

## Localizing normal modes in large molecules

Christoph R. Jacob<sup>a)</sup> and Markus Reiher<sup>b)</sup>

*Laboratorium für Physikalische Chemie, ETH Zurich, Wolfgang-Pauli-Strasse 10, 8093 Zurich, Switzerland*

(Received 19 November 2008; accepted 13 January 2009; published online 24 February 2009)

We show how vibrational spectra obtained from quantum chemical calculations can be analyzed by transforming the calculated normal modes contributing to a certain band in the vibrational spectrum to a set of localized modes. This is achieved by determining the unitary transformation that leads to modes which are maximally localized with respect to a suitably defined criterion. We demonstrate that these localized modes are more appropriate for the analysis of calculated vibrational spectra of polypeptides and proteins than the normal modes, which are usually delocalized over the whole system. Both the frequencies at which the bands in the vibrational spectra appear and the total intensities of these bands can be interpreted in terms of the localized modes. Furthermore, we show how coupling constants for the interaction between the localized modes, which can be employed to rationalize the observed band shapes, can be extracted from the calculations. © 2009 American Institute of Physics. [DOI: 10.1063/1.3077690]

### I. INTRODUCTION

Vibrational spectroscopy is an important technique for the investigation of structures of biomolecules.<sup>1</sup> It is particularly suited for studying proteins in their natural environment (i.e., in aqueous solution) and can be used in many cases where other techniques such as x-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy cannot be employed. In particular, infrared (IR) and Raman spectroscopy have been used extensively for gaining information on the secondary structure of polypeptides and proteins (for reviews, see Refs. 2–5). By employing the corresponding chiral variants, i.e., vibrational circular dichroism<sup>6</sup> (VCD) and Raman optical activity (ROA) spectroscopy,<sup>7</sup> additional information can be obtained (see, e.g., Refs. 8–12). In many cases, quantum chemical calculations can provide valuable information for interpreting experimental data and for understanding how structural changes influence the observed spectra. This is particularly true for VCD and ROA spectroscopy, where no reliable empirical rules for the prediction of the spectra are available, and accurate quantum chemical calculations are thus mandatory in order to be able to interpret the experimental spectra (see, e.g., Refs. 13–15).

However, quantum chemical calculations of vibrational spectra of biomolecules such as polypeptides and proteins pose two major problems. First, the size of the relevant systems can easily reach a few hundreds or even thousands of atoms, and full calculations of such large systems are computationally demanding if not impossible with accurate methods. Second, even if possible, such calculations provide a large amount of data that will be increasingly difficult to interpret.

The first problem can be addressed by using efficient computational methods such as linear-scaling algorithms for solving the electronic structure problem (see, e.g., Refs. 16

and 17) or by employing subsystem approaches, in which the electronic structure problem is solved by partitioning the protein into its building blocks (see, e.g., Ref. 18 and references given therein). For the efficient calculation of vibrational spectra, seminumerical differentiation schemes, which allow it to implement simple but efficient restart and parallelization facilities, can be applied.<sup>19</sup> This can be combined with a selective calculation of specific normal modes<sup>20,21</sup> or of normal modes with a high intensity,<sup>22,23</sup> which both also simplifies the interpretation of the results by avoiding the calculation of unnecessary information. Furthermore, additional approximations can be introduced in the calculation of the vibrational spectra. For instance, the molecular structure can be only partially optimized (see, e.g., Ref. 24 for an approach particularly suited for vibrational spectroscopy), it can be assumed that certain parts of the system behave as rigid blocks,<sup>25</sup> or the energy and property tensor derivatives calculated for small fragments can be transferred to larger systems.<sup>26</sup>

Nevertheless, the second problem still remains. For polypeptides and proteins, the vibrational spectra comprise a number of characteristic bands, which each consists of a number of close-lying normal modes. Each of these normal modes is usually a delocalized combination of vibrations on different amino acid residues. In general, the individual modes are not or only partially resolved in experimental spectra. Nevertheless, all normal modes and the corresponding intensities contributing to one band are required in order to explain the positions and band shapes observed experimentally. Since the normal modes may have varying contributions to the observed band shape, this hampers the analysis considerably. This is particularly true for VCD and ROA spectroscopies, where close-lying positive and negative intensities can cancel.

Here, we present a way to overcome this problem by transforming the normal modes that belong to one band of a vibrational spectrum to a set of “localized modes.” While for

<sup>a)</sup>Electronic mail: christoph.jacob@phys.chem.ethz.ch.

<sup>b)</sup>Electronic mail: markus.reiher@phys.chem.ethz.ch.

polypeptides and proteins the normal modes are combinations of vibrations on different amino acid residues, a localized mode will, in general, be dominated by a vibration on one single residue. Of course, localized modes on different but homologous residues will be very similar, i.e., they will have similar vibrational frequencies and contribute similarly to the observed intensity of the band. This leads to a picture in which the positions and the intensities of the bands in the vibrational spectra can be understood in terms of the localized modes. The delocalized normal modes then arise due to the coupling between the localized modes, and a detailed picture of these couplings can also be obtained.

Traditionally the interpretation of the vibrational spectra of polypeptides and proteins proceeds in the opposite direction. From small model systems, such as *N*-methylacetamide or di- or tripeptides, local normal modes and the corresponding vibrational frequencies and intensities are obtained that are then employed to discuss the spectra of larger polypeptides or proteins (for reviews, see, e.g., Refs. 27–29). If necessary, this is combined with simplified models to describe the coupling between these local modes, such as the transition dipole coupling (TDC) model.<sup>28</sup> However, such a procedure relies on the assumption that the vibrational frequencies and intensities of local modes found in these small model systems can be transferred to larger proteins and that the couplings between these local modes can be accurately described by simplified models, which both cannot be expected to be generally valid. In contrast, the methodology that will be presented here allows us to extract vibrational frequencies and intensities of local modes as well as the corresponding coupling constants from full quantum chemical calculations, thus including effects that are otherwise neglected. We should note that for the special case of the amide I band in proteins, the “Hessian matrix reconstruction” method by Ham *et al.*<sup>30</sup> can also be employed to obtain local amide I modes and the corresponding coupling constants from calculations on the full system. However, their method is limited to carbonyl stretching vibrations, while we will present a more general approach that can be applied to arbitrary modes.

This work is organized as follows. In Sec. II the theoretical framework will be derived. First, in Sec. II A we discuss how normal modes can be transformed to localized mode. In Sec. II B, we then show how vibrational frequencies of the localized modes and coupling constants can be determined, and finally it is discussed how the IR, Raman, VCD, and ROA intensities can be expressed in terms of localized modes in Sec. II C. The computational details are given in Sec. III, before the developed methodology is applied to the example of an  $\alpha$ -helical alanine polypeptide in Sec. IV. Finally, concluding remarks and an outlook are given in Sec. V.

## II. THEORY

### A. Localization of normal modes

Within the harmonic approximation, the normal modes and vibrational frequencies can be obtained by diagonalizing the mass-weighted molecular Hessian matrix  $\mathbf{H}^{(m)}$ , which

contains the second derivatives of the total electronic energy  $E$  (including the nuclear repulsion energy) with respect to Cartesian nuclear coordinates,<sup>1</sup> i.e.,

$$H_{i\alpha,j\beta}^{(m)} = \frac{1}{\sqrt{m_i m_j}} \left( \frac{\partial^2 E}{\partial R_{i\alpha} \partial R_{j\beta}} \right)_0, \quad (1)$$

where the subscript “0” denotes that the derivative is evaluated at the equilibrium structure  $\mathbf{R}_0$ . In the above expression,  $m_i$  is the atomic mass of nucleus  $i$  and  $R_{i\alpha}$  is the  $\alpha=x,y,z$  Cartesian component of the position of nucleus  $i$ . The matrix  $\mathbf{Q}$  containing the eigenvectors of  $\mathbf{H}^{(m)}$  can be employed to transform the Hessian matrix to diagonal form

$$\mathbf{H}^{(q)} = \mathbf{Q}^T \mathbf{H}^{(m)} \mathbf{Q}, \quad (2)$$

where  $\mathbf{H}^{(q)}$  is a diagonal matrix with the diagonal elements (i.e., the eigenvalues of  $\mathbf{H}^{(m)}$ ) equal to the squares of the angular frequencies,  $H_{pp}^{(q)} = \omega_p^2 = 4\pi^2 \nu_p^2$ , with vibrational frequencies  $\nu_p$ . The columns of the matrix  $\mathbf{Q}$  are the normal modes in terms of mass-weighted Cartesian displacements. These normal modes will be referred to as  $\mathbf{Q}_p$  and its components as  $Q_{i\alpha,p}$ . The  $\mathbf{Q}_p$  are chosen to be normalized, i.e.,  $|\mathbf{Q}_p|^2 = \sum_{i\alpha} Q_{i\alpha,p}^2 = 1$ , so that the matrix  $\mathbf{Q}$  is unitary. The normal modes in terms of not mass-weighted Cartesian displacements  $\mathbf{Q}_p^{(c)}$  can be obtained as  $Q_{i\alpha,p}^{(c)} = (1/\sqrt{m_i}) Q_{i\alpha,p}$ .

Now we consider a subset of  $k$  normal modes, e.g., all modes which are contributing to one band in the vibrational spectrum. These are collected in the matrix  $\mathbf{Q}^{\text{sub}}$ , which is of dimension  $3n \times k$ , where  $n$  is the number of atoms. By means of a unitary transformation  $\mathbf{U}$  (of dimension  $k \times k$ ), these can be transformed to a new set of modes,

$$\tilde{\mathbf{Q}}^{\text{sub}} = \mathbf{Q}^{\text{sub}} \mathbf{U} \quad (3)$$

or

$$\tilde{Q}_{i\alpha,p}^{\text{sub}} = \sum_q U_{qp} Q_{i\alpha,q}^{\text{sub}}. \quad (4)$$

In order to facilitate the interpretation of complex vibrational spectra, we want to determine the unitary transformation that yields the “most localized” modes, i.e., that maximizes  $\xi(\tilde{\mathbf{Q}}^{\text{sub}}) = \xi(\mathbf{Q}^{\text{sub}} \mathbf{U})$ , where  $\xi(\tilde{\mathbf{Q}}^{\text{sub}})$  is a function that measures how localized a set of modes  $\tilde{\mathbf{Q}}^{\text{sub}}$  is. For this function  $\xi(\tilde{\mathbf{Q}}^{\text{sub}})$ , different definitions can be employed. In analogy to the criteria used for orbital localization, we will consider two different definitions.

First, according to the orbital localization criterion of Pipek and Mezey,<sup>31</sup> which maximizes the sum of the squares of the “atomic contribution” to an orbital (measured by the atomic Mulliken populations), one can maximize the sum of the squares of the atomic contributions to the modes, i.e.,

$$\xi_{\text{Sat}}(\tilde{\mathbf{Q}}^{\text{sub}}) = \sum_{p=1}^k \sum_{i=1}^n (\tilde{C}_{ip}^{\text{sub}})^2. \quad (5)$$

In this definition,  $\tilde{C}_{ip}^{\text{sub}}$  is the contribution of nucleus  $i$  to the normal mode  $\mathbf{Q}_p^{\text{sub}}$ , which can be measured by the fraction of the kinetic energy of this atom in the normal mode as<sup>32–34</sup>

$$\tilde{C}_{ip}^{\text{sub}} = \sum_{\alpha=x,y,z} (\tilde{Q}_{i\alpha,p}^{\text{sub}})^2. \quad (6)$$

Second, in analogy to the Boys orbital localization criterion,<sup>35–37</sup> one can maximize the distance between the “centers” of the modes, which can be defined by weighting the atomic coordinates with the corresponding contributions to the normal mode as

$$\mathbf{R}_p^{\text{center}} = \sum_{i=1}^n \tilde{C}_{ip}^{\text{sub}} \mathbf{R}_i. \quad (7)$$

This leads to the localization criterion

$$\xi'_{\text{dist}}(\tilde{\mathbf{Q}}^{\text{sub}}) = \sum_{p=1}^k \sum_{q=1}^k (\mathbf{R}_p^{\text{center}} - \mathbf{R}_q^{\text{center}})^2. \quad (8)$$

However, maximizing  $\xi'_{\text{dist}}$  is equivalent to maximizing the distance of the centers of the modes from the origin (see Appendix A), so that one obtains the simpler localization criterion,

$$\xi_{\text{dist}}(\tilde{\mathbf{Q}}^{\text{sub}}) = \sum_{p=1}^k (\mathbf{R}_p^{\text{center}})^2 = \sum_{p=1}^k \left( \sum_{i=1}^n \tilde{C}_{ip}^{\text{sub}} \mathbf{R}_i \right)^2. \quad (9)$$

In the following, both the *atomic-contribution* criterion  $\xi_{\text{at}}$  and the *distance* criterion  $\xi_{\text{dist}}$  will be assessed.

Given one of these localization criteria, the optimally localized modes can be obtained using a similar procedure as for the calculation of localized orbitals. The easiest and most popular method for this purpose is to perform so-called Jacobi sweeps, as was first proposed for the determination of localized orbitals by Edmiston and Ruedenberg<sup>37</sup> (see also Ref. 31 for a summary). A brief review of the Jacobi-sweep method and a description of its application to the vibrational problem under consideration here are given in Appendix B. Of course, other algorithms for the calculation of localized orbitals that are more efficient for large systems<sup>38</sup> may easily be adopted for the determination of localized vibrational modes.

## B. Vibrational frequencies of localized modes and coupling constants

The localized modes  $\tilde{\mathbf{Q}}_p^{\text{sub}}$ , which are obtained from the normal modes by the (localizing) unitary transformation, are not eigenvectors of the mass-weighted Hessian  $\mathbf{H}^{(m)}$  and, therefore, have no direct physical significance. However, in contrast to the usually very delocalized normal modes, they involve only displacements of a few atoms, and the localized modes can thus be useful for the understanding of the influences determining the calculated vibrational spectra.

The Hessian with respect to the localized modes is not diagonal, but it is given by

$$\tilde{\mathbf{H}}^{\text{sub}} = \mathbf{U}^T \mathbf{H}^{(q),\text{sub}} \mathbf{U}, \quad (10)$$

where  $\mathbf{H}^{(q),\text{sub}}$  is a diagonal submatrix of the complete diagonalized Hessian  $\mathbf{H}^{(q)}$ , with the nonzero entries given by the squared angular frequencies  $\omega_p^2 = 4\pi^2\nu_p^2$  of those vibrations corresponding to the selected subset of normal modes. Here and in the following the tilde is used to denote quantities that

are defined with respect to the localized modes.

The diagonal elements of  $\tilde{\mathbf{H}}^{\text{sub}}$  can be used to extract fictitious vibrational frequencies  $\tilde{\nu}_p$  of the localized modes, which may be defined by the equation  $\tilde{H}_{pp}^{\text{sub}} = \tilde{\omega}^2 = 4\pi^2\tilde{\nu}_p^2$ , while the off-diagonal elements  $\tilde{H}_{pq}^{\text{sub}}$  describe the coupling between the localized modes  $p$  and  $q$ . Note that  $\tilde{\nu}_p$  does not refer to the wavenumber of the vibrational transition, but to the fictitious vibrational frequencies of the localized modes, with the tilde indicating quantities defined with respect to localized modes.

For the set of (delocalized) normal modes contributing to a certain band of the vibrational spectrum that arises due to the coupling of vibrations of similar groups, the localized modes will describe the vibrations of the individual groups, and the corresponding frequencies of the localized modes  $\tilde{\nu}_p$  will be very similar. The off-diagonal coupling elements of  $\tilde{\mathbf{H}}^{\text{sub}}$  will, in general, be small for localized modes that are centered on groups that are not spatially close. Consequently, for each localized mode only couplings with a small number of neighboring modes will be significant. Therefore, the matrix  $\tilde{\mathbf{H}}^{\text{sub}}$  will, in general, have a rather simple structure which allows it to explain how the more complicated delocalized normal modes arise.

However, since the diagonal elements of  $\mathbf{H}^{(q),\text{sub}}$  and  $\tilde{\mathbf{H}}^{\text{sub}}$ , respectively, correspond to squared angular frequencies and not to vibrational frequencies themselves, the numerical values of the off-diagonal elements of  $\tilde{\mathbf{H}}^{\text{sub}}$  cannot be interpreted as the size of the frequency splitting in the case of two degenerate modes (compare also the discussion of a related problem that arises when considering couplings between electronic excitations in Ref. 39). Therefore, it is useful to define the coupling matrix  $\tilde{\mathbf{\Omega}}$  as

$$\tilde{\mathbf{\Omega}} = \mathbf{U}^T \mathbf{\Omega} \mathbf{U}, \quad (11)$$

where  $\mathbf{\Omega}$  is a diagonal matrix with the vibrational frequencies  $\nu_p = \omega_p/2\pi$  of the considered normal modes on the diagonal. Note that  $\tilde{\mathbf{H}}^{\text{sub}}$  and  $\tilde{\mathbf{\Omega}}$  are related by

$$\tilde{\mathbf{\Omega}} = \frac{1}{2\pi} (\tilde{\mathbf{H}}^{\text{sub}})^{1/2}. \quad (12)$$

From this coupling matrix  $\tilde{\mathbf{\Omega}}$ , the vibrational frequencies of the normal modes and the transformation matrix  $\mathbf{U}$  can be obtained by diagonalization, and the eigenvectors of  $\tilde{\mathbf{\Omega}}$  are the columns of the inverse transformation  $\mathbf{U}^T$ , which give the composition of the normal modes in the basis of localized modes. In the case of two coupled localized modes, the vibrational frequencies of the corresponding normal modes are given by

$$\nu_{\pm} = \frac{\tilde{\Omega}_{11} + \tilde{\Omega}_{22}}{2} \pm \sqrt{\left( \frac{\tilde{\Omega}_{11} - \tilde{\Omega}_{22}}{2} \right)^2 + \tilde{\Omega}_{12}^2}. \quad (13)$$

For a larger number of modes, more complicated coupling patterns will arise. However, it can be seen from this expression for two localized modes that the diagonal elements  $\tilde{\Omega}_{pp}$  can be interpreted as vibrational frequencies of the localized

modes, while the off-diagonal elements  $\tilde{\Omega}_{pq}$  can be understood as coupling constants. The numerical values of these coupling constants correspond exactly to half the splitting between two degenerate coupled localized modes.

It should be noted that the frequencies of the localized modes obtained as the diagonal elements of  $\tilde{\Omega}$  are, in general, different from those calculated from the diagonal elements of  $\tilde{H}^{\text{sub}}$ . However, in the cases considered here, where the off-diagonal elements of  $\tilde{\Omega}$  are small compared to the diagonal elements, the two definitions for the frequencies of the localized modes are almost identical, i.e.,

$$\tilde{\Omega}_{ii} \approx \tilde{\nu}_i = \frac{\tilde{\omega}_i}{2\pi} = \frac{\sqrt{\tilde{H}_{ii}^{\text{sub}}}}{2\pi}. \quad (14)$$

In the cases considered in this paper, the differences are always smaller than  $0.2 \text{ cm}^{-1}$ . Similarly, the coupling constants are related to the off-diagonal elements of  $\tilde{H}^{\text{sub}}$  by

$$\tilde{\Omega}_{ij} \approx \frac{\tilde{H}_{ij}^{\text{sub}}}{4\pi^2(\tilde{\nu}_i + \tilde{\nu}_j)}. \quad (15)$$

A derivation of the approximate relations given in Eqs. (14) and (15) can be found in Appendix C.

### C. Vibrational intensities in terms of localized modes

While the vibrational frequencies of the localized modes and the couplings between them provide a tool for analyzing the position of the peaks in the vibrational spectrum, the localized modes can also be used to understand the intensities of bands in the vibrational spectrum. In the following, we will consider IR, VCD, Raman, and ROA spectroscopy. The expressions one obtains for the vibrational intensities in these different types of vibrational spectroscopy are reviewed in Appendix D.

As is apparent from the expressions given there, the measured intensity (i.e., IR absorption, scattering intensities, or the respective differences) of the  $p$ th vibrational mode is in all these cases of the form

$$I_p \propto \sum_l c_l \left( \frac{\partial P_l^{(1)}}{\partial Q_p} \right)_0 \left( \frac{\partial P_l^{(2)}}{\partial Q_p} \right)_0, \quad (16)$$

where  $P_l^{(1)}$  and  $P_l^{(2)}$  are components of the appropriate property tensors, i.e., electric dipole moment, magnetic dipole moment, electric dipole–electric dipole polarizability, electric dipole–magnetic dipole polarizability, or electric dipole–

electric quadrupole polarizability,  $Q_p$  denotes that the derivative is taken either with respect to the normal mode displacement  $Q_p$  or with respect to the normal mode velocity  $\dot{Q}_p$  in the case of VCD, and where the derivatives with respect to the normal modes or normal mode velocities are given by

$$\left( \frac{\partial \mu_\alpha}{\partial Q_p} \right)_0 = \sum_{i\beta} Q_{i\beta,p}^{(c)} \left( \frac{\partial \mu_\alpha}{\partial R_{i\beta}} \right)_0 = \sum_{i\beta} \frac{1}{\sqrt{m_i}} Q_{i\beta,p} \left( \frac{\partial \mu_\alpha}{\partial R_{i\beta}} \right)_0. \quad (17)$$

If one considers a band in the vibrational spectrum that is made up from a subset of  $k$  vibrational transitions, with the corresponding normal modes  $Q_p^{\text{sub}}$ , the total intensity of this band (i.e., the intensity integrated over the whole band) is given by

$$I^{\text{band}} \propto \sum_l c_l \sum_{p=1}^k \left( \frac{\partial P_l^{(1)}}{\partial Q_p^{\text{sub}}} \right)_0 \left( \frac{\partial P_l^{(2)}}{\partial Q_p^{\text{sub}}} \right)_0, \quad (18)$$

where the index  $p$  runs over the considered subset of vibrational modes. In this expression, it is assumed that the proportionality constants are equal for all considered transitions, i.e., that they are independent of the vibrational frequency. This is, in general, true for IR and VCD spectroscopy, but for Raman and ROA spectroscopy the proportionality constants will depend on the vibrational frequency. For details, see Refs. 19 and 40, where expressions for the absolute IR absorption as well as for the absolute Raman scattering intensity can be found. However, if one only considers one band in the Raman or ROA spectrum, for which the difference between the vibrational frequencies contributing to this band is small, it is a good approximation to assume that the proportionality constants are equal. Furthermore, it is common—in particular, in theoretical studies—to ignore the proportionality constants since they depend on the precise experimental conditions, and to consider only the appropriate combination of Raman or ROA invariants, respectively.<sup>19,41</sup> This is especially true for the purpose of analyzing the calculated intensity patterns, for which one usually analyzes the individual invariants.<sup>33</sup>

As is shown in Appendix E, the individual terms in Eq. (18) are invariant under unitary transformations, i.e., the total intensities of the considered band can also be expressed in terms of appropriate local modes as

$$I^{\text{band}} \propto \sum_l c_l \sum_{p=1}^k \left( \frac{\partial P_l^{(1)}}{\partial \tilde{Q}_p^{\text{sub}}} \right)_0 \left( \frac{\partial P_l^{(2)}}{\partial \tilde{Q}_p^{\text{sub}}} \right)_0, \quad (19)$$

where the derivatives are taken with respect to the transformed modes  $\tilde{Q}^{\text{sub}}$  or with respect to the velocity of the transformed modes  $\dot{\tilde{Q}}^{\text{sub}}$  in the case of the magnetic dipole contribution to VCD. This allows one to analyze the intensity of bands in vibrational spectra in terms of localized modes, which are in many cases much simpler than the delocalized normal modes, facilitating a qualitative understanding of intensities in terms of local contributions. Since, in general, the intensities of the different localized modes obtained for one band will be almost equal, it will in many cases be sufficient to analyze only one representative localized mode.

While the total intensity of a band in the vibrational spectrum is invariant under a transformation of the contributing normal modes, the intensities corresponding to the individual transformed modes differ from those of the normal

modes. Therefore, in order to understand the shape of the bands in the vibrational spectrum, the localized modes and their corresponding intensities alone are not sufficient, but also the couplings between them as well as the cross terms in the intensities arising due to these couplings have to be considered. However, as will be shown in Sec. IV C, these couplings, in general, have a rather simple structure, which allows it to understand the observed band shapes more easily.

### III. COMPUTATIONAL DETAILS

The structure optimization of  $\alpha$ -helical Ala<sub>20</sub> was performed with the TURBOMOLE program package<sup>42,43</sup> employing density-functional theory. The BP86 exchange-correlation functional<sup>44,45</sup> and Ahlrichs' valence triple-zeta basis set with polarization functions at all atoms (TZVP)<sup>46,47</sup> and the corresponding auxiliary basis set<sup>48,49</sup> were employed. The structure has been fully optimized. However, in order to converge to the desired local minimum, constraints fixing the positions of different groups of atoms were applied in the initial phase of the optimization. However, all these constraints were relaxed in the final stage of the optimization, and the optimized structure that was used for the calculation of the normal modes and vibrational frequencies thus corresponds to a local minimum on the potential energy surface.

The program SNF (Refs. 19 and 50) was used to calculate the normal modes and vibrational frequencies, as well as the dipole and polarizability derivatives that determine the IR absorption and the Raman scattering factor. The analytic energy gradients, which are needed for the seminumerical calculation of the harmonic force field, were calculated with TURBOMOLE for distorted structures.<sup>19,50</sup> The Raman scattering factors were calculated for an excitation wavelength of 799 nm, and it was verified that this wavelength is well away from any electronic absorption frequency of the considered molecule. All calculations were performed for the isolated molecule and solvent effects were not included.

The localization of normal modes and the related analysis routines have been implemented in an add-on package to SNF written in the PYTHON programming language. The NUMPY package<sup>51</sup> was used for efficient linear algebra operations where needed. For managing the geometric coordinates and for the assignment of atoms to individual residues and atom types in the investigated polypeptide, the OPENBABEL library<sup>52,53</sup> has been employed.

Pictures of molecular structures and normal modes were prepared with JMOL.<sup>54</sup> Plots of vibrational spectra and of group coupling matrices were produced using the MATPLOTLIB package.<sup>55</sup>

## IV. RESULTS AND DISCUSSION

### A. IR and Raman spectra of $\alpha$ -helical (Ala)<sub>20</sub>

In order to demonstrate the usefulness of localized modes for the interpretation and understanding of vibrational spectra, we study the calculated IR and Raman spectra of a polypeptide consisting of 20 S-alanine residues, denoted (Ala)<sub>20</sub>, in the conformation of an  $\alpha$ -helix. This  $\alpha$ -helical polypeptide can serve as a simple model for the vibrational

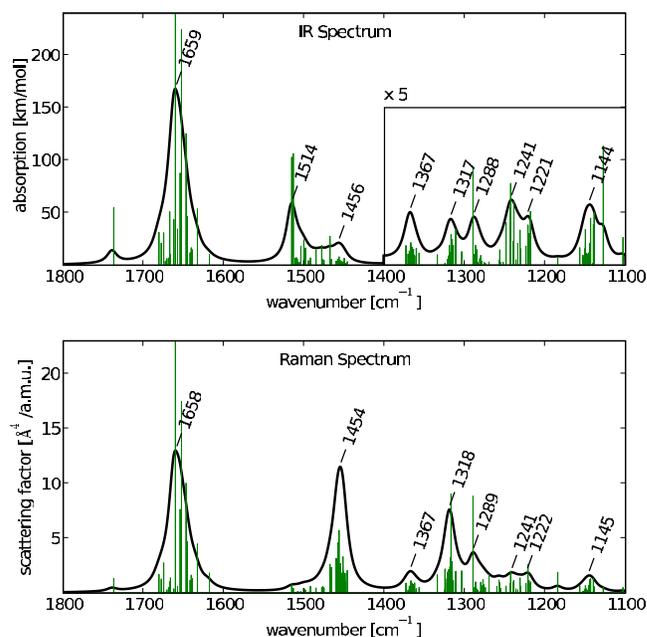


FIG. 1. (Color online) Calculated IR and Raman spectra of  $\alpha$ -helical (Ala)<sub>20</sub>. The spectra have been plotted using a Lorentzian line width of 15 cm<sup>-1</sup>. Individual peaks have been included as a line spectrum scaled by 0.2. In the IR spectrum, the region below 1400 cm<sup>-1</sup> has been magnified by a factor of 5.

spectra of proteins and allows us to study the important bands that also appear in spectra of more general polypeptides and proteins.

The molecular structure of the (Ala)<sub>20</sub> polypeptide has been obtained from a full geometry optimization. Therefore, a rather large polypeptide containing 20 alanine residues has to be used in order to model an  $\alpha$ -helix, since for smaller polypeptides the structure will otherwise tend toward a 3<sub>10</sub>-helix.<sup>56</sup> The backbone dihedral angles of the central residues are approximately  $\varphi = -63^\circ$  and  $\psi = -42^\circ$ . Note that at the termini, the helices are slightly disturbed and tend toward a 3<sub>10</sub>-helix. These problems could be circumvented by applying constraint optimization techniques in order to obtain an idealized  $\alpha$ -helix,<sup>24</sup> but then one has to be careful to ensure that the gradient with respect to the investigated normal modes vanishes, which is not always easy to achieve. Furthermore, the results might be sensitive to the chosen initial structure and the applied constraints. In addition to that, realistic models of solvated polypeptides or proteins the structure will not agree with an idealized structure, but there will be structural distortions because of end group effects as well as the solvent environment and because of differences in the amino acid side chains. Therefore, we want to demonstrate that the methodology developed here can be applied for such distorted structures. Note that for idealized models, simpler procedures that rely on the structural similarity could be used to obtain localized modes,<sup>30</sup> while our method is more generally applicable.

The calculated IR and Raman spectra of  $\alpha$ -helical (Ala)<sub>20</sub> in the region between 1800 and 1100 cm<sup>-1</sup>, which includes the most important bands that are commonly employed for investigations of polypeptides and proteins, are shown in Fig. 1. As it is common, these spectra have been

TABLE I. Assignment of modes in the calculated vibrational spectrum of  $\alpha$ -helical (Ala)<sub>20</sub> to vibrational bands. The assignment of modes to different bands is based both on the calculated wavenumbers and on the contributions of the different groups of atoms (carbonyl C=O, amide N-H, backbone C <sup>$\alpha$</sup> -H, and sidechain C <sup>$\beta$</sup> H<sub>3</sub>) to the normal modes measured according to Eq. (6), which are also listed. See text for details.

Mode no.	Range (cm <sup>-1</sup> )	Group contributions (%)				
		N-H	C=O	C <sup><math>\alpha</math></sup> -H	C <sup><math>\beta</math></sup> H <sub>3</sub>	
500	1737	2.6	96.3	0.8	0.3	Carboxyl C=O stretch
481-499	1633-1681	3-5	93-96	1-3	0-1	Amide I
480	1618	98.4	0.3	1.0	0.3	NH <sub>2</sub> bend
462-479	1475-1515	66-78	16-22	3-7	2-8	Amide II
461	1468	53.3	20.6	4.4	21.7	Amide II/C <sup><math>\beta</math></sup> H <sub>3</sub> bend
421-460	1446-1466	0-8	0-2	1-2	88-98	C <sup><math>\beta</math></sup> H <sub>3</sub> asymm. bend
401-420	1356-1373	0-2	0-2	0-2	97-99	C <sup><math>\beta</math></sup> H <sub>3</sub> symm. bend
380, 381, 383-400	1304-1334	3-13	1-15	76-89	3-5	C <sup><math>\alpha</math></sup> H bend (I)
382	1314	32.8	55.0	8.0	4.2	Carboxyl O-H bend
359, 361-379	1257-1290	3-11	3-19	65-78	8-15	C <sup><math>\alpha</math></sup> H bend (II)
340-358, 360	1184-1260	27-45	15-32	16-37	8-16	Amide III
319, 321-339	1103-1156	21-26	2-11	40-50	23-30	Skeletal C <sup><math>\alpha</math></sup> -N stretch
320	1127	31.2	45.0	18.7	5.0	Carboxyl C-O stretch

obtained by broadening the calculated individual transitions by a Lorentzian line shape with a half-width of 15 cm<sup>-1</sup> in order to obtain spectra that can be compared to those obtained experimentally. However, each band in the plotted calculated spectrum, as well as each band in the observed spectra, consists of a number of vibrational transitions, each corresponding to one normal mode. These individual transitions have been included in the spectra as lines (see Fig. 1), and it can be seen that the intensities of the individual transitions within one band in the spectrum show large differences.

In order to analyze the calculated spectra, it is first necessary to identify the individual modes that contribute to the observed bands in the vibrational spectra, i.e., the calculated normal modes have to be grouped into bands. This can be achieved by considering (a) the wavenumbers of the transitions, which for most vibrational bands cluster around the corresponding band maxima, and (b) by collecting bands for which the contributions of certain atom types to the normal modes as calculated according to Eq. (6) show a similar pattern. For the alanine polypeptides studied here, we considered the contributions of the carbonyl C=O groups, the amide N-H groups, the C <sup>$\alpha$</sup> -H groups, and the sidechain methyl C <sup>$\beta$</sup> H<sub>3</sub> groups.

The assignment of the calculated normal modes to bands is listed in Table I. While there are some single modes that originate from the end groups, in particular, the terminal carboxyl and NH<sub>2</sub> groups, for most bands there is a set of 20 normal modes that appears within a clearly defined wavenumber range and that shows similar contributions for the considered atom types. These sets of normal modes correspond to the peaks found in the calculated IR and Raman spectra and can be identified with the modes usually found in the vibrational spectrum of polypeptides in the wavenumber range considered, i.e., the amide I, II, and III modes, symmetric and asymmetric C <sup>$\beta$</sup> H<sub>3</sub> bending modes, C <sup>$\alpha$</sup> H bending modes, and skeletal C <sup>$\alpha$</sup> -N stretching modes.<sup>5,27,28</sup> Only nor-

mal mode 461 cannot be assigned to either the terminal carboxyl or NH<sub>2</sub> group or to one of these bands, but instead appears to be a mixture of amide II and C <sup>$\beta$</sup> H<sub>3</sub> bending modes.

All peaks in the IR and Raman spectra shown in Fig. 1 can be identified with one of these sets of normal modes. The wavenumber range of these sets of modes, the corresponding band maxima, as well as the total calculated IR absorption and Raman scattering factors of these bands are collected in Table II. The assignment of these groups of modes to certain types of vibrations will be discussed in more detail in Sec. IV B. Briefly, the amide I modes are dominated by C=O stretching vibrations, and the amide II and amide III modes are mainly N-H bending modes. The modes denoted as C <sup>$\alpha$</sup> -H bending and those denoted as amide III modes in Table II are usually jointly referred to as “extended amide III region,” since all these modes consist of coupled N-H bending and C <sup>$\alpha$</sup> -H bending vibrations.<sup>27</sup> For all bands, these main contributions can also be identified from the contributions of the different atom types listed in Table I.

## B. Localized modes

To rationalize the calculated spectra, each of the bands discussed in Sec. IV A can be separately analyzed in terms of localized modes. In order to explain the procedure used for this analysis and to point out general characteristics of the localized modes, we will first focus on the amide I band.

For the normal modes contributing to the amide I band, Table III lists the wavenumbers of the vibrational transitions together with the IR absorption and the Raman scattering factor. The modes span a rather large frequency range of almost 50 cm<sup>-1</sup> and show an irregular intensity pattern, in which the intensities of some modes are very strong, while those of other modes are almost negligible. These amide I normal modes are combinations of carbonyl stretching vibrations, that are—with the exception of a few modes localized

TABLE II. Wavenumber ranges, band maxima, and sum of the IR absorption and Raman scattering factor for the characteristic bands in the vibrational spectra of  $\alpha$ -helical (Ala)<sub>20</sub>. Band maxima refer to the calculated IR spectrum.

	Range (cm <sup>-1</sup> )	Maximum (cm <sup>-1</sup> )	Total IR abs. (km/mol)	Total Raman int. (Å <sup>4</sup> /amu)
Amide I	1633–1681	1659	6279.9	483.1
Amide II	1468–1515	1514	1918.6	30.3
C <sup>β</sup> H <sub>3</sub> asym. bend	1446–1466	1456	496.4	322.8
C <sup>β</sup> H <sub>3</sub> sym. bend	1356–1373	1367	248.7	45.2
C <sup>α</sup> H bend (I)	1304–1334	1317	198.8	199.9
C <sup>α</sup> H bend (II)	1257–1290	1288	215.0	101.2
Amide III	1184–1260	1221, 1241	528.4	95.8
Skeletal C <sup>α</sup> –N stretch	1103–1156	1144	353.6	47.5

on the terminal residues that appear mainly at either high or low wavenumbers—delocalized over the full helix. As representative examples, the normal modes 490 and 491 at 1660.4 and 1662.4 cm<sup>-1</sup>, respectively, are shown in Fig. 2(a).

After transformation of the amide I normal modes to a set of localized modes, the situation is rather different. The wavenumbers of the localized modes—obtained as the diagonal elements of the coupling matrix  $\Omega$  calculated according to Eq. (11)—as well as the IR absorption and the Raman scattering factors of the localized modes are listed in Table III. In this table, the localized modes are sorted according to the amino acid residue that gives the largest contribution to the mode. For the transformation to localized orbitals the atomic-contribution localization criterion has been used—a comparison of both schemes, atomic-contribution and distance localization criteria, proposed in the theory part will be presented later in Sec. IV D.

The wavenumbers of the localized modes cover a range of only approximately 15 cm<sup>-1</sup> (between 1649 and 1663 cm<sup>-1</sup>), much smaller than for the normal modes, with the exception of the localized modes at the terminal residues 3, 16, 18, and 19. Also the intensities show, in contrast to those of the normal modes, a very regular pattern with very similar IR absorption (of approximately 350 km/mol) and Raman scattering factors (of approximately 30 Å<sup>4</sup>/amu) for each localized mode. However, some irregularities for modes localized on the terminal residues occur. Note that, as discussed in Sec. II C, the total IR absorption and the total Raman scattering factor are equal for normal modes and localized modes.

In Fig. 2(b), the two localized amide I modes with the main contributions at residues 9 and 10 are shown. In contrast to the amide I normal modes these are not combinations of a number of carbonyl stretching vibrations, but each localized mode is dominated by one single carbonyl stretching

TABLE III. Wavenumbers, IR absorption, and Raman scattering factor for amide I normal modes and the corresponding localized modes of  $\alpha$ -helical (Ala)<sub>20</sub>. See text for details.

Normal modes				Localized modes			
Mode	$\nu_i$ (cm <sup>-1</sup> )	IR abs. (km/mol)	Raman int. (Å <sup>4</sup> /amu)	Residue	$\tilde{\Omega}_{ii}$ (cm <sup>-1</sup> )	IR abs. (km/mol)	Raman int. (Å <sup>4</sup> /amu)
481	1633.0	268.9	22.5	1	1658.8	329.9	28.2
482	1639.1	74.5	6.6	2	1663.1	283.1	26.0
483	1640.7	84.4	8.3	3	1637.0	361.4	30.5
484	1642.1	60.2	5.9	4	1649.4	351.8	31.0
485	1645.4	232.1	23.5	5	1659.3	329.8	29.8
486	1647.0	627.6	50.2	6	1662.0	356.3	31.2
487	1652.8	1123.5	87.4	7	1659.1	359.3	31.1
488	1654.5	439.6	38.3	8	1655.9	361.1	31.0
489	1657.7	173.1	3.2	9	1654.6	368.6	31.1
490	1660.4	2163.8	189.5	10	1654.9	360.2	29.7
491	1662.4	218.4	2.0	11	1654.8	369.8	29.1
492	1666.5	52.3	7.0	12	1655.8	367.7	29.2
493	1667.3	255.6	5.2	13	1662.1	293.9	19.8
494	1668.9	35.6	2.7	14	1649.9	387.4	29.8
495	1671.0	32.3	1.6	15	1661.4	350.0	24.6
496	1672.8	19.4	0.4	16	1676.6	274.2	16.1
497	1674.9	155.3	13.9	17	1656.4	283.5	15.8
498	1677.9	106.1	6.5	18	1676.2	252.3	11.5
499	1680.8	157.4	8.6	19	1667.6	239.6	7.6
Sum		6279.9	483.1			6279.9	483.1

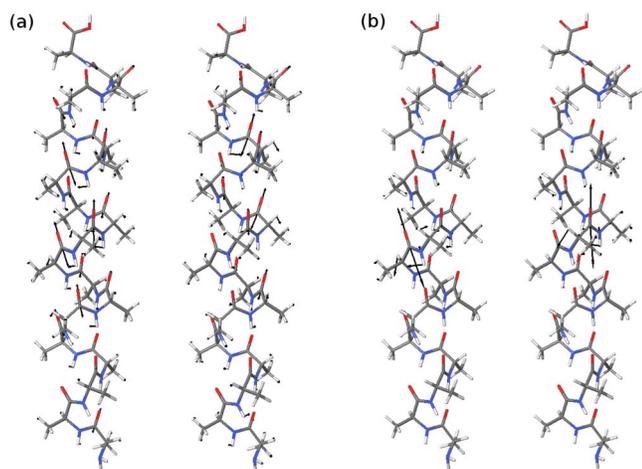


FIG. 2. (Color online) Amide I normal modes and localized modes for  $\alpha$ -helical  $(\text{Ala})_{20}$ . (a) Normal modes 490 and 491 at 1660.4 and 1662.4  $\text{cm}^{-1}$ , respectively. (b) Localized amide I modes with main contribution on residues 9 and 10. The (fictitious) wavenumbers  $\Omega_{ii}$  of these modes are 1654.6 and 1654.9  $\text{cm}^{-1}$ , respectively. See text for details.

vibration. The different localized modes involve almost identical types of displacements. Homologous atoms perform the same type of collective motion at different residues if localized modes are compared. Therefore, these localized modes exhibit very similar wavenumbers and intensities. The variations in both wavenumbers and intensities that are found for the localized modes on different residues are due to the differences in the geometric structure and the environment (e.g., different intramolecular hydrogen bonding patterns at the termini) along the helix. However, since these variations are small, it will often be sufficient to consider only one representative localized mode that can then be taken as a model for all localized modes. This localized mode and its intensity can then be further analyzed, for instance, by performing a decomposition into local contributions using the scheme pro-

posed by Hug.<sup>33</sup> It should be noted that this is not equivalent to using normal modes of smaller fragments to interpret the spectrum of larger polypeptides, since the localized modes include the effects of all residues in the full system on the electronic structure and the true normal modes of the full system can be obtained from the localized modes by a suitable transformation.

The transformation to localized modes also provides a means for visualizing the vibrations responsible for the various bands in the vibrational spectra. While the normal modes are delocalized over a large number of residues and thus—because of the large number of involved atoms—hard to visualize, the localized modes involve only a few atoms. This makes it possible to visualize the involved displacements for a small number of residues only. Figure 3 shows for each considered band in the vibrational spectra of  $\alpha$ -helical  $(\text{Ala})_{20}$  one of the localized modes with the main contribution on residues 9 or 10. Since the contributions of atoms in all other residues are small for these localized modes, only the atoms of residues 9 and 10 have to be shown, which simplifies the pictures significantly compared to Fig. 2.

As already discussed above, the amide I modes are dominated by the carbonyl C=O stretching vibration. However, as Fig. 3 shows there are also smaller contributions of N-H bending and C $^{\alpha}$ -H bending vibrations. The amide II mode is an out-of-phase combination of the N-H bending and the C-N stretching vibrations of the amide group. The asymmetric and symmetric C $^{\beta}$ H $_3$  bending modes can also be easily identified in the Fig. 3 and are hardly coupled to any other vibrations. For the modes in the extended amide III region the situation is more complicated. The modes labeled as C $^{\alpha}$ -H bending (I) are mainly a deformation of the C $^{\alpha}$ -H group in the direction of the C $^{\alpha}$ -H bond, while those labeled C $^{\alpha}$ -H bending (II) are dominated by a deformation perpendicular to the C $^{\alpha}$ -N bond. However, both modes also have

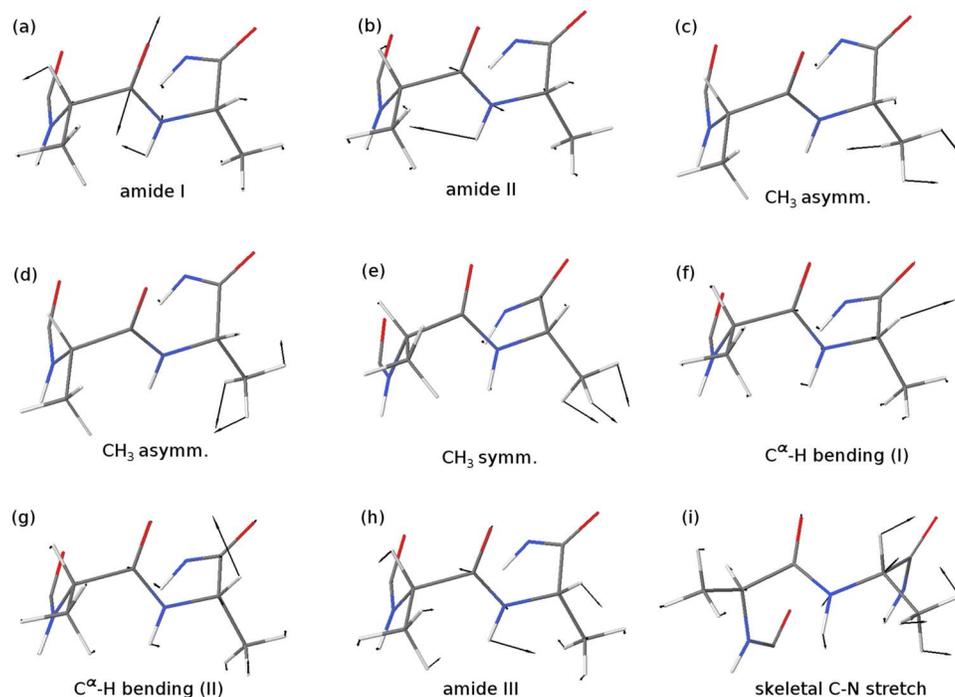


FIG. 3. (Color online) Localized modes of all considered bands in the vibrational spectra of  $\alpha$ -helical  $(\text{Ala})_{20}$ . Only residues 9 and 10 are shown, the remaining parts of the helix are left out for clarity. For each band, one of the localized modes with the main contribution on these residues is shown. The contribution of atoms in the remaining residues to these modes is negligible.

TABLE IV. Coupling constants for the interaction of the localized modes on residue 8 with those on residues 9–12. For the asymmetric  $C^\beta H_3$  bending vibration, only the localized mode with the lower wavenumber is considered on each residue. Both the coupling constants given by the matrix elements of  $\tilde{\Omega}$  and those calculated approximately from the interaction of the transition dipole moments of the localized modes (labeled TDC) are given. See text for details. All values in  $\text{cm}^{-1}$ .

	$\tilde{\Omega}$ matrix elements				TDC			
	8 $\leftrightarrow$ 9	8 $\leftrightarrow$ 10	8 $\leftrightarrow$ 11	8 $\leftrightarrow$ 12	8 $\leftrightarrow$ 9	8 $\leftrightarrow$ 10	8 $\leftrightarrow$ 11	8 $\leftrightarrow$ 12
Amide I	7.9	-2.3	-4.2	-0.4	10.2	-2.0	-7.2	-0.7
Amide II	3.3	-3.6	0.6	0.8	2.5	-0.8	0.3	0.5
$C^\beta H_3$ asymmetric bend	0.1	0.2	0.2	0.0	0.0	0.0	0.0	0.0
$C^\beta H_3$ symmetric bend	0.2	0.1	0.2	-0.0	-0.1	0.0	-0.1	0.0
$C^\alpha H$ bend (I)	1.1	0.3	-0.1	-0.3	0.0	0.0	0.0	-0.1
$C^\alpha H$ bend (II)	2.7	-0.8	1.3	-0.1	-0.3	0.0	0.0	-0.2
Amide III	7.5	1.0	-0.6	0.8	1.9	-0.3	-0.2	0.0
Skeletal $C^\alpha$ -N stretch	1.3	-0.8	-0.3	-0.5	0.4	-0.0	-0.1	-0.1

significant contributions of a N–H bending as well as of an asymmetric  $C^\beta H_3$  bending vibration. The amide III mode is an in-phase combination of the N–H bending and C–N stretching vibrations of the amide group, with some  $C^\alpha$ –H bending (approximately perpendicular to the  $C^\alpha$ –N bond) and asymmetric  $C^\beta H_3$  bending contributions. Finally, the skeletal  $C^\alpha$ –N stretching mode not only contains a significant amount of a  $C^\alpha$ –N stretching vibration but also has significant contributions of  $C^\alpha$ –H and asymmetric  $C^\beta H_3$  bending.

### C. Coupling constants

Additional information is provided by the coupling constants between the localized modes on different residues, i.e., the off-diagonal elements of the matrix  $\tilde{\Omega}$ . These describe the interaction between the localized modes, and therefore determine how the localized modes combine to normal modes. While for an analysis of the total intensity of a certain band the localized modes and their intensities are sufficient, a detailed description of the band shapes will require that these coupling constants are taken into consideration.

Similar to what was noted for the localized modes, where modes on different residues are similar and show similar vibrational frequencies and intensities, also the coupling constants are similar for localized modes on different residues and only depend on the distance between the coupled residues. Therefore, the coupling matrix  $\tilde{\Omega}$  has a rather simple structure, with almost equal coupling constants on the secondary diagonals. Deviations only occur for the terminal residues, for which deviations of the coupling constants are found. Therefore, it will usually be sufficient to consider only the couplings of one representative localized mode with the localized modes on its neighboring residues.

The coupling constants for the interaction of the localized mode on residue 8 with those on its neighboring residues are given in Table IV for all considered vibrational bands of  $(\text{Ala})_{20}$ . Only the couplings between localized modes that are separated by up to three residues are given, the magnitude of the coupling constants between localized modes that are further apart are in all cases smaller than  $1 \text{ cm}^{-1}$ . It should be noted that the sign of the coupling

constants depends on the choice of the (arbitrary) phase of the localized modes. We choose this phase such that in all cases, the nearest-neighbor coupling constants are positive, which in the case of the amide I vibrations leads to localized modes with the same phase of the carbonyl stretching vibration, but which results in localized modes with alternating phases in some other cases.

For the different bands, both the magnitudes of the coupling constants and the qualitative patterns (i.e., the relative sign and magnitudes of the couplings between different residues) are rather different. The largest coupling constants are found for the amide I band, where the largest couplings occur between nearest-neighbor residues, with a coupling constant of  $\tilde{\Omega}_{i,i+1} = 7.9 \text{ cm}^{-1}$ . For residues that are further apart, the coupling constants are negative. For the second nearest neighbor, the coupling constant is  $\tilde{\Omega}_{i,i+2} = -2.3 \text{ cm}^{-1}$ , which is significantly smaller than for the nearest neighbor, and for the third nearest neighbor, which corresponds to one complete turn of the  $\alpha$ -helix, the coupling constants increase again to  $\tilde{\Omega}_{i,i+3} = -4.3 \text{ cm}^{-1}$ , i.e., the larger third nearest-neighbor coupling reflects the  $\alpha$ -helical structure of the system under study. For residues that are more than three residues apart, the coupling constants are significantly smaller.

Also for the amide II band and the extended amide III region significant coupling constants are found, while the couplings are small for the skeletal  $C^\alpha$ –N stretching band and almost zero for the symmetric and asymmetric  $C^\beta H_3$  bending modes. Note the different qualitative coupling patterns found for the different bands. For instance, while for the amide I band the largest coupling constant is the nearest-neighbor one, with the other couplings being significantly smaller, for the amide II band the first and second nearest-neighbor coupling constants are of similar magnitude, but with opposite sign. This will lead to distinct differences in the normal modes and the intensity distributions within the bands.

While with our approach it is possible to obtain vibrational frequencies of localized modes and coupling constants for the interaction between them from full quantum chemical calculations, empirical methods for the calculation of vibrational spectra often proceed in the other direction by employ-

ing frequencies of local modes and coupling constants that are either determined empirically or from calculations on small systems. The most prominent example where such a model has been successfully applied is for the amide I band in polypeptides and proteins, where it is well established that the shape and position of the amide I band are largely influenced by the coupling between the individual carbonyl stretching vibrations. This was first recognized by Krimm and co-workers,<sup>57,58</sup> who showed that the splitting of the amide I band in  $\beta$ -sheet polypeptides can be explained by using a simple dipole coupling model. Such empirical TDC coupling constants have since then been used in numerous theoretical studies.<sup>58–60</sup> However, the TDC model neglects the effects of through-bond couplings, and it has been shown that for residues that are close to each other, the TDC model leads to significant deviations of the resulting vibrational frequencies.<sup>61,62</sup>

In contrast, our approach allows one to extract coupling constants from full quantum chemical calculations that include all these effects and that do not rely on any empirical models. It is therefore instructive to compare these coupling constants obtained as off-diagonal elements of  $\tilde{\mathbf{Q}}$ , which can be considered exact within the given computational methodology, to those one obtains by using a the TDC model,<sup>28</sup> i.e., by assuming that the coupling constants are determined by an interaction of the transition dipole moments of the localized modes. Within such a simple dipole approximation, the off-diagonal elements of  $\tilde{\mathbf{H}}^{\text{sub}}$  are given by

$$\tilde{H}_{pq}^{\text{sub}} = |\boldsymbol{\mu}_p||\boldsymbol{\mu}_q|X_{pq}, \quad (20)$$

where  $\boldsymbol{\mu}_p = (\partial\boldsymbol{\mu}/\partial\tilde{Q}_p^{\text{sub}})_0$  is the derivative of the dipole moment with respect to normal mode  $p$  (i.e., the transition dipole moment) and

$$X_{pq} = \frac{\mathbf{e}_p \cdot \mathbf{e}_q - 3(\mathbf{e}_p \cdot \mathbf{e}_{pq})(\mathbf{e}_q \cdot \mathbf{e}_{pq})}{R_{pq}^3} \quad (21)$$

is the orientational factor, which depends on the orientation of the two transition dipoles with respect to each other. In this expression  $\mathbf{e}_p$  and  $\mathbf{e}_q$  are the unit vectors in the directions of  $\boldsymbol{\mu}_p$  and  $\boldsymbol{\mu}_q$ , respectively,  $R_{pq}$  is the distance between the two transition dipoles, and  $\mathbf{e}_{pq}$  is the unit vector along the line connecting the two dipoles.

Studies applying such a TDC model have to apply values for the transition dipole vectors that have either been determined empirically or from calculations on model systems, and rely on a suitable parametrization of the location and direction of these transition dipoles.<sup>28,59</sup> In our case it is possible to extract these parameters directly from the localized modes. The transition dipole vectors  $\boldsymbol{\mu}_p$  and  $\boldsymbol{\mu}_q$  can be taken directly from the calculations, and we assume that these dipoles are located at the centers of the localized modes defined according to Eq. (7). The calculated off-diagonal elements of  $\tilde{\mathbf{H}}^{\text{sub}}$  can then be converted to coupling constants according to Eq. (15). These coupling constants then correspond to the pure dipole–dipole coupling between the localized modes.

The coupling constants calculated using the TDC model are included in Table IV. For the amide I band the TDC

coupling constants are roughly in qualitative agreement with the calculated off-diagonal elements of  $\tilde{\mathbf{Q}}$ . The largest coupling constant is found for the nearest-neighbor coupling, and the coupling constants for residues that are further apart are negative, with the third nearest-neighbor coupling being the largest in magnitude. However, there are significant quantitative differences between the exact coupling constants and the TDC values. In particular, the magnitude of the TDC coupling constants is in all cases too large, and the absolute value of the fourth-nearest-neighbor coupling constant is larger for the second nearest neighbor, while the exact value is smaller by a factor of almost 7. These differences can be mainly attributed to through-bond couplings, which are neglected in the TDC model. Note that in our calculations the magnitude of the transition dipole moment is approximately  $3.1 \text{ D } \text{\AA}^{-1} \text{ amu}^{-1/2}$  for the considered localized modes, while empirical TDC calculations for the amide I band usually use a smaller value of  $2.73 \text{ D } \text{\AA}^{-1} \text{ amu}^{-1/2}$ .<sup>59</sup> This lowers the absolute value of the calculated coupling constants and leads to a better agreement, i.e., the parameters used in empirical TDC calculations are scaled such that they partly account for through-bond couplings.

For the amide II band, the TDC coupling constants have the correct sign, but are qualitatively different from the exact ones. While for the latter, the second-nearest-neighbor coupling constant is the largest in magnitude, with the TDC model the nearest-neighbor coupling is larger in magnitude than the second nearest-neighbor one. For the extended amide III region as well as for the skeletal  $\text{C}^\alpha\text{--N}$  stretching band, the TDC coupling constants are significantly too small and even have the wrong sign in some cases. In addition to the neglect of through-bond couplings for these vibrations that are not dominated by a single stretching or bending vibration, also the neglect of higher transition multipole moments becomes significant.

## D. Comparison of atomic-contribution and distance localization criteria

The two localization criteria  $\xi_{\text{at}}$  [Eq. (5)] and  $\xi_{\text{dist}}$  [Eq. (9)] are defined in rather different ways as explained in the theory part of this work. The first minimizes—in analogy to the Pipek–Mezey orbital localization criterion—the sum of the squares of the atomic contributions to the localized modes, while the latter—in analogy to the Boys orbital localization criterion—maximizes the distance between the centers of the localized modes. Therefore, it is necessary to compare the localized modes resulting from these two different criteria.

To measure how similar the localized modes  $\tilde{Q}_p^{\text{sub,(at)}}$  and  $\tilde{Q}_p^{\text{sub,(dist)}}$  (obtained by localization according to the atomic-contribution and distance criteria, respectively) are, one can calculate the overlap of the two localized modes, which can be defined as<sup>63</sup>

$$O_p = \left( \sum_{i\alpha} \tilde{Q}_{i\alpha,p}^{\text{sub,(at)}} \tilde{Q}_{i\alpha,p}^{\text{sub,(dist)}} \right)^2, \quad (22)$$

where the square has to be taken since the phase of the modes is arbitrary. For identical modes, the overlap is 1,

TABLE V. Comparison of the localized modes obtained with the atomic-contribution and with the distance localization criteria. For each band in the vibrational spectrum of  $\alpha$ -helical (Ala)<sub>20</sub> the square root of the minimum overlap, the maximum absolute differences in the diagonal and off-diagonal element of the coupling matrix  $\tilde{\Omega}$ , and the maximum absolute differences in the IR absorption and Raman scattering factor with respect to the localized modes are listed.

Band	$\min \sqrt{O_p}$	$\max \Delta\tilde{\Omega}_{ii} $ (cm <sup>-1</sup> )	$\max \Delta\tilde{\Omega}_{ij} $ (cm <sup>-1</sup> )	$\max \Delta(\text{IR}) $ (km/mol)	$\max \Delta(\text{Raman}) $ (Å <sup>4</sup> /amu)
Amide I	0.9999	0.04	0.07	1.18	0.07
Amide II	0.9986	0.40	1.71	2.07	0.02
C <sup>β</sup> H <sub>3</sub> asymmetric bend	0.8952	1.06	2.26	10.18	1.02
C <sup>β</sup> H <sub>3</sub> symmetric bend	0.9999	0.00	0.03	0.04	0.01
C <sup>α</sup> H bend (I)	0.9999	0.02	0.17	0.33	0.09
C <sup>α</sup> H bend (II)	0.9988	0.25	0.43	0.59	0.35
Amide III	0.9982	1.16	0.85	0.92	0.13
Skeletal C <sup>α</sup> -N stretch	0.9994	0.18	0.95	1.41	0.16

while it is 0 for orthogonal modes. In addition to the overlap, we will also compare the wavenumbers of the localized modes (i.e., the diagonal elements of the coupling matrix  $\tilde{\Omega}$ ), the coupling constants (i.e., the off-diagonal elements of  $\tilde{\Omega}$ ), as well as the IR absorption and Raman scattering factor with respect to the localized modes.

In Table V, the localized modes obtained according to the two different localization criteria for each considered band in the vibrational spectra of  $\alpha$ -helical (Ala)<sub>20</sub> are compared. For each of these bands, the minimum overlap of the localized modes as well as the maximum absolute differences in the diagonal and off-diagonal elements of  $\tilde{\Omega}$  and in the IR absorption and Raman scattering factors with respect to the localized modes are shown.

For all bands except for C<sup>β</sup>H<sub>3</sub> asymmetric bending, which will be considered separately below, the square root of the overlap of the localized modes obtained according to the two different localization criteria is larger than 0.99, i.e., the calculated localized modes are almost identical. This can also be seen from the differences in the wavenumbers of the localized modes, which are in all cases smaller than 1.1 cm<sup>-1</sup>. Furthermore, the differences in the IR absorption are smaller than 2.1 km/mol, and those in the Raman scattering factor smaller than 0.4 Å<sup>4</sup>/amu, which is a relative error of less than 1%. The coupling constants are somewhat more sensitive to the differences in the localized modes, with a maximum difference of 1.7 cm<sup>-1</sup> for the amide II band and of up to 1.0 cm<sup>-1</sup> for the other bands.

For the C<sup>β</sup>H<sub>3</sub> asymmetric bending vibrations, the overlap of the localized modes obtained according to the two different localization criteria is only 0.90, and also for the IR absorption, the Raman scattering factors, and the coupling constants larger errors than for the other bands are found. However, these can be easily understood, since there are two almost degenerate asymmetric C<sup>β</sup>H<sub>3</sub> bending modes on the side chain of each residue [cf. Figs. 3(c) and 3(d)]. This is similar to what is known from orbital localization. While Pipek–Mezey-localized orbitals preserve the  $\sigma$ -/ $\pi$ -separation for lone pairs and double bonds, the Boys-localized orbitals show “rabbit ears” and “banana bonds,” respectively.<sup>64</sup>

In summary, both the atomic-contribution and distance localization criteria yield very similar localized modes for cases in which the normal modes are combinations of similar vibrations on different residues, i.e., the resulting localized modes are similar modes located on different residues. Differences between the two criteria only occur when localizing modes that involve the same atoms within one residue, such as it is the case for the asymmetric C<sup>β</sup>H<sub>3</sub> bending modes. In these cases, we consider the atomic-contribution localization criterion more suitable, since—similar as for orbital localization—it tends to produce localized modes that resemble the normal modes calculated for a single residue more closely.

## V. CONCLUSIONS

We have developed a methodology for the analysis of calculated vibrational spectra of polypeptides and proteins in terms of localized modes. These localized modes are obtained by determining the unitary transformation of the normal modes within one band of the spectrum that maximizes a suitably defined localization criterion. For this localization criterion, two different choices  $\xi_{\text{at}}$  and  $\xi_{\text{dist}}$  have been considered, and it is found that for normal modes that arise as delocalized combinations of vibrations on different residues the resulting localized modes are not sensitive to the chosen localization criterion. Hence, there is often no need to actually differentiate between the atomic contributions and distance localized modes. Instead, we may choose one of the localization procedures and call the resulting modes simply localized modes.

While in the present work we focused on the calculation of localized modes, it is also possible to determine the unitary transformation that is optimal with respect to another criterion. This could, for instance, be used to extract special modes such as the low-frequency helical breathing mode considered in Ref. 56, which are not directly accessible otherwise.

The localized modes have, even though they have no direct physical significance, a number of features that make them more suited for the analysis of vibrational spectra than the normal modes. First, the localized modes are limited to a

specific part of the system and to a small number of atoms, in contrast to the normal modes, which are usually delocalized over the whole system, limited to a specific part of the system and to a small number of atoms. This simplifies their visualization and allows it to easily investigate the vibrations contributing to a certain band in the vibrational spectrum. Second, for the normal modes within one band of the vibrational spectrum of a polypeptide or protein, the localized modes on the different amino acid residues will be very similar vibrations. Therefore, for many purposes it will be sufficient to look at only one representative localized mode for each band found in the spectrum if one wants to analyze the vibrational frequencies at which certain bands appear in the spectrum or shifts in these vibrational frequencies caused by structural changes.

As we have shown, the total intensity of a band in the IR or Raman spectrum is invariant under a unitary transformation of the contributing normal modes and, therefore, it can equally well be analyzed in terms of the intensities of the localized modes. Since these will, in general, all show similar contributions, it is sufficient to consider the intensity of only one localized mode in order to study the factors determining the total intensities of certain bands. The same will be true for VCD and ROA spectroscopies, even though this was not exploited in the present work but is in progress in our laboratory. While the total intensity of a band in the IR and Raman spectra is given by the sum of the intensities of the localized modes, the precise band shape is determined by the couplings between the localized modes, for which coupling constants can be extracted from full quantum chemical calculations using the methodology presented here.

While traditionally vibrational spectra of polypeptides and proteins are interpreted in terms of local modes and local mode vibrational frequencies determined for small model systems, we have shown how such parameters can be extracted from full quantum chemical calculations. These could then possibly be used to parametrize improved empirical models. The same holds for the coupling constants describing the interaction between localized modes, for which mostly the TDC model is applied. However, as we have shown for the example of  $\alpha$ -helical (Ala)<sub>20</sub>, such empirical coupling constants are in many cases not even qualitatively correct.

The analysis of calculated vibrational spectra of polypeptides and proteins in terms of localized modes will make it possible to analyze the influence of secondary structure changes on the vibrational spectra in detail. However, such an investigation is beyond the scope of the present work, and a study of the IR and Raman spectra of helical alanine polypeptides will be presented elsewhere.<sup>65</sup>

## ACKNOWLEDGMENTS

C.R.J. acknowledges funding by a Rubicon scholarship of the Netherlands Organization for Scientific Research (NWO). This work has been supported by the Swiss National Science Foundation SNF (Project No. 200020-121870). The authors would like to thank Sandra Lubner for help with the (Ala)<sub>20</sub> calculations and for valuable discussions.

## APPENDIX A: EQUIVALENCE OF $\xi'_{\text{dist}}$ AND $\xi_{\text{dist}}$

In this appendix we show that the unitary transformation  $U$  that maximizes  $\xi'_{\text{dist}}(\mathbf{Q}^{\text{sub}}U)$  also maximizes  $\xi_{\text{dist}}(\mathbf{Q}^{\text{sub}}U)$ , i.e., the two localization criteria are equivalent. First,  $\xi'_{\text{dist}}(\mathbf{Q}^{\text{sub}}U) = \xi'_{\text{dist}}(\tilde{\mathbf{Q}}^{\text{sub}})$  can be rewritten as

$$\xi'_{\text{dist}}(\tilde{\mathbf{Q}}^{\text{sub}}) = \sum_{p=1}^k \sum_{q=1}^k (\mathbf{R}_p^{\text{center}} - \mathbf{R}_q^{\text{center}})^2 \quad (\text{A1})$$

$$= \sum_{p=1}^k \sum_{q=1}^k [(\mathbf{R}_p^{\text{center}})^2 - 2\mathbf{R}_p^{\text{center}}\mathbf{R}_q^{\text{center}} + (\mathbf{R}_q^{\text{center}})^2] \quad (\text{A2})$$

$$= k \sum_{p=1}^k (\mathbf{R}_p^{\text{center}})^2 + k \sum_{q=1}^k (\mathbf{R}_q^{\text{center}})^2 - 2 \sum_{p=1}^k \sum_{q=1}^k \mathbf{R}_p^{\text{center}}\mathbf{R}_q^{\text{center}} \quad (\text{A3})$$

$$= 2k \sum_{p=1}^k (\mathbf{R}_p^{\text{center}})^2 - 2 \left( \sum_{p=1}^k \mathbf{R}_p^{\text{center}} \right)^2 \quad (\text{A4})$$

$$= 2k \sum_{p=1}^k (\mathbf{R}_p^{\text{center}})^2 - 2 \left[ \sum_{\beta} \left( \sum_{p=1}^k R_{p,\beta}^{\text{center}} \right)^2 \right]. \quad (\text{A5})$$

The maximization of the first term is equivalent to the maximization of  $\xi_{\text{dist}}$ , while the second term is invariant under unitary transformations of the modes because

$$\xi'^{(\beta)}_{\text{dist}}(\tilde{\mathbf{Q}}^{\text{sub}}) = \xi'^{(\beta)}_{\text{dist}}(\mathbf{Q}^{\text{sub}}U) = \sum_p \mathbf{R}_{p,\beta}^{\text{center}} = \sum_p \sum_i R_{i\beta} \sum_{\alpha} (\tilde{Q}_{i\alpha,p}^{\text{sub}})^2 \quad (\text{A6})$$

$$= \sum_p \sum_{rs} U_{pr} U_{ps} \sum_i R_{i\beta} \sum_{\alpha} Q_{i\alpha,r}^{\text{sub}} Q_{i\alpha,s}^{\text{sub}} \quad (\text{A7})$$

$$= \sum_{rs} \left( \sum_p U_{pr} U_{ps} \right) \sum_i R_{i\beta} \sum_{\alpha} Q_{i\alpha,r}^{\text{sub}} Q_{i\alpha,s}^{\text{sub}} \quad (\text{A8})$$

$$= \sum_{rs} \delta_{rs} \sum_i R_{i\beta} \sum_{\alpha} Q_{i\alpha,r}^{\text{sub}} Q_{i\alpha,s}^{\text{sub}} \quad (\text{A9})$$

$$= \sum_r \sum_i R_{i\beta} \sum_{\alpha} (Q_{i\alpha,r}^{\text{sub}})^2 = \xi'^{(\beta)}_{\text{dist}}(\mathbf{Q}^{\text{sub}}), \quad (\text{A10})$$

is invariant under unitary transformations. This can also be seen by defining a matrix  $\mathbf{K}^{(\beta)}$  as

$$\mathbf{K}_{pq}^{(\beta)} = \sum_i R_{i\beta} \sum_{\alpha} Q_{i\alpha,p}^{\text{sub}} Q_{i\alpha,q}^{\text{sub}} \quad (\text{A11})$$

Then we have

$$\xi'^{(\beta)}_{\text{dist}}(\mathbf{Q}^{\text{sub}}) = \text{Tr} \mathbf{K}^{(\beta)}, \quad (\text{A12})$$

$$\xi_{\text{dist}}^{(\beta)}(\tilde{\mathcal{Q}}^{\text{sub}}) = \xi_{\text{dist}}^{(\beta)}(\mathcal{Q}^{\text{sub}}\mathbf{U}) = \text{Tr}(\mathbf{U}^T \mathbf{K}^{(\beta)} \mathbf{U}), \quad (\text{A13})$$

and  $\xi_{\text{dist}}^{(\beta)}(\mathcal{Q}^{\text{sub}}) = \xi_{\text{dist}}^{(\beta)}(\mathcal{Q}^{\text{sub}}\mathbf{U})$  because the trace of a matrix is invariant under unitary transformations.

## APPENDIX B: JACOBI-SWEEP METHOD FOR THE CALCULATION OF LOCALIZED MODES

The unitary transformation that—according to a given localization measure—yields the optimally localized modes, i.e., the matrix  $\mathbf{U}$  that maximizes the function  $\xi(\mathcal{Q}^{\text{sub}}\mathbf{U})$ , can be determined by using so-called Jacobi sweeps, as proposed by Edmiston and Ruedenberg<sup>37</sup> for the determination of localized orbitals.

The transformation matrix  $\mathbf{U}$  is expressed as consecutive Jacobi rotations,

$$\mathbf{U} = \mathbf{U}^{(1)}\mathbf{U}^{(2)} \dots \mathbf{U}^{(m)}, \quad (\text{B1})$$

where  $\mathbf{U}^{(i)}$  are two by two rotations (among modes  $p$  and  $q$ ) with the entries

$$U_{pp}^{(i)} = U_{qq}^{(i)} = \cos \gamma, \quad (\text{B2})$$

$$U_{pq}^{(i)} = -U_{qp}^{(i)} = \sin \gamma, \quad (\text{B3})$$

$$U_{rs}^{(i)} = \delta_{rs} \quad \text{for all other entries.} \quad (\text{B4})$$

The optimal transformation is then built up by performing sweeps over all pairs of modes and choosing the optimal rotation angle  $\gamma$  in each step. Hence, the rotation matrix  $\mathbf{U}^{(i)}$  is determined such that  $\xi(\mathcal{Q}^{(i-1)}\mathbf{U}^{(i)})$  is maximized, where  $\mathcal{Q}^{(i-1)} = \mathcal{Q}^{\text{sub}}(\mathbf{U}^{(1)}\mathbf{U}^{(2)} \dots \mathbf{U}^{(i-1)})$  are the (partially localized) modes after  $(i-1)$  Jacobi rotations. This procedure is repeated until a Jacobi sweep does not lead to a further increase in  $\xi$ .

For the two considered localization measures  $\xi_{\text{at}}$  and  $\xi_{\text{dist}}$ , the function  $\xi(\mathcal{Q}^{(i-1)}\mathbf{U}^{(i)})$  can be expressed in the form

$$\xi(\mathcal{Q}^{(i-1)}\mathbf{U}^{(i)}) = \sum_r \sum_{stuv} U_{sr}^{(i)} U_{tr}^{(i)} U_{ur}^{(i)} U_{vr}^{(i)} K_{stuv}^{(i-1)}, \quad (\text{B5})$$

where

$$K_{\text{stuv}}^{\text{at}(i-1)} = \sum_{k=1}^n \left( \sum_{\alpha=x,y,z} Q_{k\alpha,s}^{(i-1)} Q_{k\alpha,t}^{(i-1)} \right) \left( \sum_{\alpha=x,y,z} Q_{k\alpha,u}^{(i-1)} Q_{k\alpha,v}^{(i-1)} \right), \quad (\text{B6})$$

$$K_{\text{stuv}}^{\text{dist}(i-1)} = \left( \sum_{k=1}^n \mathbf{R}_i \sum_{\alpha=x,y,z} Q_{k\alpha,s}^{(i-1)} Q_{k\alpha,t}^{(i-1)} \right) \times \left( \sum_{k=1}^n \mathbf{R}_i \sum_{\alpha=x,y,z} Q_{k\alpha,u}^{(i-1)} Q_{k\alpha,v}^{(i-1)} \right). \quad (\text{B7})$$

Utilizing that  $\mathbf{U}^{(i)}$  is a Jacobi rotation which mixes modes  $p$  and  $q$  with rotation angle  $\gamma$ , Eq. (B5) simplifies to

$$\xi(\mathcal{Q}^{(i-1)}\mathbf{U}^{(i)}) = \xi(\mathcal{Q}^{(i-1)}) + A_{pq} - A_{pq} \cos 4\gamma + B_{pq} \sin 4\gamma, \quad (\text{B8})$$

with

$$\xi(\mathcal{Q}^{(i-1)}) = \sum_r K_{rrrr}^{(i-1)} \quad (\text{B9})$$

and

$$A_{pq} = K_{ppqq}^{(i-1)} - \frac{1}{4} [K_{pppp}^{(i-1)} + K_{qqqq}^{(i-1)} - 2K_{ppqq}^{(i-1)}], \quad (\text{B10})$$

$$B_{pq} = K_{ppqp}^{(i-1)} - K_{qqqp}^{(i-1)}. \quad (\text{B11})$$

If one now defines the angle  $\alpha$  with  $-\pi/4 \leq \alpha < \pi/4$  by the two equations

$$\sin 4\alpha = \frac{B_{pq}}{\sqrt{A_{pq}^2 + B_{pq}^2}}, \quad (\text{B12})$$

$$\cos 4\alpha = \frac{-A_{pq}}{\sqrt{A_{pq}^2 + B_{pq}^2}}, \quad (\text{B13})$$

one obtains

$$\xi(\mathcal{Q}^{(i-1)}\mathbf{U}^{(i)}) = \xi(\mathcal{Q}^{(i-1)}) + A_{pq} + \sqrt{A_{pq}^2 + B_{pq}^2} \cos 4(\gamma - \alpha), \quad (\text{B14})$$

which is maximized for  $\gamma = \alpha$ . Note that both Eqs. (B12) and (B13) are needed in order to define the angle  $\alpha$  uniquely, since either of the two equations has two solutions between  $-\pi/4$  and  $\pi/4$ .

## APPENDIX C: APPROXIMATE RELATIONS BETWEEN $\tilde{\mathbf{H}}^{\text{sub}}$ AND $\tilde{\mathbf{\Omega}}$

The Hessian matrix with respect to localized modes  $\tilde{\mathbf{H}}^{\text{sub}}$  and the coupling matrix  $\tilde{\mathbf{\Omega}}$  are related by [cf. Eq. (12)],

$$\tilde{\mathbf{H}}^{\text{sub}} = 4\pi^2 \tilde{\mathbf{\Omega}}^2, \quad (\text{C1})$$

and it follows for the diagonal elements

$$\tilde{H}_{ii}^{\text{sub}} = 4\pi^2 \sum_k \tilde{\Omega}_{ik}^2 \quad (\text{C2})$$

$$= 4\pi^2 \left( \tilde{\Omega}_{ii}^2 + \sum_{k \neq i} \tilde{\Omega}_{ik}^2 \right) \quad (\text{C3})$$

$$= 4\pi^2 \tilde{\Omega}_{ii}^2 \left( 1 + \frac{1}{4\pi^2} \sum_{k \neq i} \frac{\tilde{\Omega}_{ik}^2}{\tilde{\Omega}_{ii}^2} \right). \quad (\text{C4})$$

If we assume that  $\tilde{\Omega}_{ik} \ll \tilde{\Omega}_{ii}$  (for  $k \neq i$ ), which usually holds if one considers the localized modes for one band in the vibrational spectrum, i.e., if the differences between the vibration frequencies of the considered normal modes is small, the second term is negligible and we get

$$\tilde{H}_{ii}^{\text{sub}} \approx 4\pi^2 \tilde{\Omega}_{ii}^2 \quad (\text{C5})$$

from which Eq. (14) immediately follows.

For the off-diagonal elements we have

$$\tilde{H}_{ij}^{\text{sub}} = 4\pi^2 \sum_k \tilde{\Omega}_{ik} \tilde{\Omega}_{jk} \quad (\text{C6})$$

$$= 4\pi^2 \left( \tilde{\Omega}_{ij}(\tilde{\Omega}_{ii} + \tilde{\Omega}_{jj}) + \sum_{k \neq i,j} \tilde{\Omega}_{ik}\tilde{\Omega}_{jk} \right), \quad (C7)$$

and it follows

$$\frac{\tilde{H}_{ij}^{\text{sub}}}{4\pi^2(\tilde{\Omega}_{ii} + \tilde{\Omega}_{jj})} = \tilde{\Omega}_{ij} + \sum_{k \neq i,j} \frac{\tilde{\Omega}_{ik}\tilde{\Omega}_{jk}}{4\pi^2(\tilde{\Omega}_{ii} + \tilde{\Omega}_{jj})}, \quad (C8)$$

where under the assumptions made above, the second term on the right-hand side is very small compared to the first one, so that

$$\tilde{\Omega}_{ij} \approx \frac{\tilde{H}_{ij}}{4\pi^2(\tilde{\Omega}_{ii} + \tilde{\Omega}_{jj})} \quad (C9)$$

$$\approx \frac{\tilde{H}_{ij}}{4\pi^2(\tilde{v}_i + \tilde{v}_j)}, \quad (C10)$$

which is the same as Eq. (15).

#### APPENDIX D: VIBRATIONAL INTENSITIES IN IR, RAMAN, VCD, AND ROA SPECTROSCOPIES

In IR spectroscopy, the absorption of the  $p$ th vibrational mode is proportional to the square of the derivative of the dipole moment with respect to the corresponding normal mode,<sup>19</sup>

$$A_p^{\text{IR}} \propto \left( \frac{\partial \boldsymbol{\mu}}{\partial Q_p} \right)_0^2 = \sum_{\alpha} \left( \frac{\partial \mu_{\alpha}}{\partial Q_p} \right)_0^2. \quad (D1)$$

The rotational strengths, which determine the differences in the absorption of left- and right-circularly polarized lights that are measured in VCD spectroscopy, are given by the imaginary part of the products of electric and magnetic transition dipole moments. Within the usual approximations,<sup>66,67</sup> the measured difference in absorption is proportional to the product of the derivative of the electric dipole moment with respect to the normal mode and the imaginary part of the derivative of the magnetic dipole moment with respect to the normal mode velocity,<sup>1</sup>

$$R_p^{\text{VCD}} \propto \text{Im} \left[ \left( \frac{\partial \boldsymbol{\mu}}{\partial Q_p} \right)_0 \left( \frac{\partial \mathbf{m}}{\partial \dot{Q}_p} \right)_0 \right] = i \sum_{\alpha} \left( \frac{\partial \mu_{\alpha}}{\partial Q_p} \right)_0 \left( \frac{\partial m_{\alpha}}{\partial \dot{Q}_p} \right)_0. \quad (D2)$$

The scattering intensity measured in Raman spectroscopy is proportional to a linear combination of the Raman invariants,<sup>19,68</sup>

$$I_p^{\text{Raman}} \propto [c_1 a_p^2 + c_2 \gamma_p^2], \quad (D3)$$

where the coefficients  $c_1$  and  $c_2$  depend on the geometries of the scattering experiment. The term on the right-hand side of the above expression is the so-called scattering factor  $S$ . The isotropic and anisotropic Raman invariants are given by

$$a_p^2 = \frac{1}{9} \left[ \left( \frac{\partial \alpha_{xx}}{\partial Q_p} \right)_0 + \left( \frac{\partial \alpha_{yy}}{\partial Q_p} \right)_0 + \left( \frac{\partial \alpha_{zz}}{\partial Q_p} \right)_0 \right]^2 \\ = \frac{1}{9} \sum_{\alpha\beta} \left( \frac{\partial \alpha_{\alpha\alpha}}{\partial Q_p} \right)_0 \left( \frac{\partial \alpha_{\beta\beta}}{\partial Q_p} \right)_0 \quad (D4)$$

$$\gamma_p^2 = \frac{1}{2} \sum_{\alpha\beta} \left[ 3 \left( \frac{\partial \alpha_{\alpha\beta}}{\partial Q_p} \right)_0^2 - \left( \frac{\partial \alpha_{\alpha\alpha}}{\partial Q_p} \right)_0 \left( \frac{\partial \alpha_{\beta\beta}}{\partial Q_p} \right)_0 \right]. \quad (D5)$$

The intensity differences measured in ROA spectroscopy are proportional to a linear combination of the ROA invariants,<sup>41</sup>

$$I_p^{\text{ROA}} = (I_R^{\text{Raman}} - I_L^{\text{Raman}})_p \\ \propto [c_1 \alpha G' + c_2 \beta (G')^2 + c_3 \beta (A)^2], \quad (D6)$$

where again the coefficients  $c_i$  depend on the experimental setup. The ROA invariants are given by

$$\alpha G' = \frac{1}{9} \sum_{\alpha\beta} \left( \frac{\partial \alpha_{\alpha\beta}}{\partial Q_p} \right)_0 \left( \frac{\partial G'_{\alpha\beta}}{\partial Q_p} \right)_0 \quad (D7)$$

$$\beta (G')^2 = \frac{1}{2} \sum_{\alpha\beta} \left[ 3 \left( \frac{\partial \alpha_{\alpha\beta}}{\partial Q_p} \right)_0 \left( \frac{\partial G'_{\alpha\beta}}{\partial Q_p} \right)_0 - \left( \frac{\partial \alpha_{\alpha\alpha}}{\partial Q_p} \right)_0 \left( \frac{\partial G'_{\beta\beta}}{\partial Q_p} \right)_0 \right] \quad (D8)$$

$$\beta (A)^2 = \omega_L \sum_{\alpha\beta\gamma\delta} \epsilon_{\alpha\gamma\delta} \left( \frac{\partial \alpha_{\alpha\beta}}{\partial Q_p} \right)_0 \left( \frac{\partial A_{\gamma\beta\delta}}{\partial Q_p} \right)_0, \quad (D9)$$

where  $\omega_L$  is the angular frequency of the incident laser beam and  $\epsilon_{\alpha\gamma\delta}$  is an element of the antisymmetric Levi-Civita tensor.

As can be seen from these expressions, the IR absorption and the Raman scattering intensity as well as the intensity differences measured in VCD and ROA spectroscopy can all be given in the form of Eq. (16).

#### APPENDIX E: INVARIANCE OF THE SUM OF VIBRATIONAL INTENSITIES UNDER UNITARY TRANSFORMATIONS

As shown in Sec. II C, the total intensity of a band in the vibrational spectrum depends on terms of the form

$$L_l = \sum_{p=1}^k \left( \frac{\partial P_l^{(1)}}{\partial Q_p^{\text{sub}}} \right)_0 \left( \frac{\partial P_l^{(2)}}{\partial Q_p^{\text{sub}}} \right)_0 \quad (E1)$$

$$= \sum_{p=1}^k \sum_{i\alpha} \sum_{j\beta} Q_{i\alpha,p}^{\text{sub},(c)} Q_{j\beta,p}^{\text{sub},(c)} \left( \frac{\partial P_l^{(1)}}{\partial R_{i\alpha}} \right)_0 \left( \frac{\partial P_l^{(2)}}{\partial R_{j\beta}} \right)_0, \quad (E2)$$

where in the second line the derivatives are taken either with respect to the Cartesian displacement  $R_{i\alpha}$  or with respect to the Cartesian component of the velocity  $\dot{R}_{i\alpha}$ . By defining the matrix  $\mathbf{K}^{(l)}$  as

$$K_{pq}^{(l)} = \sum_{i\alpha} \sum_{j\beta} Q_{i\alpha,p}^{\text{sub},(c)} Q_{j\beta,q}^{\text{sub},(c)} \left( \frac{\partial P_l^{(1)}}{\partial R_{i\alpha}} \right)_0 \left( \frac{\partial P_l^{(2)}}{\partial R_{j\beta}} \right)_0, \quad (E3)$$

this can be expressed as

$$L_l = \text{Tr} \mathbf{K}^{(l)}. \quad (E4)$$

With respect to the transformed modes we have

$$\tilde{L}_l = \sum_{p1}^k \left( \frac{\partial P_l^{(1)}}{\partial \tilde{Q}_p^{\text{sub},(c)}} \right)_0 \left( \frac{\partial P_l^{(2)}}{\partial \tilde{Q}_p^{\text{sub},(c)}} \right)_0 \quad (\text{E5})$$

$$= \sum_{p=1}^k \sum_{i\alpha} \sum_{j\beta} \sum_{rs} U_{pr} U_{ps} Q_{i\alpha,r}^{\text{sub},(c)} Q_{j\beta,s}^{\text{sub},(c)} \left( \frac{\partial P_l^{(1)}}{\partial R_{i\alpha}} \right)_0 \left( \frac{\partial P_l^{(2)}}{\partial R_{i\alpha}} \right)_0 \quad (\text{E6})$$

$$= \text{Tr}(\mathbf{U}^T \mathbf{K}^{(l)} \mathbf{U}) = \text{Tr} \mathbf{K}^{(l)} = L_l, \quad (\text{E7})$$

where in the last line we used that the trace of a matrix is invariant under unitary transformations.

- <sup>1</sup>C. Herrmann and M. Reiher, *Top. Curr. Chem.* **268**, 85 (2007).
- <sup>2</sup>P. R. Carey, *J. Biol. Chem.* **274**, 26625 (1999).
- <sup>3</sup>G. J. Thomas, Jr., *Biopolymers* **67**, 214 (2002).
- <sup>4</sup>E. Vass, M. Hollosi, F. Besson, and R. Buchet, *Chem. Rev. (Washington, D.C.)* **103**, 1917 (2003).
- <sup>5</sup>A. Barth, *Biochim. Biophys. Acta, Bioenerg.* **1767**, 1073 (2007).
- <sup>6</sup>T. A. Keiderling, in *Circular Dichroism: Principles and Applications*, 2nd ed., edited by N. Berova, K. Nakanishi, and R. W. Woody (Wiley-VCH, New York, 2000), pp. 621–666.
- <sup>7</sup>F. Zhu, N. W. Isaacs, L. Hecht, and L. D. Barron, *Structure* **13**, 1409 (2005).
- <sup>8</sup>R. A. G. D. Silva, J. Kubelka, P. Bour, S. M. Decatur, and T. A. Keiderling, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8318 (2000).
- <sup>9</sup>J. Kubelka, R. A. G. D. Silva, and T. A. Keiderling, *J. Am. Chem. Soc.* **124**, 5325 (2002).
- <sup>10</sup>I. H. McColl, E. W. Blanch, A. C. Gill, A. G. O. Rhie, M. A. Ritchie, L. Hecht, K. Nielsen, and L. D. Barron, *J. Am. Chem. Soc.* **125**, 10019 (2003).
- <sup>11</sup>I. H. McColl, E. W. Blanch, L. Hecht, N. R. Kallenbach, and L. D. Barron, *J. Am. Chem. Soc.* **126**, 5076 (2004).
- <sup>12</sup>I. H. McColl, E. W. Blanch, L. Hecht, and L. D. Barron, *J. Am. Chem. Soc.* **126**, 8181 (2004).
- <sup>13</sup>J. Kapitán, V. Baumruk, and P. Bouř, *J. Am. Chem. Soc.* **128**, 2438 (2006).
- <sup>14</sup>J. Kapitán, F. Zhu, L. Hecht, J. Gardiner, D. Seebach, and L. D. Barron, *Angew. Chem.* **120**, 6492 (2008).
- <sup>15</sup>Ch. R. Jacob, S. Luber, and M. Reiher, *ChemPhysChem* **9**, 2177 (2008).
- <sup>16</sup>C. Ochsenfeld, J. Kussmann, and D. S. Lambrecht, in *Reviews in Computational Chemistry* (Wiley-VCH, New York, 2007), Vol. 23, pp. 1–82.
- <sup>17</sup>J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, *Comput. Phys. Commun.* **167**, 103 (2005).
- <sup>18</sup>Ch. R. Jacob and L. Visscher, *J. Chem. Phys.* **128**, 155102 (2008).
- <sup>19</sup>J. Neugebauer, M. Reiher, C. Kind, and B. A. Hess, *J. Comput. Chem.* **23**, 895 (2002).
- <sup>20</sup>M. Reiher and J. Neugebauer, *J. Chem. Phys.* **118**, 1634 (2003).
- <sup>21</sup>C. Herrmann, J. Neugebauer, and M. Reiher, *New J. Chem.* **31**, 818 (2007).
- <sup>22</sup>K. Kiewisch, J. Neugebauer, and M. Reiher, *J. Chem. Phys.* **129**, 204103 (2008).
- <sup>23</sup>S. Luber, J. Neugebauer, and M. Reiher, *J. Chem. Phys.* **130**, 064105 (2009).
- <sup>24</sup>P. Bouř and T. A. Keiderling, *J. Chem. Phys.* **117**, 4126 (2002).
- <sup>25</sup>A. Ghysels, D. Van Neck, V. Van Speybroeck, T. Verstraelen, and M. Waroquier, *J. Chem. Phys.* **126**, 224102 (2007).
- <sup>26</sup>P. Bouř, J. Sopková, L. Bednárová, P. Maloň, and T. A. Keiderling, *J. Comput. Chem.* **18**, 646 (1997).
- <sup>27</sup>M. Diem, *Introduction to Modern Vibrational Spectroscopy* (Wiley-

- Interscience, New York, 1993).
- <sup>28</sup>S. Krimm and J. Bandekar, *Adv. Protein Chem.* **38**, 181 (1986).
- <sup>29</sup>R. Schweitzer-Stenner, *J. Raman Spectrosc.* **32**, 711 (2001).
- <sup>30</sup>S. Ham, S. Cha, J.-H. Choi, and M. Cho, *J. Chem. Phys.* **119**, 1451 (2003).
- <sup>31</sup>J. Pipek and P. G. Mezey, *J. Chem. Phys.* **90**, 4916 (1989).
- <sup>32</sup>E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra* (Dover, New York, 1980).
- <sup>33</sup>W. Hug, *Chem. Phys.* **264**, 53 (2001).
- <sup>34</sup>W. Hug and J. Haesler, *Int. J. Quantum Chem.* **104**, 695 (2005).
- <sup>35</sup>S. F. Boys, *Rev. Mod. Phys.* **32**, 296 (1960).
- <sup>36</sup>J. M. Foster and S. F. Boys, *Rev. Mod. Phys.* **32**, 300 (1960).
- <sup>37</sup>C. Edmiston and K. Ruedenberg, *Rev. Mod. Phys.* **35**, 457 (1963).
- <sup>38</sup>J. E. Subotnik, Y. Shao, W. Liang, and M. Head-Gordon, *J. Chem. Phys.* **121**, 9220 (2004).
- <sup>39</sup>J. Neugebauer, *J. Phys. Chem. B* **112**, 2207 (2008).
- <sup>40</sup>M. Reiher, J. Neugebauer, and B. A. Hess, *Z. Phys. Chem.* **217**, 91 (2003).
- <sup>41</sup>L. D. Barron, *Molecular Light Scattering and Optical Activity*, 2nd ed. (Cambridge University Press, Cambridge, 2004).
- <sup>42</sup>R. Ahlrichs *et al.*, TURBOMOLE, <http://www.turbomole.com>.
- <sup>43</sup>R. Ahlrichs, M. Bär, M. Häser, H. Horn, and Ch. Kölmel, *Chem. Phys. Lett.* **162**, 165 (1989).
- <sup>44</sup>A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).
- <sup>45</sup>J. P. Perdew, *Phys. Rev. B* **33**, 8822 (1986).
- <sup>46</sup>A. Schäfer, C. Huber, and R. Ahlrichs, *J. Chem. Phys.* **100**, 5829 (1994).
- <sup>47</sup>TURBOMOLE basis set library, <ftp://ftp.chemie.uni-karlsruhe.de/pub/basen>.
- <sup>48</sup>K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs, *Chem. Phys. Lett.* **240**, 283 (1995).
- <sup>49</sup>TURBOMOLE auxiliary basis sets, <ftp://ftp.chemie.uni-karlsruhe.de/pub/jbasen>.
- <sup>50</sup>J. Neugebauer, C. Herrmann, S. Luber, and M. Reiher, SNF 4.0, A program for the quantum chemical calculation of vibrational spectra, <http://www.theochem.ethz.ch/software/snf>.
- <sup>51</sup>T. Oliphant *et al.*, NUMPY, A python library for numerical computations, <http://www.scipy.org/NumPy>.
- <sup>52</sup>The OPEN BABEL package, <http://openbabel.sourceforge.net/>.
- <sup>53</sup>R. Guha, M. T. Howard, G. R. Hutchison, P. Murray-Rust, H. Rzepa, C. Steinbeck, J. Wegner, and E. L. Willighagen, *J. Chem. Inf. Model.* **46**, 991 (2006).
- <sup>54</sup>JMOL, An open-source molecule viewer, <http://jmol.sourceforge.net>.
- <sup>55</sup>MATPLOTLIB, A python 2D plotting library, <http://matplotlib.sourceforge.net/>.
- <sup>56</sup>C. Herrmann, K. Ruud, and M. Reiher, *ChemPhysChem* **7**, 2189 (2006).
- <sup>57</sup>S. Krimm and Y. Abe, *Proc. Natl. Acad. Sci. U.S.A.* **69**, 2788 (1972).
- <sup>58</sup>W. H. Moore and S. Krimm, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 4933 (1975).
- <sup>59</sup>H. Torii and M. Tasumi, *J. Chem. Phys.* **96**, 3379 (1992).
- <sup>60</sup>P. Hamm, M. Lim, and R. M. Hochstrasser, *J. Phys. Chem. B* **102**, 6123 (1998).
- <sup>61</sup>S. Cha, S. Ham, and M. Cho, *J. Chem. Phys.* **117**, 740 (2002).
- <sup>62</sup>J. Kubelka, J. Kim, P. Bour, and T. A. Keiderling, *Vib. Spectrosc.* **42**, 63 (2006).
- <sup>63</sup>W. Hug and M. Fedorovsky, *Theor. Chem. Acc.* **119**, 113 (2008).
- <sup>64</sup>F. Jensen, *Introduction to Computational Chemistry*, 2nd ed. (Wiley, Chichester, 2007).
- <sup>65</sup>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* (submitted).
- <sup>66</sup>P. J. Stephens, *J. Phys. Chem.* **89**, 748 (1985).
- <sup>67</sup>P. J. Stephens, *J. Phys. Chem.* **91**, 1712 (1987).
- <sup>68</sup>D. A. Long, *The Raman Effect: A Unified Treatment of the Theory of Raman Scattering by Molecules* (Wiley, Chichester, 2001).