven snapshots were extracted. For both the isolated polypeptide structures extracted from these snapshots as well as for the polypeptides solvated in water, partial geometry optimizations were performed using density-functional theory (DFT). For the solvated polypeptides, a simple QM/MM setup in which explicit water molecules contribute classically was employed. Details on the molecular dynamics and the subsequent quantum-chemical calculations are given in the Section "Computational Details" at the end of this article. This resulted in eleven structures per test case, except for the $(Ala)_{15}$ β-strand and Trpzip2, where the geometry optimizations failed for one snapshot and thus only ten structures could be optimized.

For each of these structures, the molecular Hessians were calculated, in which the relevant amide I modes were identified and subjected to a localization, which provides local modes (which can be assigned to the individual peptide groups), local mode frequencies, and coupling constants between local modes (see Section 2 for details). Overall, the resulting test set obtained from our quantum-chemical calculations consists of 822 amide I local mode frequencies, 758 nearest-neighbor coupling constants, 694 second neighbor coupling constants. The reference full data set containing all optimized molecular structures as well as local-mode frequencies and coupling constants extracted from the quantum-chemical calculations is available at Ref. 67.

For comparison, the local mode frequencies and coupling constants were calculated using vibrational exciton models. In particular, we applied three different vibrational frequency maps (Jansen [43], Skinner [44], and Tokmakoff [45]) in combination with the GLDP map for nearest-neighbor shifts [42]. For the coupling constants, we employ Jansen's GLDP map [42] for the nearest-neighbor couplings and the TDC model of Torii and Tasumi [66] for other couplings (see Section 2 for details). These vibrational maps were

applied as implemented in the AIM program [47, 68]. Further details are given in the Section "Computational Details" at the end of this article.

3.2 Local mode frequencies for isolated polypeptides

First, we consider the isolated polypeptide structures and compare the local-mode frequencies calculated using DFT with those predicted by vibrational frequency maps. Fig. 2 visualizes this comparison for each of the six test cases and for the three considered frequency maps. Each panel plots the reference value for the local-mode frequencies calculated using DFT on the horizontal axis, and the corresponding local-mode frequency predicted by the vibrational map on the vertical axis. For each test case, all snapshots are included, and the colors of the dots are used to indicate the position of the corresponding peptide unit in the polypeptide chain. The local modes corresponding to the terminal capping groups have been excluded in the plots. Some outliers, mainly found for Trpzip2, might be outside of the plotted range.

Table I: Mean error $\mathrm{ME}^{\mathrm{global}}$ [in cm^{-1}] over all polypeptide structures in the test sets of isolated and solvated polypeptides, respectively, and for the three considered vibrational frequency maps. The mean values listed here are used to shift the local-mode vibrational frequencies predicted by these maps.

	Jansen	Skinner	Tokmakoff
isolated	+32.9	+0.8	+34.4
solvated	+14.6	−19.4	+16.3

To aid the visual comparison of the local-mode frequencies and to remove systematic bias, we have shifted the local-mode frequencies predicted by each vibrational maps by the mean error calculated over all six structures (see Table I), i.e., $\omega_i^{\mathrm{map,shifted}} = \omega_i^{\mathrm{map}} - \mathrm{ME}^{\mathrm{global}}$, where ω_i^{map} is the local-mode frequency predicted by the vibrational map, and $\mathrm{ME}^{\mathrm{global}}$ is

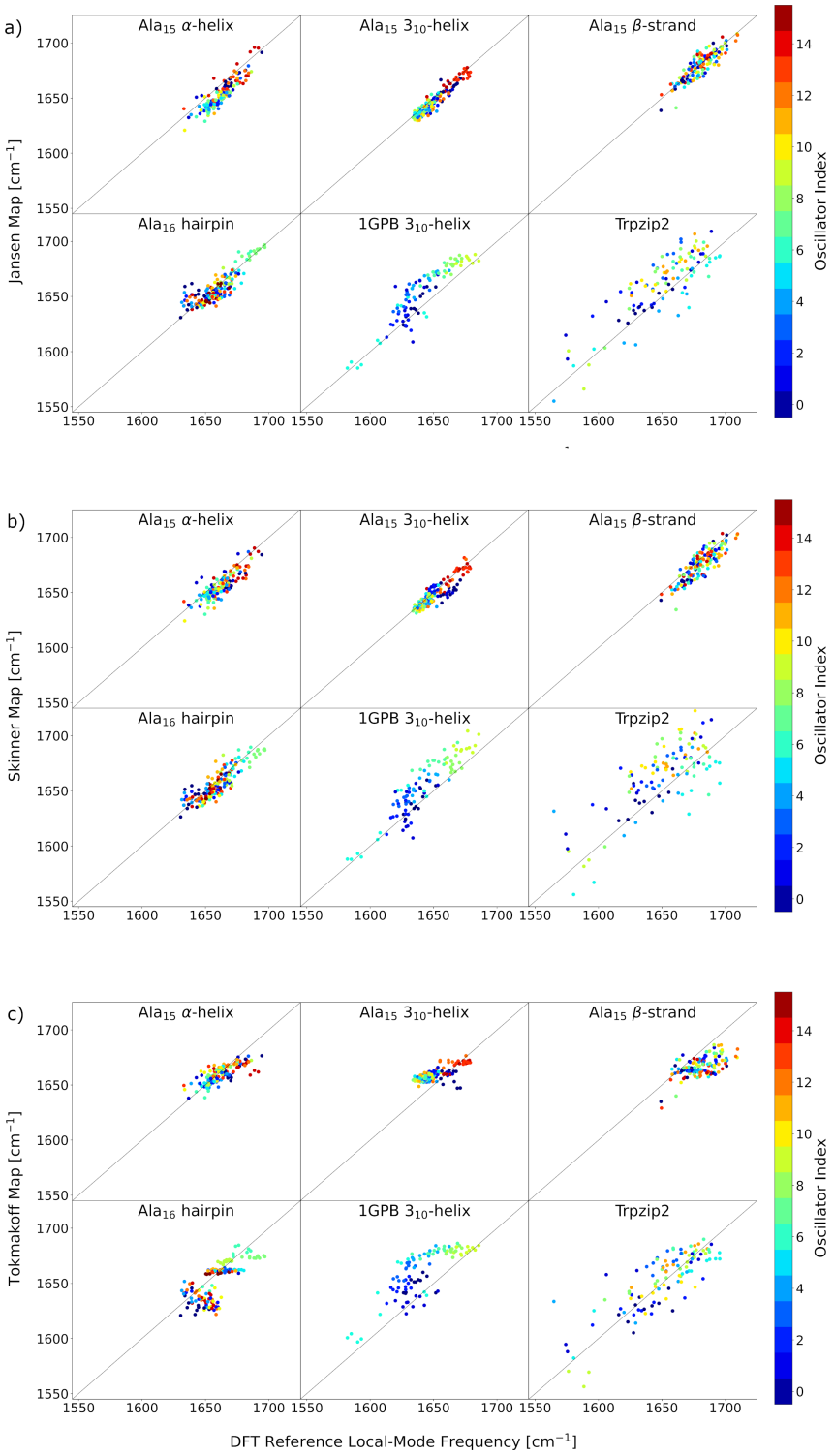

Figure 2: Comparison of the amide I local-mode frequencies [in cm^{-1}] predicted by the (a) Jansen, (b) Skinner, and (c) Tokmakoff vibrational map, $\omega_i^{\mathrm{map,shifted}}$ (shown on the vertical axis) and the reference values extracted from quantum-chemical calculations via a localization of normal modes (shown on the horizontal axis) for **isolated polypeptides**. Each panel shows the local-mode frequencies for all considered structural snapshots of one test case. The color code indicates the position of the peptide group along the chain.

the mean error calculated over the full test set. For a quantitative comparison, Table II further lists the mean errors,

$$\mathrm{ME} = \frac{1}{N}\sum_{i=1}^{N}\left(\omega_i^{\mathrm{map,shifted}} - \omega_i^{\mathrm{ref}}\right), \tag{14}$$

and the root mean square deviation (RMSD),

$$\mathrm{RMSD} = \frac{1}{N}\sqrt{\sum_{i=1}^{N}\left(\omega_i^{\mathrm{map,shifted}} - \omega_i^{\mathrm{ref}} - \mathrm{ME}\right)^2} \tag{15}$$

for each of the considered structures separately. Here, the sums run over the N local modes included in the test set for each structure, $\omega_i^{\mathrm{map,shifted}}$ are the (shifted) local-mode frequencies predicted by the vibrational frequency map, and ω_i^{ref} are the corresponding reference values obtained from the quantum-chemical calculation via the localization of the normal modes. If the prediction of the vibrational map were perfect (except for a systematic shift), all dots in the panels of Fig. 2 would lie on the diagonal. The ME is visible as a shift of all dots parallel to the diagonal, whereas the RMSD is visible as the spread of the dots around this shifted diagonal.

To put the size of the RMSD into perspective, Table. I also contains the RMSD of the reference frequencies,

$$\mathrm{RMSD_{Ref}} = \frac{1}{N}\sqrt{\sum_{i=1}^{N}\left(\omega_i^{\mathrm{ref}} - \bar{\omega}_i^{\mathrm{ref}}\right)^2} \tag{16}$$

with the average of the reference frequencies, $\bar{\omega}_i^{\mathrm{ref}} = \frac{1}{N}\sum_{i=1}^{N}\omega_i^{\mathrm{ref}}$. Moreover, we have calculated to coefficient of determination,

$$R^2 = 1 - \frac{(\mathrm{RMSD})^2}{(\mathrm{RMSD_{Ref}})^2}. \tag{17}$$

For the Jansen model (Fig. 2a), there is a very good correlation between predicted and reference local-mode frequencies for the four polyalanine models. For the α-helix, there is a systematic shift of ca. 9 cm^{-1}, while for the other three models, the ME is below

Table II: Mean error (ME), root mean square deviation (RMSD), and coefficient of determination (R^2) of the amide I local-mode frequencies of **isolated polypeptides** predicted by the vibrational frequency maps of Jansen, Skinner, and Tokmakoff, in comparison to reference values extracted from quantum-chemical calculations via a localization of the amide I normal modes. For comparison, the $\mathrm{RMSD_{Ref}}$ of the reference values is also included. All values are given in $\mathrm{cm^{-1}}$.

		Jansen			Skinner			Tokmakoff		
	$\mathrm{RMSD_{Ref}}$	ME	RMSD	R^2	ME	RMSD	R^2	ME	RMSD	R^2
$\mathrm{Ala_{15}}$ α-helix	11.4	−9.1	5.8	0.96	−6.1	6.1	0.96	−2.3	7.3	0.94
$\mathrm{Ala_{15}}$ 3_{10}-helix	11.5	−3.2	3.8	0.98	−2.7	5.1	0.97	+8.7	8.3	0.93
$\mathrm{Ala_{15}}$ β-strand	12.0	−0.7	6.1	0.96	−5.1	6.8	0.95	−14.4	9.9	0.89
$\mathrm{Ala_{16}}$ β-hairpin	14.1	−1.6	7.7	0.94	−3.0	7.3	0.94	−6.6	11.5	0.86
1GPB 3_{10}-helix	21.4	+11.7	10.6	0.88	+11.3	11.2	0.86	+18.9	13.6	0.80
Trpzip2	30.6	+10.0	18.1	0.65	+13.1	20.7	0.54	+2.3	17.0	0.69

3.2 $\mathrm{cm^{-1}}$. In all four cases, the RMSD is comparable and amounts to 3.8–7.7 $\mathrm{cm^{-1}}$, and the coefficient of determination is larger than 0.9 in all four cases. We note that the color coding in the figure nicely highlights that local modes in similar positions in the polypeptide chain appear at similar frequencies. For instance, for the α-helix and the 3_{10}-helix, the local-mode frequencies for the peptide close to both ends of the chain are shifted to higher wavenumbers, while those in the middle of the chain appear at lower wavenumbers. For the β-hairpin, the local-modes that lie within the turn are shifted to particularly high wavenumbers.

The remaining two test cases, the 1GPB 3_{10}-helix and Trpzip2, feature a large variety of amino acids in their protein chains. For both polypeptides, the local-mode frequencies are spread over a larger range compared to the polyalanine models. While the correlation between predicted and reference local-mode frequencies becomes worse, it still remains

reasonable. For both models, there is a systematic shift of ca. 10–12 cm^{-1}, and the RMSD amounts to 10.6 cm^{-1} for 1GPB and to 18.1 cm^{-1} for Trpzip2. While for 1GPB, R^2 is still as large as 0.88, it drops to only 0.65 for Trpzip2.

For the Skinner model (Fig. 2b), we find a similar picture. For the four polyalanine models, there is a very good correlation between the predicted and the reference frequencies. The MEs amount to between -2.7 and -6.1 cm^{-1}, and the RMSDs between 5.1 and 7.3 cm^{-1} are very similar to those of the Jansen model. Again, the coefficient of determination is larger than 0.9 for all four polypeptide models. Also for the 1GPB 3_{10}-helix and Trpzip2, the results are rather similar to those of the Jansen model, with slightly larger RMSDs and slightly lower R^2 of 0.86 and 0.54, respectively.

For the Tokmakoff model (Fig. 2c), the predictions of the vibrational map cover a smaller range of wavenumbers. For the four polyalanine models, the agreement between predicted and reference local-mode frequencies becomes significantly worse, and in the plots, they appear to correlate with a slope different from one. The systematic shifts (MEs) now vary between +8.7 and -14.4 cm^{-1}, and the RMSDs increase to ca. 7.3–11.5 cm^{-1}. Nevertheless, for the 1GPB 3_{10}-helix and for Trpzip2, the overall correlation is comparable to the other two models, with RMSDs of 13.6 and 17.0 cm^{-1} and R^2 of 0.80 and 0.69, respectively.

Overall, our results show that when considering isolated polypeptides, for the local-mode frequencies all three vibrational maps are in reasonable agreement with the reference values extracted from quantum-chemical calculations. For the Tokmakoff model, the correlation between predicted frequencies and the reference values is slightly worse, in particular for the polyalanine test cases. For all models, the performance is worse for the more realistic test cases that feature different amino acids than for the polyalanine models.

3.3 Local mode frequencies for solvated polypeptides

Next, we turn to the models of polypeptides solvated in water, and again compare the local-mode frequencies predicted by the Jansen, Skinner, and Tokmakoff maps to reference values extracted from quantum-chemical calculations. Note that because in the quantum-chemical calculations, the solvent water molecules are described as point charges, the electrostatics of the solvent are described consistently in both the quantum-chemical calculations and when applying the vibrational maps.

Table III: Mean error (ME), root mean square deviation (RMSD), and coefficient of determination (R^2) of the amide I local-mode frequencies of **solvated polypeptides** predicted by the vibrational frequency maps of Jansen, Skinner, and Tokmakoff, in comparison to reference values extracted from quantum-chemical calculations via a localization of the amide I normal modes. For comparison, the $\mathrm{RMSD_{Ref}}$ of the reference values is also included. All values are given in $\mathrm{cm^{-1}}$.

		Jansen			Skinner			Tokmakoff		
	$\mathrm{RMSD_{Ref}}$	ME	RMSD	R^2	ME	RMSD	R^2	ME	RMSD	R^2
$\mathrm{Ala_{15}}$ α-helix	13.1	−3.2	11.7	0.41	−2.7	13.0	0.28	+5.3	17.9	< 0
$\mathrm{Ala_{15}}$ 3_{10}-helix	13.1	+2.8	10.9	0.49	+2.4	12.4	0.34	+16.3	18.8	< 0
$\mathrm{Ala_{15}}$ β-strand	14.7	−14.3	14.8	0.06	−14.1	17.5	< 0	−22.7	22.8	< 0
$\mathrm{Ala_{16}}$ β-hairpin	16.7	−5.6	17.0	< 0	−7.6	17.7	< 0	−11.5	20.5	< 0
1GPB 3_{10}-helix	15.5	+18.5	13.5	0.22	+18.1	15.5	< 0	+24.0	19.9	< 0
Trpzip2	15.3	+10.4	19.5	< 0	+3.5	26.3	< 0	−5.8	20.3	< 0

Fig. 3 plots the correlation for the three vibrational maps and each of the considered test cases. Again, all local-mode frequencies predicted by the maps have been shifted by the mean errors over all structures, which are listed in Table I. The MEs, RMSDs, and R^2 for each of the test cases are given in Table III.

For the Jansen model (see Fig. 3a), we immediately notice that the correlation is signifi-

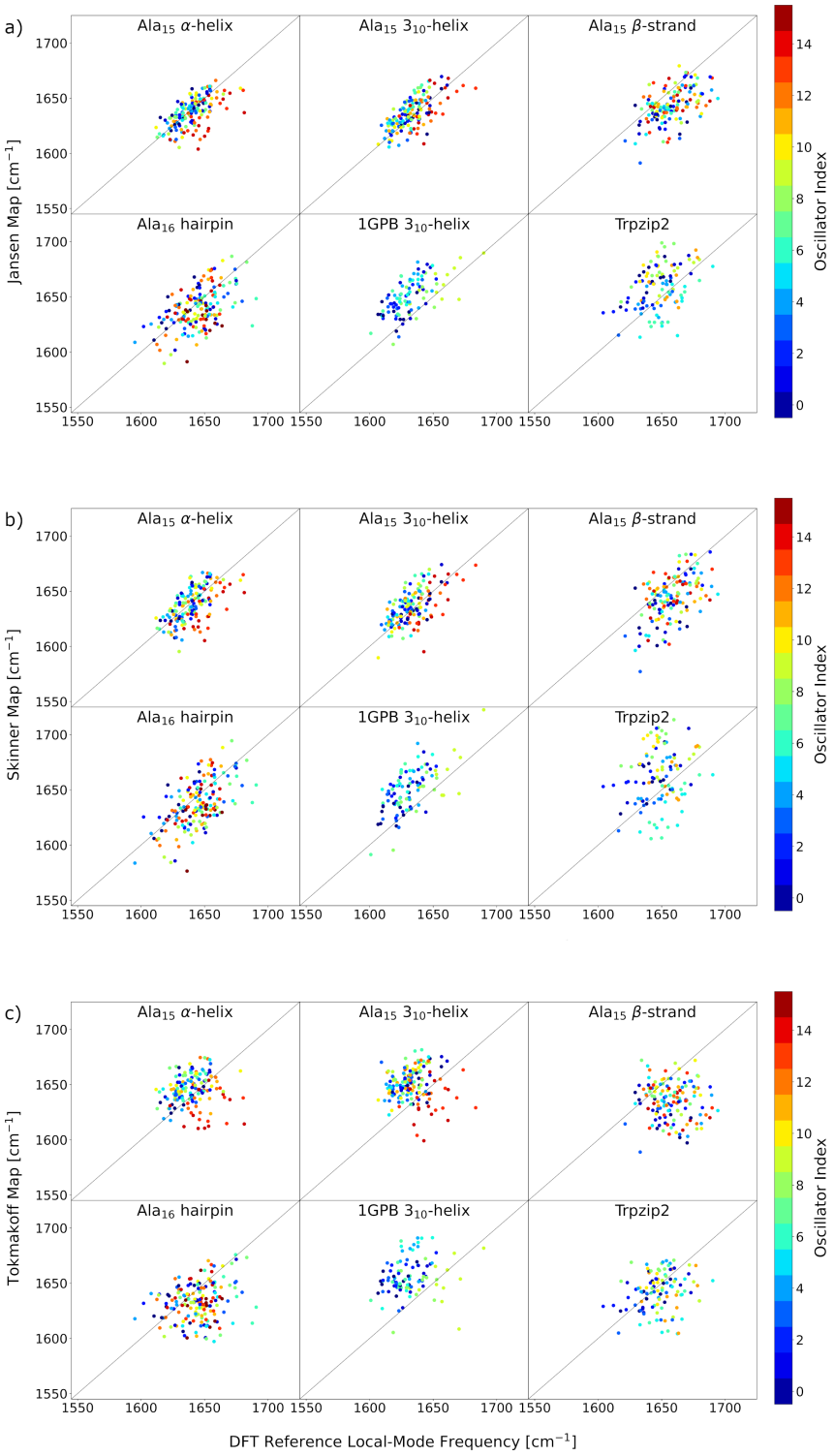

Figure 3: Comparison of the amide I local-mode frequencies [in cm^{-1}] predicted by the (a) Jansen, (b) Skinner, and (c) Tokmakoff vibrational map, $\omega_i^{\mathrm{map,shifted}}$ (shown on the vertical axis) and the reference values extracted from quantum-chemical calculations via a localization of normal modes (shown on the horizontal axis) for **solvated polypeptides**. Each panel shows the local-mode frequencies for all considered structural snapshots of one test case. The color code indicates the position of the peptide group along the chain.

cantly worse compared to the case of the isolated polypeptides. Visually, the scatter of the dots in the plots is drastically increased for the four polyalanine models. This is confirmed by the RMSDs, which increase by at least a factor of two and amount to 10.9–17.0 cm^{-1}. At the same time, R^2 drops below 0.5 in all cases, and becomes as small as 0.06 for the β-strand and even smaller than zero for the β-hairpin. While for the quantum-chemical reference values, some dependence of the local-mode frequencies on the position in the peptide chain is still present (e.g., for the polyalanine α-helix, the local modes close to the C-terminus appear at higher wavenumbers), such a dependence is mostly lost in the predictions of the vibrational map. For the remaining two test cases, the 1GPB 3_{10}-helix and Trpzip2, the RMSDs are only slightly increased compared to the case of isolated polypeptides. However, the local-mode frequencies now cover a smaller range, which makes the overall correlation appear worse in the plots, and results in an R^2 of 0.22 and of below zero, respectively.

For the Skinner model (see Fig. 3b), we obtain a rather similar picture. For the polyalanine models, the RMSDs increase to ca. 13.0–17.7 cm^{-1}, which is at least twice as large as for the isolated polypeptide models. Again, when looking at the plots and at the values obtained for R^2, it appears that all correlation between the reference values and the vibrational maps is lost. Also for the 1GPB 3_{10}-helix and for Trpzip2, there is now a noticeable increase of the RMSD (from 11.2 to 15.5 cm^{-1} and from 15.5 to 26.3 cm^{-1}, respectively) of the RMSDs compared to the isolated polypeptides, and R^2 drops below zero for both test cases.

For the Tokmakoff model, the plots in Fig. 3c indicate hardly any remaining correlation between the local-mode vibrational frequencies predicted by the map and the reference values provided by quantum-chemical calculation. For all six considered structures, the RMSD amounts to around 20 cm^{-1}, which is larger to the RMSD of the reference local-mode frequencies themselves. Consequently, R^2 is below zero for all test cases.

While for the isolated polypeptide models, all three considered vibrational maps were in reasonable agreement with the reference values obtained from quantum-chemical calculation, the picture changes dramatically when looking at solvated polypeptides. Here, we find that for the solvated polypeptides, there is hardly any correlation between the local-mode frequencies predicted by the vibrational maps and the reference values extracted from quantum-chemical calculations. This is particularly striking, as in both the quantum-chemical calculations and when evaluating the vibrational maps, the solvent molecules are consistently treated as classical point charges.

3.4 Coupling constants

Besides the local-mode frequencies, the coupling constants between local modes are another essential ingredient of vibrational exciton models. Therefore, we also compare the coupling constants predicted by vibrational maps commonly used in simulations of protein 2D-IR spectra to the reference values extracted from our quantum-chemical calculations.

This comparison is shown in Fig. 4 for both the isolated and the solvated polypeptides. These plots are restricted to the comparison of the nearest-neighbor couplings and of the second neighbor couplings. To avoid complications arising from a phase difference between the local modes in both calculations, we only compare the absolute values of the coupling constants. In contrast to the local-mode frequencies, we do not apply a shift by the global mean error, and instead show the unshifted data. For a quantitative comparison, Tables IV and V list the MEs, RMSDs, and R^2 for each of the considered test cases for the isolated and the solvated polypeptide models, respectively.

Looking at the plots in Fig. 4, one can clearly see the different coupling patterns for different secondary structure elements. For the polyalanine α-helix, the nearest-neighbor couplings are generally larger than the second-neighbor couplings, while for the polyalanine

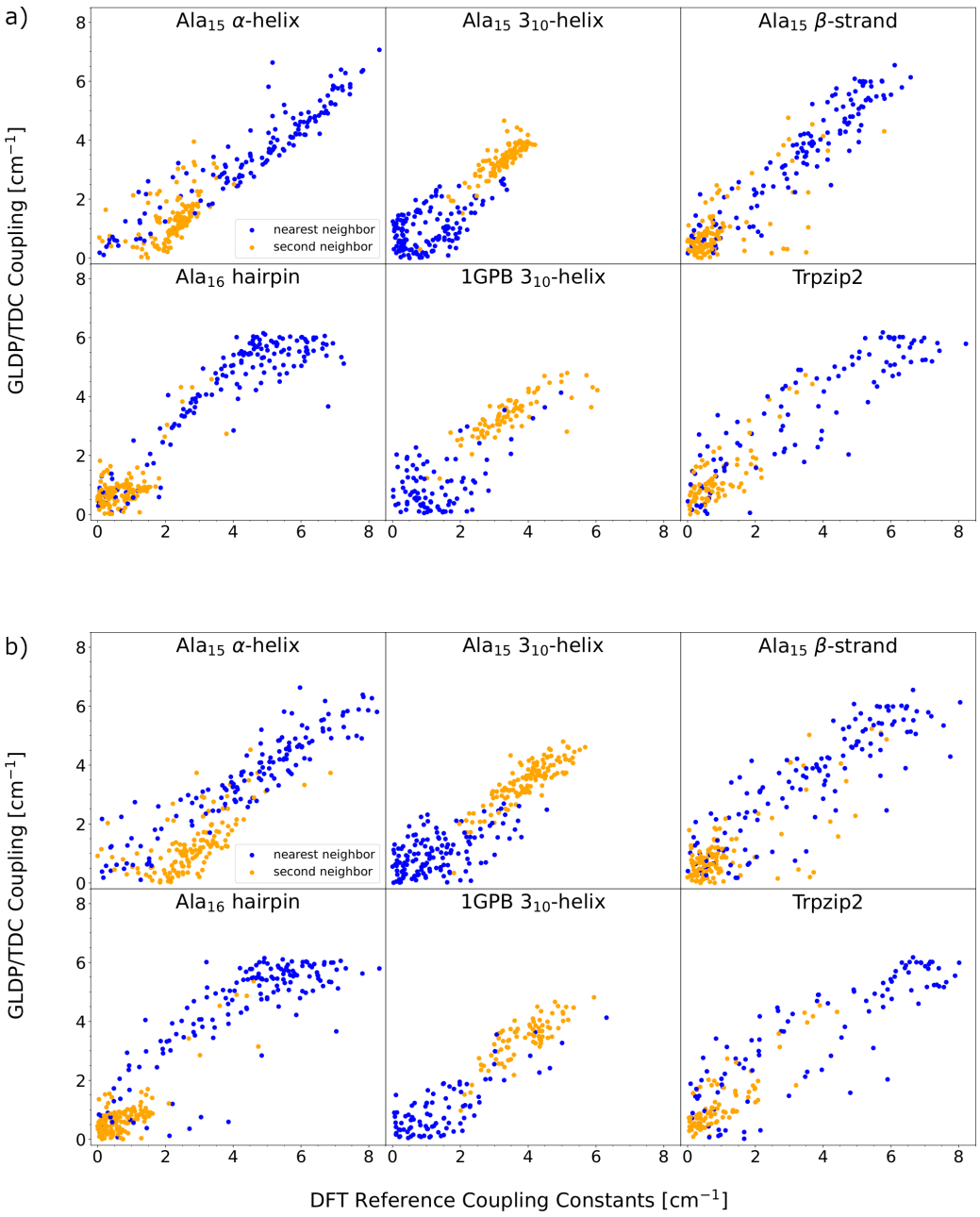

Figure 4: Comparison of the absolute values of the amide I nearest neighbor (blue dots) and second neighbor (orange dots) coupling constants [in cm^{-1}] predicted by the GLDP model for the nearest neighbors and by the TDC model of Torii and Tasumi for all other couplings (shown on the vertical axis) to the reference values extracted from quantum-chemical calculations via a localization of normal modes (shown on the horizontal axis) for (a) isolated polypeptides and (b) solvated polypeptides. Each panel shows the local-mode frequencies for all considered structural snapshots of one test case.

Table IV: Mean error (ME), root mean square deviation (RMSD), and coefficient of determination (R^2) of the absolute values of the amide I local-mode couplings of **isolated polypeptides** predicted with the GLDP model for nearest neighbors and the TDC model of Torii and Tasumi for all other couplings, in comparison to reference values extracted from quantum-chemical calculations via a localization of the amide I normal modes. For comparison, the $\mathrm{RMSD}_{\mathrm{Ref}}$ of the reference values is also included. All values are given in cm^{-1}.

	nearest neighbor				second neighbor			
	$\mathrm{RMSD}_{\mathrm{Ref}}$	ME	RMSD	R^2	$\mathrm{RMSD}_{\mathrm{Ref}}$	ME	RMSD	R^2
Ala_{15} α-helix	2.11	−1.11	0.74	0.92	0.87	−0.92	0.67	0.96
Ala_{15} 3_{10}-helix	0.82	−0.08	0.75	0.91	0.74	+0.07	0.28	0.99
Ala_{15} β-strand	1.79	+0.15	0.61	0.94	2.37	+0.02	0.80	0.94
Ala_{16} β-hairpin	1.89	+0.33	0.89	0.90	1.95	+0.12	0.55	0.97
1GPB 3_{10}-helix	1.01	−0.37	0.96	0.86	1.69	−0.08	0.59	0.97
Trpzip2	2.56	−0.28	1.50	0.65	3.29	+0.33	0.55	0.97

and 1GPB 3_{10}-helices, this pattern is reversed. For the polyalanine β-strand, the polyalanine β-hairpin, and Trpzip2, the second neighbor couplings are significantly smaller than for the helices.

Overall, for all considered test cases we find a good correlation between the coupling constants predicted by the GLDP and TDC models and the quantum-chemical reference values. For the isolated polypeptides (see Fig. 4a and Table IV), we notice that there are some distinct systematic errors for the different test cases. For the α-helix, all couplings are slightly underestimated by the vibrational maps, which is reflected by the MEs of -1.11 and -0.92 for the nearest neighbor and second neighbor couplings, respectively. For the polyalanine β-hairpin, the nearest neighbor couplings between 2 and 5 cm^{-1} appear to be systematically overestimated, while for the polyalanine 3_{10}-helix, they feature a bimodal distribution in which some couplings are systematically underestimated by the

Table V: Mean error (ME), root mean square deviation (RMSD), and coefficient of determination (R^2) of the absolute values of the amide I local-mode couplings of **solvated polypeptides** predicted with the GLDP model for nearest neighbors and the TDC model of Torii and Tasumi for all other couplings, in comparison to reference values extracted from quantum-chemical calculations via a localization of the amide I normal modes. For comparison, the $\mathrm{RMSD_{Ref}}$ of the reference values is also included. All values are given in $\mathrm{cm^{-1}}$.

	nearest neighbor				second neighbor			
	$\mathrm{RMSD_{Ref}}$	ME	RMSD	R^2	$\mathrm{RMSD_{Ref}}$	ME	RMSD	R^2
$\mathrm{Ala_{15}}$ α-helix	2.21	−0.96	0.85	0.90	1.06	−1.47	0.82	0.94
$\mathrm{Ala_{15}}$ 3_{10}-helix	1.01	−0.26	0.77	0.91	1.30	−0.50	0.40	0.99
$\mathrm{Ala_{15}}$ β-strand	2.15	−0.12	1.01	0.85	2.44	−0.09	0.87	0.94
$\mathrm{Ala_{16}}$ β-hairpin	2.02	+0.12	1.11	0.82	2.06	−0.02	0.47	0.98
1GPB 3_{10}-helix	1.22	−0.45	0.77	0.91	1.40	−0.45	0.46	0.98
Trpzip2	2.64	−0.47	1.35	0.74	3.45	+0.23	0.52	0.98

vibrational map, whereas others are systematically overestimated. Nevertheless, with the exception of the polyalanine α-helix, the systematic shifts reflected by the MEs are below 0.4 $\mathrm{cm^{-1}}$.

Except for Trpzip2, the RMSD of the nearest-neighbor and second-neighbor coupling constants is smaller than 1 $\mathrm{cm^{-1}}$ for all test cases, and the coefficient of determination R^2 is larger than 0.86 and 0.94 for the first and second nearest-neighbor couplings, respectively. For Trpzip2, the RMSD for the nearest-neighbor couplings is slightly larger and amounts to 1.50 $\mathrm{cm^{-1}}$, which leads to a drop of R^2 to 0.65. However, for the second neighbors, the accuracy of the vibrational map remains as good as for the other five test cases also for Trpzip2.

When comparing to the results for the solvated polypeptides (see Fig. 4b and Table V),

we can find the same trends. The results are overall very similar, and the correlation between the coupling constants predicted by the vibrational maps and the quantum-chemical reference values remains very good. For the nearest-neighbor couplings, the coefficient of determination R^2 is larger than 0.82 for all test cases, with the exception of Trpzip2 for which it amounts to only 0.74. For the second-neighbor couplings, R^2 is larger than 0.94 for all six test cases.

Overall, we find a good correlation between the coupling constants predicted by the considered maps, both for the nearest-neighbor and for the second-neighbor couplings, even though we observe some systematic errors for specific secondary structure elements. Our results are comparable for both isolated and solvated polypeptides, which demonstrates that the influence of the electrostatic environment is rather small for the coupling constants.

4 Conclusions and Outlook

We have applied a novel approach for the critical assessment of vibrational maps that are used to parametrize the local-mode frequencies and coupling constants that enter vibrational exciton models, which are in turn widely used for the simulation of biomolecular 2D-IR spectra. We have demonstrated that a localization of normal modes can be used to extract local-mode frequencies and coupling constants from quantum-chemical calculations, which can be directly compared to those predicted by vibrational maps. This makes it possible to assess the accuracy of the different ingredients of vibrational exciton models, which might be masked in a direct comparison of predicted spectra with experimental data.

While quantum-chemical calculations for small model systems have previously been used as an essential tool in the parametrization of vibrational maps, our approach is not re-

stricted to specific models, and thus opens up the possibility of obtaining references for realistic test cases, such as polypeptides and proteins. Here, we have a constructed a test set consisting of MD snapshots of polypeptides and small proteins, and have obtained quantum-chemical reference values for local-mode frequencies and coupling constants in both the isolated and solvated systems.

The comparison of local-mode frequencies and coupling constants predicted by commonly used vibrational maps to this quantum-chemical reference data allows us to draw a number of conclusions concerning the quality of these maps.

First, we find a rather good correlation for the coupling constants, both in the isolated and in the solvated polypeptides. Thus, it appears that these are well described by the existing maps, which are able to predict the distinct coupling patterns of different secondary structure elements. Second, we also find a good correlation for the local-mode frequencies in isolated polypeptides. In this case, the vibrational frequency maps are dominated by the nearest-neighbor shifts, for which all considered maps use the same model.

However, both for the local-mode frequencies in the isolated polypeptide models and for the nearest-neighbor couplings, we notice a that the correlation is worse for test cases beyond polyalanine, i.e., for polypeptides featuring a larger variety of amino acids. This might indicate that there is room for improvement by parametrizing amino-acid-specific nearest-neighbor maps, both for the frequency shifts and for the couplings.

Third, we find that for solvated polypeptides, there is hardly any correlation between the local-mode frequencies predicted by the considered vibrational maps and the quantum-chemical reference data. Only for the Jansen map, there is some correlation for the helical polyalanine test cases. This is in stark contrast to the situation for the isolated polypeptides, and indicates that the electrostatic maps are not able to model the effect

of the electrostatic environment of the solvent on the local-mode frequencies. Thus, we believe there is a need to revisit these electrostatic maps.

The poor performance of the electrostatic maps in comparison to quantum-chemical reference data is particularly striking, as these maps are widely used for the simulation of 2D-IR spectra, and have shown good agreement with experimental spectra in many studies. This might indicate that the local-mode frequencies (and their change due to the solvent environment) are actually less important for the prediction of 2D-IR spectra than the coupling constants.

The quantum-chemical methodology employed here opens the door to a new approach for the construction and validation of vibrational maps. Instead of relying on data for small model systems, it now becomes possible to use reference data for realistic systems (i.e., solvated polypeptides and proteins). We plan to further explore this avenue in our future work.

With respect to the electrostatic maps, our results suggest that a critical reassessment of the existing approaches for the construction of electrostatic maps is necessary. To this end, we suggest two complementary lines of research. First, it might be warranted to reinvestigate the functional form underlying electrostatic maps, and possibly go beyond a linear dependence on the local electric field and field gradient, and to assess the dependence of the electrostatic shifts on local mode parameters in more details.

Second, one can pursue a data-driven approach by employing the methodology put forward in this work to generate larger reference data sets, which can then be used to parametrize vibrational maps using machine learning [69–71]. Work along both of these lines is currently in progress in our research group.

Computational Details

The starting molecular structures for the four polyalanine test cases were constructed using VMD [72] with the Molefacture plugin by using pre-defined backbone angles. For the 3_{10}-helix cutout from glycogen phosphorylase B, we used residues 515 to 524 from the PDB structure (PDB-Code: 1GPB), and for the tryptophan zipper 2, we also used the PDB structure as starting point (PDB-Code: 1LE1) All structures were capped with acetyl group (ACE) at the N-termini and an N-methyl group at the C-termini.

For each structure a multi-stage equilibration and optimization procedure was employed. First, a molecular dynamic (MD) simulation was performed using Gromacs [73–75] and the AMBER99SB-ILDN force field [76] and the TIP3P water model [77]. The padding of the cuboid water box around the peptide was set to 1 nm, resulting in simulation boxes containing 2000–4000 water molecules depending on the aspect ratio of the peptide. If necessary, the net charge was equalized by replacing one of the water molecules with a chloride anion. All structures were equilibrated at room temperature and standard pressure. As no restraints were applied to the initial idealized structures, and the polypeptides changed their conformations during the following MD runs. For the alanine α-helix only the first 2.5 ns of the trajectory were used for extracting snapshots, while for all other systems, snapshots were extracted from the whole 10 ns trajectories. For all extracted snapshots, the peptides were re-centered in their water boxes before the quantum-chemical calculations.

Subsequently, the molecular structures of all snapshots were optimized using DFT as implemented in the TURBOMOLE program package [78–80] employing the BP86 exchange–correlation functional [81,82] in combination with the Karlsruhe def2-TZVP basis-set [83], following a pre-optimization with the def2-SVP basis set.

For the solvated polypeptides, we used a simple QM/MM setup with electrostatic em-

bedding, in which the coordinates of the solvent molecules were kept fixed. All solvent molecules were modeled classically as TIP3P point charges [77] with a Lennard-Jones potential. We used atom-specific Lennard-Jones parameters by Freindorf and Gao, which have been optimized for QM/MM geometry optimizations based on the TIP3P model [84]. For the isolated polypeptides, the water environment was deleted from the optimized snapshots of the solvated polypeptides, and the structures were re-optimized in vacuo with constrained backbone dihedral angles. All optimized molecular structures are included in the supplementary data set [67].

For the quantum-chemical calculations of the harmonic vibrational frequencies and normal modes, we used MoViPac/SNF [85, 86] in combination with TURBOMOLE, applying the same setup as for the geometry optimizations (i.e., including the electrostatic potential of the point charges, the corresponding interaction energy, and the interaction with the Lennard-Jones potential. All frequency calculations were performed for the N-deuterated peptides. For the localization of normal modes, we employed the LocVib package [55, 87] with the atomic-contribution localization measure.

For the calculation of the local-mode frequencies and coupling constants using vibrational maps, we applied the AIM program [47, 68], using the maps and parameters as described above. To allow for the treatment of the ACE and NME caps in AIM, we used a locally modified version [88]. For the 1GPB helix and for Trpzip2, the chloride ion was deleted from the corresponding PDB file before the AIM calculation. Input scripts for AIM are included in the supplementary data set [67].

Data Availability

A data set containing all optimized molecular structures of the isolated and solvated polypeptides in our test set, all local-mode frequencies and coupling constants obtained

quantum-chemically and with vibrational maps, input files for SNF and AIM, and Jupyter notebooks for data analysis and for generating all plots and tables contained within this article, are available in Zenodo at DOI: 10.5281/zenodo.14181231 (Ref. 67).

Author Contributions

Anna M. van Bodegraven: Methodology (supporting), Software (supporting), Investigation (equal), Visualization (lead), Data Curation (supporting), Writing – Original Draft (lead), Writing – Review and Editing (equal) **Kevin Focke**: Methodology (supporting), Software (lead), Investigation (equal), Visualization (supporting), Data Curation (lead) Writing – Original Draft (supporting), Writing – Review and Editing (equal) **Mario Wolter**: Conceptualization (supporting), Methodology (lead), Supervision (lead), Software (supporting), Investigation (supporting), Data Curation (supporting), Writing – Original Draft (supporting), Writing – Review and Editing (equal) **Christoph R. Jacob**: Conceptualization (lead), Methodology (supporting), Supervision (supporting), Writing – Original Draft (supporting), Writing – Review and Editing (lead)

Conflicts of Interest

The authors have no conflicts to disclose.

Acknowledgements

Funding from the Deutsche Forschungsgemeinschaft (Project JA 2329/2-2) is gratefully acknowledged. We thank Prof. Thomas La Cour Jansen and Kim E. van Adrichem (University of Groningen) for valuable support with the AIM package.

References

[1] E. Vass, M. Hollosi, F. Besson, and R. Buchet, *Vibrational Spectroscopic Detection of Beta- and Gamma-Turns in Synthetic and Natural Peptides and Proteins*, Chem. Rev. **103**, 1917 (2003).

[2] T. A. Keiderling, Peptide and Protein Conformational Studies with Vibrational Circular Dichroism and Related Spectroscopies, in *Circular Dichroism: Principles and Applications*, edited by N. Berova, K. Nakanishi, and R. W. Woody, pages 621–666, Wiley-VCH, New York, 2nd edition, 2000.

[3] A. Barth, *Infrared spectroscopy of proteins*, Biochim. Biophys. Acta, Bioenerg. **1767**, 1073 (2007).

[4] F. Siebert and P. Hildebrandt, *Vibrational Spectroscopy in Life Science*, Wiley-VCH, Weinheim, 1 edition, 2007.

[5] M. Cho, *Coherent Two-Dimensional Optical Spectroscopy*, Chem. Rev. **108**, 1331 (2008).

[6] N. T. Hunt, *2D-IR spectroscopy: ultrafast insights into biomolecule structure and function*, Chem. Soc. Rev. **38**, 1837 (2009).

[7] P. Hamm and M. Zanni, *Concepts and Methods of 2D Infrared Spectroscopy*, Cambridge University Press, Cambridge, 1st edition, 2011.

[8] M. Cho, editor, *Coherent Multidimensional Spectroscopy*, volume 226 of *Springer Series in Optical Sciences*, Springer, Singapore, 2019, DOI: 10.1007/978-981-13-9753-0.

[9] P. Hamm, M. Lim, and R. M. Hochstrasser, *Structure of the Amide I Band of Peptides Measured by Femtosecond Nonlinear-Infrared Spectroscopy*, J. Phys. Chem. B **102**, 6123 (1998).

[10] C. T. Middleton, P. Marek, P. Cao, C.-c. Chiu, S. Singh, A. M. Woys, J. J. de Pablo, D. P. Raleigh, and M. T. Zanni, *Two-dimensional infrared spectroscopy reveals the complex behaviour of an amyloid fibril inhibitor*, Nat. Chem. **4**, 355 (2012).

[11] A. Remorino and R. M. Hochstrasser, *Three-Dimensional Structures by Two-Dimensional Vibrational Spectroscopy*, Acc. Chem. Res. **45**, 1896 (2012).

[12] C. R. Baiz and A. Tokmakoff, *Structural Disorder of Folded Proteins: Isotope-Edited 2D IR Spectroscopy and Markov State Modeling*, Biophys. J. **108**, 1747 (2015).

[13] A. Huerta-Viga, S. Amirjalayer, S. R. Domingos, H. Meuzelaar, A. Rupenyan, and S. Woutersen, *The structure of salt bridges between Arg+ and Glu- in peptides investigated with 2D-IR spectroscopy: Evidence for two distinct hydrogen-bond geometries*, J. Chem. Phys. **142**, 212444 (2015).

[14] A. Ghosh, J. S. Ostrander, and M. T. Zanni, *Watching Proteins Wiggle: Mapping Structures with Two-Dimensional Infrared Spectroscopy*, Chem. Rev. **117**, 10726 (2017).

[15] M. K. Petti, J. P. Lomont, M. Maj, and M. T. Zanni, *Two-Dimensional Spectroscopy Is Being Used to Address Core Scientific Questions in Biology and Materials Science*, J. Phys. Chem. B **122**, 1771 (2018).

[16] Z. Li, X. Feng, K. Yuan, and X.-X. Zhang, *Identification of Binding Sites in Copper(II)-Peptide Complexes Using Infrared Spectroscopy*, J. Phys. Chem. B **128**, 1884 (2024).

[17] M. C. Thielges and M. D. Fayer, *Protein Dynamics Studied with Ultrafast Two-Dimensional Infrared Vibrational Echo Spectroscopy*, Acc. Chem. Res. **45**, 1866 (2012).

[18] J. K. Chung, M. C. Thielges, and M. D. Fayer, *Conformational Dynamics and Stability of HP35 Studied with 2D IR Vibrational Echoes*, J. Am. Chem. Soc. **134**, 12118 (2012).

[19] K. C. Jones, C. S. Peng, and A. Tokmakoff, *Folding of a heterogeneous ?-hairpin peptide from temperature-jump 2D IR spectroscopy*, Proc. Natl. Acad. Sci. U. S. A. **110**, 2828 (2013).

[20] C.-J. Feng and A. Tokmakoff, *The dynamics of peptide-water interactions in dialanine: An ultrafast amide I 2D IR and computational spectroscopy study*, J. Chem. Phys. **147**, 085101 (2017).

[21] J. M. Schmidt-Engler, L. Blankenburg, R. Zangl, J. Hoffmann, N. Morgner, and J. Bredenbeck, *Local dynamics of the photo-switchable protein PYP in ground and signalling state probed by 2D-IR spectroscopy of –SCN labels*, Phys. Chem. Chem. Phys. **22**, 22963 (2020).

[22] J. M. Schmidt-Engler, R. Zangl, P. Guldan, N. Morgner, and J. Bredenbeck, *Exploring the 2D-IR repertoire of the –SCN label to study site-resolved dynamics and solvation in the calcium sensor protein calmodulin*, Phys. Chem. Chem. Phys. **22**, 5463 (2020).

[23] W. Weng, A. B. Weberg, R. Gera, N. C. Tomson, and J. M. Anna, *Probing Ligand Effects on the Ultrafast Dynamics of Copper Complexes via Midinfrared Pump–Probe and 2DIR Spectroscopies*, J. Phys. Chem. B **125**, 12228 (2021).

[24] S. H. Rutherford, M. J. Baker, and N. T. Hunt, *2D-IR spectroscopy of proteins in H2O—A Perspective*, J. Chem. Phys. **158**, 030901 (2023).

[25] C. Herrmann and M. Reiher, *First-Principles Approach to Vibrational Spectroscopy of Biomolecules*, Top. Curr. Chem. **268**, 85 (2007).

[26] V. Barone, *The virtual multifrequency spectrometer: a new paradigm for spectroscopy*, WIREs Comput. Mol. Sci. **6**, 86 (2016).

[27] M. Kowalewski, B. P. Fingerhut, K. E. Dorfman, K. Bennett, and S. Mukamel, *Simulating Coherent Multidimensional Spectroscopy of Nonadiabatic Molecular Processes: From the Infrared to the X-ray Regime*, Chem. Rev. **117**, 12165 (2017).

[28] Ch. R. Jacob, S. Luber, and M. Reiher, *Calculated Raman Optical Activity Signatures of Tryptophan Side Chains*, ChemPhysChem **9**, 2177 (2008).

[29] Ch. R. Jacob, S. Luber, and M. Reiher, *Analysis of Secondary Structure Effects on the IR and Raman Spectra of Polypeptides in Terms of Localized Vibrations*, J. Phys. Chem. B **113**, 6558 (2009).

[30] Ch. R. Jacob, S. Luber, and M. Reiher, *Understanding the Signatures of Secondary-Structure Elements in Proteins with Raman Optical Activity Spectroscopy*, Chem. Eur. J. **15**, 13491 (2009).

[31] T. Weymuth, Ch. R. Jacob, and M. Reiher, *Identifying Protein β-Turns with Vibrational Raman Optical Activity*, ChemPhysChem **12**, 1165 (2011).

[32] Ch. R. Jacob, *Theoretical Study of the Raman Optical Activity Spectra of 3_{10}-Helical Polypeptides*, ChemPhysChem **12**, 3291 (2011).

[33] W. Zhuang, T. Hayashi, and S. Mukamel, *Coherent Multidimensional Vibrational Spectroscopy of Biomolecules: Concepts, Simulations, and Challenges*, Angew. Chem. Int. Ed. **48**, 3750 (2009).

[34] T. l. C. Jansen, S. Saito, J. Jeon, and M. Cho, *Theory of coherent two-dimensional vibrational spectroscopy*, J. Chem. Phys. **150**, 100901 (2019).

[35] P. T. Panek and Ch. R. Jacob, *Efficient Calculation of Anharmonic Vibrational Spectra of Large Molecules with Localized Modes*, ChemPhysChem **15**, 3365 (2014).

[36] P. T. Panek and Ch. R. Jacob, *Anharmonic Theoretical Vibrational Spectroscopy of Polypeptides*, J. Phys. Chem. Lett. **7**, 3084 (2016).

[37] C. König, *Tailored multilevel approaches in vibrational structure theory: A route to quantum mechanical vibrational spectra for complex systems*, Int. J. Quantum Chem. **121**, e26375 (2021).

[38] A. A. Zhanserkeev, E. L. Yang, and R. P. Steele, *Accelerating Anharmonic Spectroscopy Simulations via Local-Mode, Multilevel Methods*, J. Chem. Theory Comput. **19**, 5572 (2023).

[39] J. Brüggemann, M. Wolter, and Ch. R. Jacob, *Quantum-chemical calculation of two-dimensional infrared spectra using localized-mode VSCF/VCI*, J. Chem. Phys. **157**, 244107 (2022).

[40] J. Brüggemann, M. Chekmeneva, M. Wolter, and Ch. R. Jacob, *Structural Dependence of Extended Amide III Vibrations in Two-Dimensional Infrared Spectra*, J. Phys. Chem. Lett. **14**, 9257 (2023).

[41] C. R. Baiz et al., *Vibrational Spectroscopic Map, Vibrational Spectroscopy, and Intermolecular Interaction*, Chem. Rev. **120**, 7152 (2020).

[42] T. l. C. Jansen, A. G. Dijkstra, T. M. Watson, J. D. Hirst, and J. Knoester, *Modeling the amide I bands of small peptides*, J. Chem. Phys. **125**, 044312 (2006).

[43] T. l. C. Jansen and J. Knoester, *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).

[44] L. Wang, C. T. Middleton, M. T. Zanni, and J. L. Skinner, *Development and Validation of Transferable Amide I Vibrational Frequency Maps for Peptides*, J. Phys. Chem. B **115**, 3713 (2011).

[45] M. Reppert and A. Tokmakoff, *Electrostatic frequency shifts in amide I vibrational spectra: Direct parameterization against experiment*, J. Chem. Phys. **138**, 134116 (2013).

[46] A. V. Cunha, A. S. Bondarenko, and T. L. C. Jansen, *Assessing Spectral Simulation Protocols for the Amide I Band of Proteins*, J. Chem. Theory Comput. **12**, 3982 (2016).

[47] K. E. van Adrichem and T. L. C. Jansen, *AIM: A Mapping Program for Infrared Spectroscopy of Proteins*, J. Chem. Theory Comput. **18**, 3089 (2022).

[48] M. F. DeCamp, L. DeFlores, J. M. McCracken, A. Tokmakoff, K. Kwac, and M. Cho, *Amide I Vibrational Dynamics of N-Methylacetamide in Polar Solvents: The Role of Electrostatic Interactions*, J. Phys. Chem. B **109**, 11016 (2005).

[49] A. V. Cunha, E. Salamatova, R. Bloem, S. J. Roeters, S. Woutersen, M. S. Pshenichnikov, and T. L. C. Jansen, *Interplay between Hydrogen Bonding and Vibrational Coupling in Liquid N-Methylacetamide*, J. Phys. Chem. Lett. **8**, 2438 (2017).

[50] E. Salamatova, A. V. Cunha, R. Bloem, S. J. Roeters, S. Woutersen, T. L. C. Jansen, and M. S. Pshenichnikov, *Hydrophobic Collapse in N-Methylacetamide–Water Mixtures*, J. Phys. Chem. A **122**, 2468 (2018).

[51] J.-H. Choi, H. Lee, K.-K. Lee, S. Hahn, and M. Cho, *Computational spectroscopy of ubiquitin: Comparison between theory and experiments*, J. Chem. Phys. **126**, 045102 (2007).

[52] S. Roy, J. Lessing, G. Meisl, Z. Ganim, A. Tokmakoff, J. Knoester, and T. L. C. Jansen, *Solvent and conformation dependence of amide I vibrations in peptides and proteins containing proline*, J. Chem. Phys. **135**, 234507 (2011).

[53] E. Małolepsza and J. E. Straub, *Empirical Maps For The Calculation of Amide I Vibrational Spectra of Proteins From Classical Molecular Dynamics Simulations*, J. Phys. Chem. B (2014).

[54] A. S. Bondarenko and T. L. C. Jansen, *Application of two-dimensional infrared spectroscopy to benchmark models for the amide I band of proteins*, J. Chem. Phys. **142**, 212437 (2015).

[55] Ch. R. Jacob and M. Reiher, *Localizing normal modes in large molecules*, J. Chem. Phys. **130**, 084106 (2009).

[56] M. W. D. Hanson-Heine, F. S. Husseini, J. D. Hirst, and N. A. Besley, *Simulation of Two-Dimensional Infrared Spectroscopy of Peptides Using Localized Normal Modes*, J. Chem. Theory Comput. **12**, 1905 (2016).

[57] P. T. Panek, A. A. Hoeske, and Ch. R. Jacob, *On the choice of coordinates in anharmonic theoretical vibrational spectroscopy: Harmonic vs. anharmonic coupling in vibrational configuration interaction*, J. Chem. Phys. **150**, 054107 (2019).

[58] T. Weymuth, Ch. R. Jacob, and M. Reiher, *A Local-Mode Model for Understanding the Dependence of the Extended Amide III Vibrations on Protein Secondary Structure*, J. Phys. Chem. B **114**, 10649 (2010).

[59] C. König, M. B. Hansen, I. H. Godtliebsen, and O. Christiansen, *FALCON: A method for flexible adaptation of local coordinates of nuclei*, J. Chem. Phys. **144**, 074108 (2016).

[60] E. Kraka, W. Zou, and Y. Tao, *Decoding chemical information from vibrational spectroscopy data: Local vibrational mode theory*, WIREs Comput. Mol. Sci. (2020).

[61] V. Liégeois, Ch. R. Jacob, B. Champagne, and M. Reiher, *Analysis of Vibrational Raman Optical Activity Signatures of the $(TG)_N$ and $(GG)_N$ Conformations of Iso-*

tactic Polypropylene Chains in Terms of Localized Modes, J. Phys. Chem. A **114**, 7198 (2010).

[62] X. Cheng and R. P. Steele, *Efficient anharmonic vibrational spectroscopy for large molecules using local-mode coordinates*, J. Chem. Phys. **141**, 104105 (2014).

[63] E. L. Klinting, C. König, and O. Christiansen, *Hybrid Optimized and Localized Vibrational Coordinates*, J. Phys. Chem. A **119**, 11007 (2015).

[64] P. T. Panek and Ch. R. Jacob, *On the benefits of localized modes in anharmonic vibrational calculations for small molecules*, J. Chem. Phys. **144**, 164111 (2016).

[65] B. Ziegler and G. Rauhut, *Localized Normal Coordinates in Accurate Vibrational Structure Calculations: Benchmarks for Small Molecules*, J. Chem. Theory Comput. **15**, 4187 (2019).

[66] H. Torii and M. Tasumi, *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 (1998).

[67] A. M. van Bodegraven, K. Focke, M. Wolter, and Ch. R. Jacob, Data Set: Benchmarking of Vibrational Exciton Models Against Quantum-Chemical Localized-Mode Calculations, 2024, DOI: 10.5281/zenodo.14181231.

[68] K. E. van Adrichem and T. L. C. Jansen, AIM, Version 1.0.2, 2023, URL: https://github.com/Kimvana/AIM.

[69] K. Kwac, H. Freedman, and M. Cho, *Machine Learning Approach for Describing Water OH Stretch Vibrations*, J. Chem. Theory Comput. **17**, 6353 (2021).

[70] X. Zhang, X. Chen, and D. G. Kuroda, *Computing the frequency fluctuation dynamics of highly coupled vibrational transitions using neural networks*, J. Chem. Phys. **154**, 164514 (2021).

[71] J. Fan, H. Lan, W. Ning, R. Zhong, F. Chen, G. Yan, and K. Cai, *Modeling amide-I vibrations of alanine dipeptide in solution by using neural network protocol*, Spectrochim. Acta Part A **268**, 120675 (2022).

[72] W. Humphrey, A. Dalke, and K. Schulten, *VMD — Visual Molecular Dynamics*, J. Mol. Graphics **14**, 33 (1996).

[73] GROMACS Version 5.0.6, 2016, URL: http://www.gromacs.org/.

[74] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, *GROMACS: Fast, flexible, and free*, J. Comput. Chem. **26**, 1701 (2005).

[75] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl, *GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers*, SoftwareX **1-2**, 19 (2015).

[76] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw, *Improved side-chain torsion potentials for the Amber ff99SB protein force field*, Proteins: Struct., Funct., Bioinf. **78**, 1950 (2010).

[77] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *Comparison of simple potential functions for simulating liquid water*, J. Chem. Phys. **79**, 926 (1983).

[78] R. Ahlrichs, M. Bär, M. Häser, H. Horn, and C. Kölmel, *Electronic structure calculations on workstation computers: The program system turbomole*, Chem. Phys. Lett. **162**, 165 (1989).

[79] S. G. Balasubramani et al., *TURBOMOLE: Modular program suite for ab initio quantum-chemical and condensed-matter simulations*, J. Chem. Phys. **152**, 184107 (2020).

[80] Y. J. Franzke et al., *TURBOMOLE: Today and Tomorrow*, J. Chem. Theory Comput. **19**, 6859 (2023).

[81] A. D. Becke, *Density-functional exchange-energy approximation with correct asymptotic behavior*, Phys. Rev. A **38**, 3098 (1988).

[82] J. P. Perdew, *Density-functional approximation for the correlation energy of the inhomogeneous electron gas*, Phys. Rev. B **33**, 8822 (1986).

[83] F. Weigend and R. Ahlrichs, *Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy*, Phys. Chem. Chem. Phys. **7**, 3297 (2005).

[84] M. Freindorf and J. Gao, *Optimization of the Lennard-Jones parameters for a combined ab initio quantum mechanical and molecular mechanical potential using the 3-21G basis set*, J. Comput. Chem. **17**, 386 (1996).

[85] J. Neugebauer, M. Reiher, C. Kind, and B. A. Hess, *Quantum chemical calculation of vibrational spectra of large molecules—Raman and IR spectra for Buckminsterfullerene*, J. Comput. Chem. **23**, 895 (2002).

[86] T. Weymuth, M. P. Haag, K. Kiewisch, S. Luber, S. Schenk, Ch. R. Jacob, C. Herrmann, J. Neugebauer, and M. Reiher, *MoViPac: Vibrational spectroscopy with a robust meta-program for massively parallel standard and inverse calculations*, J. Comput. Chem. **33**, 2186–2198 (2012).

[87] Ch. R. Jacob, P. T. Panek, T. Bergmann, J. Brüggemann, M. O. Welzel, and M. Wolter, LocVib — Python tools for localizing normal modes: Version 1.3, 2023, DOI: 10.5281/zenodo.8115103, URL: https://github.com/chjacob-tubs/LocVib.

[88] Aim, locally modified Version 1.0.2, 2024, URL: https://github.com/MarioWolter/AIM/tree/add_caps.