

Vibrational Analysis of Constrained Molecular Systems

Published as part of The Journal of Physical Chemistry A *virtual special issue* “Krishnan Raghavachari Festschrift”.

Ankur K. Gupta* and Krishnan Raghavachari



Cite This: *J. Phys. Chem. A* 2024, 128, 28–40



Read Online

ACCESS |



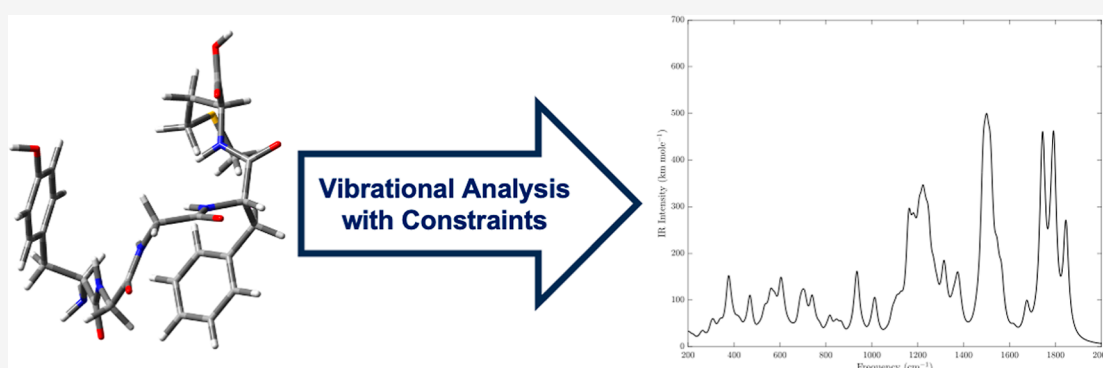
Metrics & More



Article Recommendations



Supporting Information



ABSTRACT: Vibrational spectroscopy, including infrared (IR), Raman spectroscopy, and vibrational circular dichroism, is instrumental in determining the structure and composition of molecules. These techniques are highly sensitive to molecular conformations. However, full molecular optimization, necessary for theoretical vibrational spectra, can lead to unintended conformational changes, especially in large biomolecules like polypeptides. To address this, dihedral angle constraints can be imposed during optimization to preserve the molecule’s native conformation. Constraint-optimized molecular geometries, not being true stationary points in the full configurational space, pose challenges for traditional vibrational analysis. We address this by considering such geometries as subspace minima, reformulating vibrational analysis to incorporate constraints. Normal modes and spectra consistent with these constraints are obtained by projecting the force constant matrix onto a space orthogonal to the constrained coordinates. This method, illustrated by the example of enkephalin, yields $3N - 6 - m$ nonzero frequencies after constraint projection, demonstrating its applicability to biomolecules with flexible conformations. Our approach offers a comprehensive mathematical framework to compute vibrational spectra of molecules with conformationally flexible subunits under environmental constraints.

1. INTRODUCTION

Vibrational spectroscopy is a crucial technique for uncovering molecular structures and compositions but is often confronted with the complexity of experimental spectra. While the experimental spectra hold unique information about the chemical bonds in a molecule, the inherent spectral bandwidths can cause the overlap of spectral features, complicating their analysis. Thus, theoretical approaches are commonly used to interpret the origins and details of vibrational spectral signatures, with cross-validation against experimental data being a standard practice.

Molecular vibrations are typically described using a harmonic oscillator model, which assumes that atoms oscillate about their equilibrium positions, necessitating the molecule to be at its minimum energy configuration.¹ Therefore, molecular geometry optimization is required before performing frequency calculations to obtain spectra. However, the actual molecular

geometry may differ from this theoretical minimum, especially when the complete chemical system is not accounted for in the computations. This is observed in cases like catalysts trapped in zeolite sieves, molecules within nanocapsules, and biomolecules in dense cellular environments.

Peptide molecules exhibit vibrational interactions that are highly sensitive to their conformation, which in turn can be influenced by the surrounding environment. These influences may cause peptides to assume conformations that do not correspond to their most energetically favorable states.

Received: June 29, 2023

Revised: November 20, 2023

Accepted: November 21, 2023

Published: December 28, 2023



Theoretically capturing every facet of these complex biochemical environments poses a significant computational challenge, leading to quantum mechanical calculations often being performed on truncated chemical structures in the gas phase or with simple solvation models. Neglecting the molecule's actual chemical environment can result in substantial structural changes during optimization. As a result, achieving full geometry optimization while maintaining the experimentally observed structure often proves to be an incongruent objective for conformationally flexible molecules.

In response to these challenges, the field has seen the development of various innovative approaches. A notable example is the Cartesian Coordinate Transfer (CCT) method, introduced by Bour and colleagues, which has been utilized for molecules containing as many as 10,000 atoms.^{2–5} Given the impracticality of brute force optimization and frequency calculations for large systems with current computational resources, the CCT method leverages the transferability of tensor derivatives from molecular fragments to their parent system, thereby circumventing the need for intensive calculations on the entire parent molecule. However, CCT is not without its limitations, as it employs a nonequilibrium molecular geometry, leading to nonzero forces. Furthermore, the full molecular Hessian in CCT contains inherent approximations, primarily due to its construction from aggregated Hessians of individual molecular fragments or subsystems. As a result, this approach might yield several imaginary frequencies when diagonalizing the CCT Hessian. Despite these limitations, CCT has been proven effective in accurately predicting experimental spectra for medium to very large molecular systems.

In a related research, Polavarapu and colleagues investigated a conformationally flexible tetrapeptide, WUGW, which undergoes a notable conformational shift during optimization.⁶ To preserve its observed conformation, the peptide's dihedral angles were constrained. Frequencies were then obtained by diagonalizing the Hessian at the partially optimized structure. It is important to note that since this constrained geometry does not constitute a true minimum, the resulting forces would be nonzero, which might lead to some imaginary frequencies. Nonetheless, the comparison between the theoretical and experimental spectra was reported to be satisfactory.

In our current work, we aim to establish a protocol that preserves the molecule's native conformation through constraints while yielding a mathematically consistent Hessian and, consequently, accurate vibrational spectra. For studying large molecular systems, environmental restrictions can be represented as internal coordinate constraints within the molecule of interest. Therefore, we impose geometric constraints during optimization to maintain the molecule's native conformation. To incorporate these constraints into the Hessian, we have adapted a technique by Lu et al.,^{7,8} which involves projecting out internal coordinate constraints from the Hessian matrix. This adaptation enables us to develop new protocols for obtaining vibrational spectra of constrained molecules.

2. METHODS: VIBRATIONS IN A CONSTRAINED MOLECULE

Understanding molecular vibrations is best achieved through the lens of classical formalism. To determine the fundamental frequencies of molecular vibrational motions, it is essential to establish an appropriate frame of reference. The Casimir–

Eckart conditions, also known as Sayvetz conditions, provide a framework for defining an instantaneous coordinate system that neutralizes any rotational motion resulting from vibrations.^{9,10} Within this framework, the molecular kinetic and potential energies can be expressed, which are then utilized to formulate a set of equations of motion in the Lagrangian form.

When formulating the potential energy function for deriving the Lagrangian, it is crucial that the first-order term (the gradient or force) in the truncated Taylor expansion of the potential is zero. This condition holds only when the molecule is in its minimum energy (equilibrium) configuration. Consequently, solving the secular equation derived from the Lagrangian simplifies to the diagonalization of the Hessian matrix. This process yields the fundamental frequencies and their corresponding normal modes.

In practice, the six eigenvalues corresponding to translations and rotations obtained from the diagonalized Hessian, as calculated by quantum chemistry software, are not exactly zero.¹¹ Frequencies for translational modes are typically around 0.01 cm^{-1} but can reach up to 50 cm^{-1} for rotational coordinates. This discrepancy occurs because molecular geometry is optimized to a finite gradient value rather than precisely zero. Additionally, grid-related numerical issues in density functional theory methods, particularly for heavier elements, also contribute to this variance.

For most large molecules of practical interest, frequencies of the pure low-frequency vibrational modes may overlap with those of the “rotational” modes, potentially resulting in unreliable values for the $3N - 6$ fundamental frequencies. To circumvent potential ambiguities, translational and rotational coordinates are systematically projected out from the Cartesian force constant matrix prior to its diagonalization.

The secular equation, which facilitates the derivation of fundamental frequencies, is valid if the molecular geometry corresponds to a stationary point on the $3N$ -dimensional potential energy surface. Our study focuses on molecules that undergo substantial conformational changes during their optimization, rendering the entire optimization process unphysical. As such, these systems pose significant challenges for theoretical investigation, especially when using conformationally sensitive techniques such as vibrational spectroscopy.

In practical scenarios, constraints must be applied to the molecule to maintain its native conformation. The molecular configuration achieved through this constrained optimization does not represent a stationary point in the full $3N$ -dimensional configuration space (R^{3N}). Nonetheless, it can be considered as a stationary point within the subspace defined by the unconstrained degrees of freedom. This approach allows for a generalized normal-mode analysis of the molecule, preserving its original conformation while allowing a partially relaxed geometry in a reduced vector space.

Constraints are typically represented as curvilinear internal coordinates, such as interatomic distances, bond angles, and dihedral angles, due to their ease of visualization, interpretation, and implementation. The force constant matrix, consistent with these constraints, can be calculated once the projection operator for the constrained coordinates is identified. This process is detailed in eq A7 in the Appendix, where we provide a full derivation of the constrained Hessian formalism. In the main text, we focus on outlining the key concepts.

To formulate the necessary projection operator, as outlined in eq A6, the constraints must match the basis of the Hessian matrix, typically defined in mass-weighted Cartesian displacement coordinates. Consequently, for small-amplitude vibrations, a transformation from Cartesian to internal displacement coordinates is required. This transformation is given in standard treatments as follows:¹

$$\mathbf{s} = \mathbf{B}\mathbf{x} \quad (1)$$

where vectors $\mathbf{s}$ and $\mathbf{x}$ represent the sets of internal and Cartesian displacement coordinates, respectively.

The so-called “Wilson $\mathbf{B}$ matrix”, determined using molecular geometry, is generally a rectangular matrix. These Wilson vectors are utilized to construct the projection operators in accordance with eq A6, corresponding to the internal coordinate constraints imposed on the molecule. For a molecule with m independent internal coordinate constraints, the corresponding projection operator is formulated as follows:

$$\mathbf{P}^C = \sum_{i=1}^m \mathbf{b}_i \mathbf{b}_i^T \quad (2)$$

where $\mathbf{b}_i$ represents the Wilson vector for the i th internal coordinate. When constraints are not independent, an orthogonal set of vectors can be obtained through QR factorization.

Additionally, it is necessary to ensure that these constrained vectors are orthogonal to translational and rotational normal modes before constructing the projection operator using eq 2. Following this, the force constant matrix and the gradient vector can be projected onto a space that is independent of both the imposed constraints and the rotational and translational movements, as detailed in eqs A7 and A8, using the following equations:

$$\mathbf{f}^{\text{RTC}'} = (\mathbf{I} - \mathbf{P}^C)(\mathbf{I} - \mathbf{P}^{\text{RT}})\mathbf{f}(\mathbf{I} - \mathbf{P}^{\text{RT}})(\mathbf{I} - \mathbf{P}^C) \quad (3)$$

$$\mathbf{g}^{\text{RTC}'} = (\mathbf{I} - \mathbf{P}^C)(\mathbf{I} - \mathbf{P}^{\text{RT}})\mathbf{g} \quad (4)$$

When the projected Hessian matrix ($\mathbf{f}^{\text{RTC}'}$) is diagonalized, it results in $3N - 6 - m$ nonzero eigenvalues, indicating a reduction in the vibrational space dimensionality for the molecule. If a molecule is optimized with m constraints, the projected gradient (as per eq 4) should theoretically be close to zero, within numerical precision. Additionally, all eigenvalues of the projected Hessian (as per eq 3) should be positive, signifying a minimum in the reduced space. The vibrational modes derived from the projected force constant matrix ($\mathbf{f}^{\text{RTC}'}$) ensure that the constrained coordinates remain stationary. Consequently, within the projected Hessian framework, coordinates deviating from their equilibrium positions do not engage in vibrational motion, a conclusion consistent with physical principles.

The vibrational spectrum of a molecular system is graphically depicted as a function of its fundamental frequencies and associated intensities. These intensities are calculated from molecular property tensors, including dipole moment derivatives, atomic axial tensors, and polarizability derivatives. Standard quantum chemistry packages typically provide these tensors based on the Cartesian coordinate system. However, for constructing a spectral plot (fundamental frequencies versus intensities), these tensors must be transformed to the basis of normal coordinates. Therefore, all property tensors are projected onto the normal modes

obtained from the projected force constant matrix ($\mathbf{f}^{\text{RTC}'}$). This process enables the calculation of the derivative of a physical property, A , with respect to a normal coordinate, Q_i , as demonstrated by the following formula:¹²

$$\frac{dA}{dQ_i} = \sum_{k=1}^{3N} \frac{\partial A}{\partial x_k} l_{ki} \quad (5)$$

In this equation, x_k represents the Cartesian displacement coordinates and l_{ki} are the eigenvector components corresponding to the normal mode Q_i .

3. COMPUTATIONAL DETAILS

For vibrational analysis of conformationally flexible molecules, such as polypeptides, the typical approach involves full geometry optimization either in vacuo or within an implicit solvent environment. This optimization can be performed using any standard quantum chemistry package; in our study, we utilized Gaussian 16¹³ for all computations. After optimization, the molecular geometry in its optimized state is compared with its initial configuration to assess the extent of conformational changes. For larger polypeptides, significant structural changes may not be immediately visually evident. Thus, we employed the root-mean-square deviation (RMSD) of atomic positions to quantify structural alterations. In a molecule with N atoms, if vector $\mathbf{X}$ represents the coordinates of the initial conformation and $\mathbf{X}'$ those of the optimized conformation, the RMSD between them is calculated as follows:

$$\text{RMSD} = \sqrt{\frac{1}{N} \sum_{i=1}^N \|\mathbf{X}_i - \mathbf{X}'_i\|^2} \quad (6)$$

This measurement assumes proper superposition of the two structures. This measurement assumes that the two structures are properly superimposed. This measurement assumes proper superposition of the two structures. RMSD is usually expressed in angstroms (Å). Typically, an RMSD value less than 1.5 Å indicates a close resemblance between the optimized geometry and its original structure, while an RMSD greater than 2.5 Å suggests a significant deviation of the optimized molecule's conformation from its native state.¹⁴ The RMSD value can vary with the level of theory used; ideally, it should approach zero as the theoretical model increasingly reflects real system behaviors. Consequently, a balanced level of theory [B3LYP/6-31G(d)] was selected for our computations, considering both accuracy and computational efficiency.

To define a set of curvilinear internal coordinates for constraining a molecule during its optimization process, we took into consideration that the energy changes resulting from alterations in bond lengths and angles are typically greater than those associated with changes in dihedral angles. Consequently, dihedral angles are more prone to deviate from their initial values during geometry optimization, potentially altering the polypeptide's conformation. Therefore, our protocol involved freezing appropriate dihedral angles during energy minimization to preserve the molecule's initial geometry. In the examples presented in this paper, this protocol entailed constraining all dihedral angles associated with single bonds involving only heavy (non-hydrogen) atoms. Torsions associated with partial double bonds were not constrained, owing to their higher energy barriers to rotation. Likewise, dihedral angles in five-membered rings or smaller were

excluded due to inherent constraints within their cyclic structure. Following constrained optimization, both the force constant matrix and gradient were projected (using eqs 3 and 4) onto a subspace orthogonal to that spanned by dihedral angle constraints. The resulting Hessian was then diagonalized to derive the fundamental frequencies and corresponding eigenvectors.

4. RESULTS AND DISCUSSION

4.1. Illustrative Example: Enkephalin (1PLW). The process of obtaining vibrational spectra for geometrically constrained molecular systems is best illustrated with a specific peptide example. For our study, we chose enkephalin (PDB ID: 1PLW), a neuropeptide prevalent in the central nervous system, known for interacting with nerve cell membranes and adopting a conformation suitable for binding to opiate receptors.¹⁵ Its structure, depicted in Figure 1, was obtained

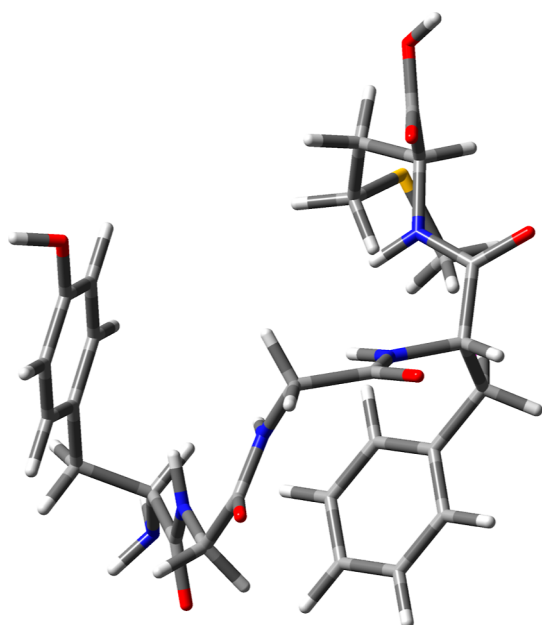


Figure 1. Representation of enkephalin's molecular geometry (PDB ID: 1PLW) as sourced from the Protein Data Bank (PDB). The structure illustrates the spatial arrangement of atoms in enkephalin as determined experimentally.

from the PDB file, determined using multidimensional proton NMR in solution. Among the 80 available conformers of 1PLW in the PDB database, we selected conformer 1 for our quantum chemistry simulations. Preliminary structural adjustments included protonating the carboxylic terminus and deprotonating the amine group, neutralizing the structure for computational analysis. The RMSD between the initial PDB structure of 1PLW and its fully optimized geometry, considering all atoms, was 2.64 Å. Consequently, we reoptimized it with 16 dihedral angle constraints. This partial optimization resulted in a reduced RMSD of 0.86 Å, indicating preservation of the peptide's conformation. The IR spectrum derived from the unprojected force constant matrix is presented in Figure 3a, where the most intense peaks near 1800, 1500, and 1200 cm^{-1} correspond to the amide I, II, and III vibrational modes. Amide I is primarily related to the C=O bond stretch, while amide II and III modes are combinations of C–N stretching and H–N–C bending. These modes are

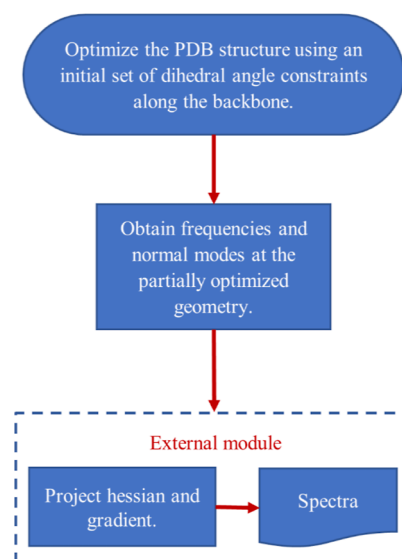


Figure 2. Flowchart illustrating the protocol for obtaining vibrational spectra consistent with imposed constraints. The process begins with the constrained optimization of a molecule, employing an initial set of dihedral angle constraints for every rotatable bond. Next, the Hessian matrix is computed at the partially optimized geometry, which is then diagonalized to compute the frequencies and normal modes, resulting in the raw (unprojected) vibrational spectrum. This Hessian is subsequently processed with an external script to align the spectrum with the imposed constraints. This step involves projecting the Hessian onto a subspace orthogonal to these constraints, as per eq 3, followed by diagonalization to generate a spectrum that accurately reflects the constraints.

crucial for identifying secondary structural changes in proteins due to their sensitivity to noncovalent interactions and peptide geometry.

The Hessian matrix and the gradient vector, derived from the partially relaxed geometry, were projected using eqs 3 and 4, respectively. A flowchart visually representing our projection protocol is illustrated in Figure 2. After projection, the RMS gradient decreased from 1.7×10^{-3} to 5.2×10^{-6} a.u., and all eigenvalues of the projected Hessian were positive, indicating that the constrained optimized structure represents a minimum in the subspace orthogonal to the constrained vectors. The resulting projected IR spectrum is displayed in Figure 3b. When comparing the spectra in Figure 3a,b, the spectrum obtained after projection shows a broader distribution of absorption peaks. This effect arises from excluding the constrained degrees of freedom from the vibrational space. In the unprojected case, all internal (or Cartesian) coordinates could vibrate harmonically, despite some not being at their equilibrium positions. However, by projecting constraints from the Hessian, these constrained coordinates are omitted from vibrations, leading to modified normal modes and potential shifts in frequencies and intensities. The extent of these shifts generally relates to the involvement of a constrained internal coordinate in the original normal mode. The projection may also affect the three amide modes, given their partial dependence on dihedral angle movements.

4.2. Minimizing Coupling Interference between Constrained Coordinates and Amide Modes. While our earlier-described procedure yields frequencies that are mathematically consistent with imposed constraints, it is important to note that amide modes may be affected due to

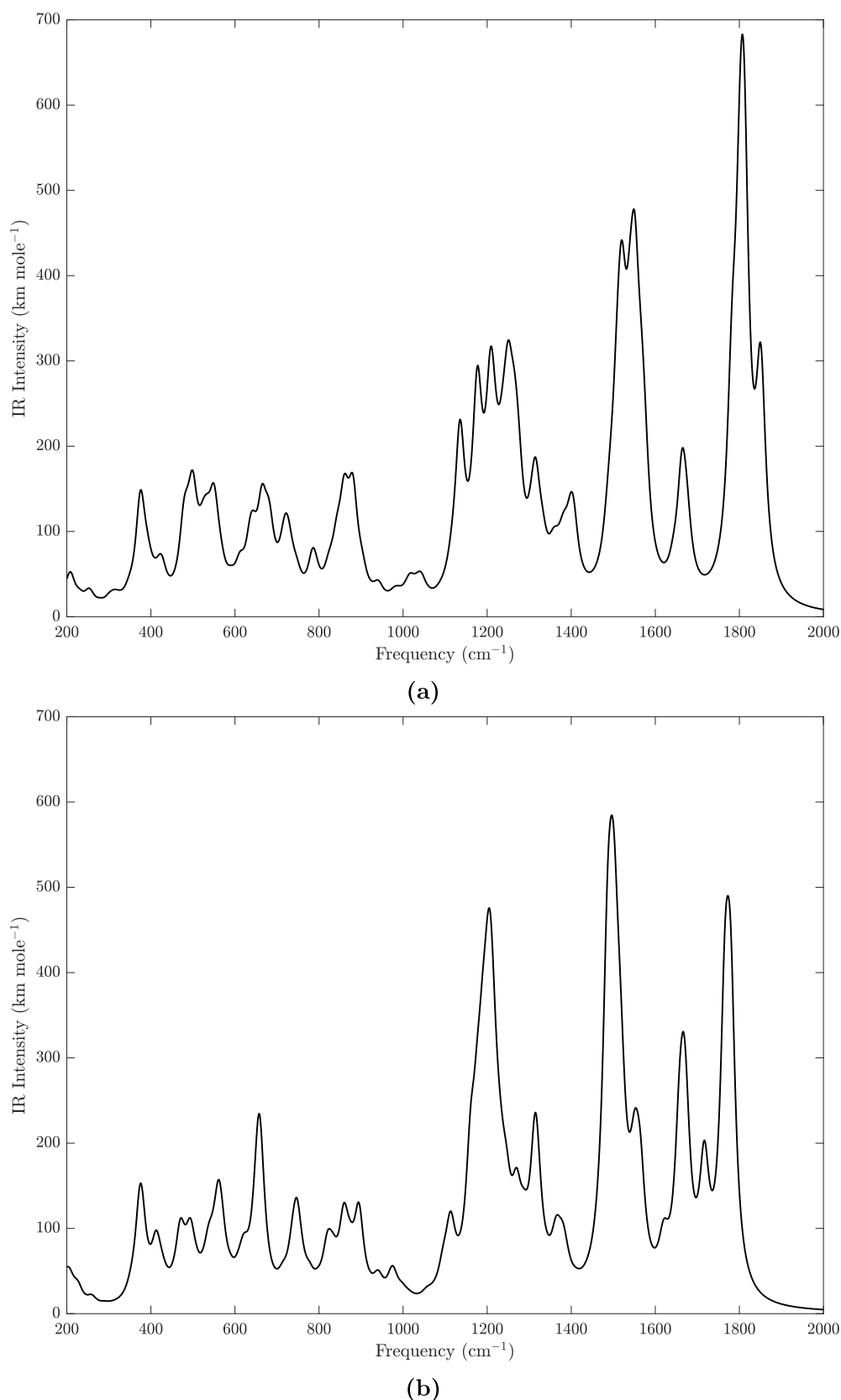


Figure 3. Infrared (IR) spectra of enkephalin (1PLW) with a full-width at half maximum (FWHM) of 14 cm⁻¹ (a) displays the IR spectrum derived from the unprojected Hessian matrix (raw spectra), while (b) shows the IR spectrum from the Hessian matrix projected using eq 3, following the protocol illustrated in Figure 2. Both spectra are obtained under the optimization involving 16 dihedral angle constraints.

their partial dependence on the constrained dihedral angles. Therefore, it can be advantageous to choose dihedral angle

constraints that minimally interfere with high-intensity normal modes (notably, 1100–1900 cm⁻¹ for amide vibrations), while

still keeping the RMSD from the original PDB structure low. As outlined in Figure 5, we have developed a systematic

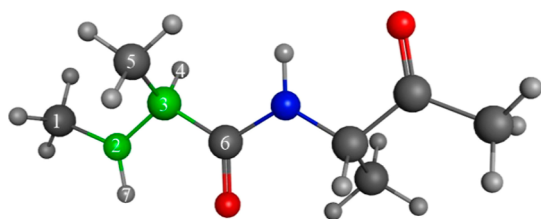


Figure 4. Diagrammatic representation of a dipeptide molecule illustrating the determination of the Degree of Coupling (DOC) parameter. This figure exemplifies the process of constraining rotation along the bond between atoms 2 and 3. Among the six potential dihedral angles available for this bond (1–2–3–4, 1–2–3–5, 1–2–3–6, 7–2–3–4, 7–2–3–5, and 7–2–3–6), one is selected based on the DOC analysis. The chosen dihedral angle is the one with the smallest DOC value, ensuring minimal interference with the molecule's normal modes.

protocol to select dihedral angle constraints for every rotatable bond between heavy (non-hydrogen) atoms in a molecule. For enkephalin, this process led to the identification of 16 constraints, as discussed in Section 4.1. After imposing these constraints, the molecule undergoes optimization and the Hessian matrix is computed for the partially optimized structure. This computation enables projection of the resultant normal modes onto the internal coordinate space, allowing for the quantification of each internal coordinate's contribution to specific normal modes, known as the relative weight (RW). Recognizing the importance of high-intensity modes for practical applications, we aimed to develop a metric that accentuates the contribution of internal coordinates to these significant modes. This was achieved by scaling the internal coordinates' contributions to normal modes (their RWs) with the corresponding infrared intensity values. Consequently, for each dihedral angle constraint, or any other internal coordinate, denoted as d_j , we established a measure termed the degree of coupling (DOC).

$$\text{DOC}(d_j) = \sum_i (RW)_i \cdot I_i \quad (7)$$

Here, $(RW)_i$ represents the relative weight of the dihedral angle d_j in the i th normal mode (between 1100 and 1900 cm^{-1}), correlated with the infrared intensity I_i . Equation 7 clarifies that a dihedral angle constraint with a higher DOC value suggests stronger coupling with high-intensity modes, while a lower DOC value indicates weaker coupling. Therefore, for constrained optimization, it is preferable to select internal coordinates with smaller DOC values as constraints. To apply this, the DOC value is computed for every potential dihedral angle associated with rotation along a single bond. The dihedral angle demonstrating the lowest DOC is then chosen to constrain the rotation around that specific bond.

Take, for instance, a dipeptide shown in Figure 4. To constrain rotation along the bond between atoms 2 and 3, one of the six potential dihedral angles (1–2–3–4, 1–2–3–5, 1–2–3–6, 7–2–3–4, 7–2–3–5, or 7–2–3–6) can be frozen. However, each angle's contribution to normal modes varies. Therefore, we select the dihedral angle with the smallest DOC to minimize its coupling with normal modes, resulting in a spectrum with minimal lateral shifts. It's crucial to understand

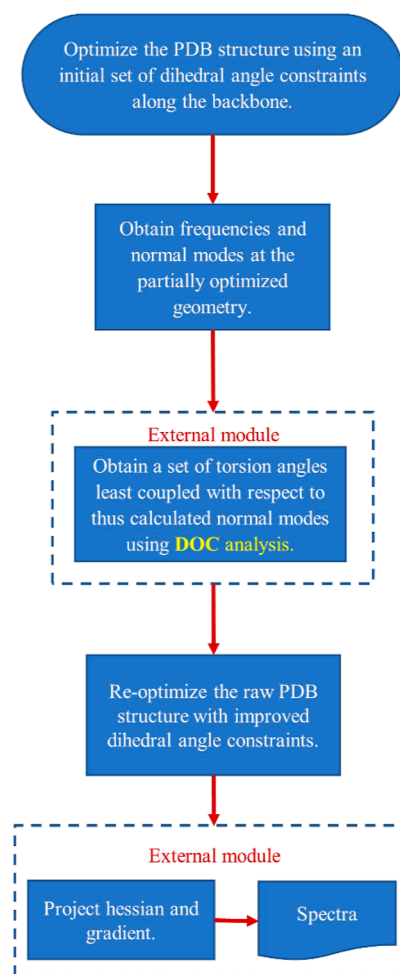


Figure 5. Flowchart illustrating the protocol for obtaining vibrational spectra consistent with constraints, refined by the Degree of Coupling (DOC) metric (eq 7). The process initiates with the constrained optimization of a molecule, applying dihedral angle constraints to each rotatable bond. Subsequent to the computation and diagonalization of the Hessian at the partially optimized geometry, the DOC value is calculated for every dihedral angle associated with single rotatable bonds. This step identifies a new set of dihedral angles with the lowest DOC values, thereby minimizing interference with high-intensity normal modes. The molecule is then reoptimized using these optimized constraints. The resulting Hessian from this new partially optimized geometry is further processed to align the vibrational spectrum with these refined constraints. This involves projecting the Hessian onto a subspace orthogonal to the constraints as per eq 3, followed by its diagonalization to generate a spectrum that accurately reflects these constraints.

that the DOC analysis is highly influenced by the nature of normal modes, which depends on the molecule's conformation and geometry. Thus, applying this protocol to a raw PDB structure, where atomic positions are unrelaxed, is inappropriate as it leads to inaccurate representations of the normal modes compared to those of a relaxed geometry from constrained optimization. Given this, our first step involves performing a constrained optimization and full Hessian calculation on the target molecule. This is done using a set of initial guess constraints, such as all rotatable single bond dihedral angle constraints. Then, a DOC analysis on the partially optimized geometry identifies a refined set of constraints that minimally interfere with normal modes. The

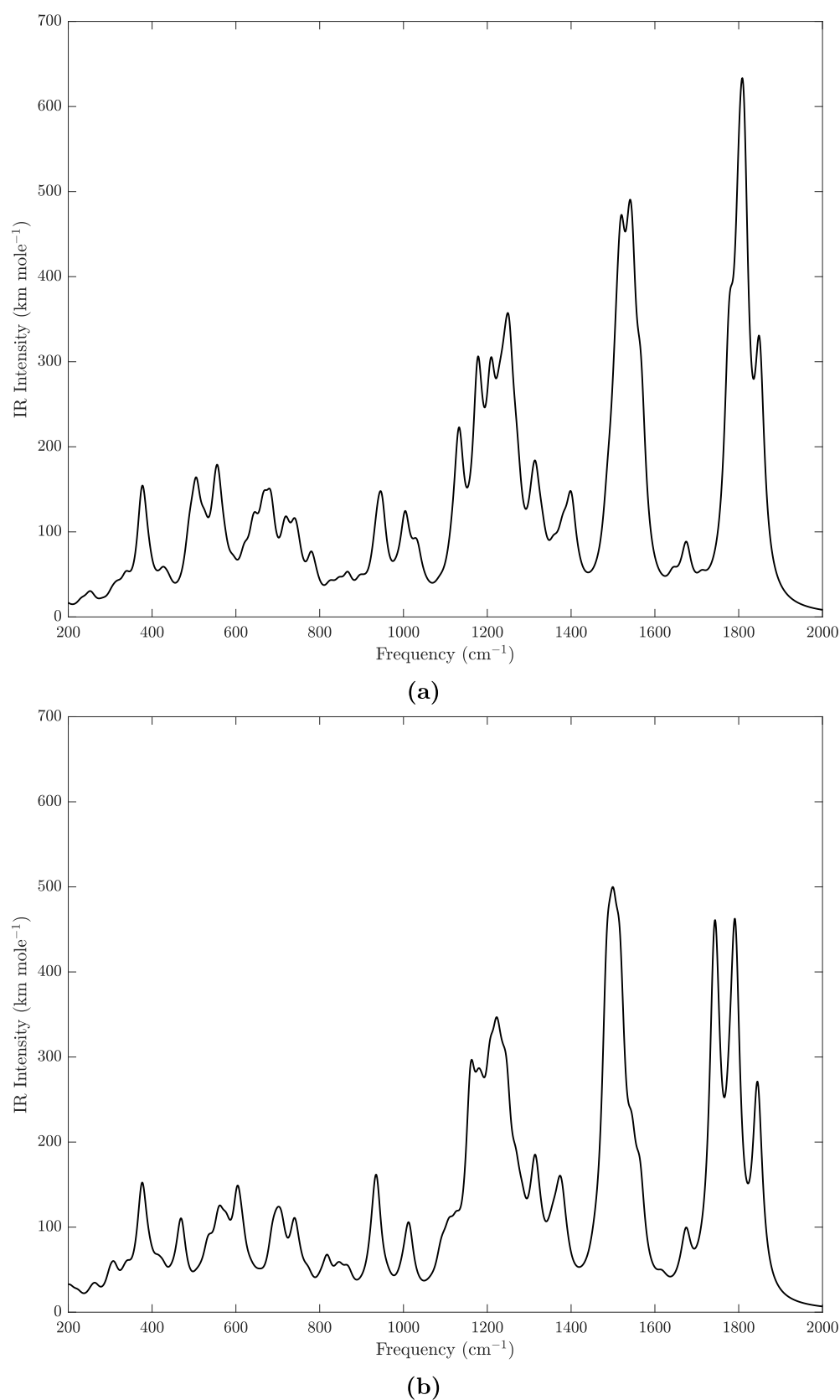
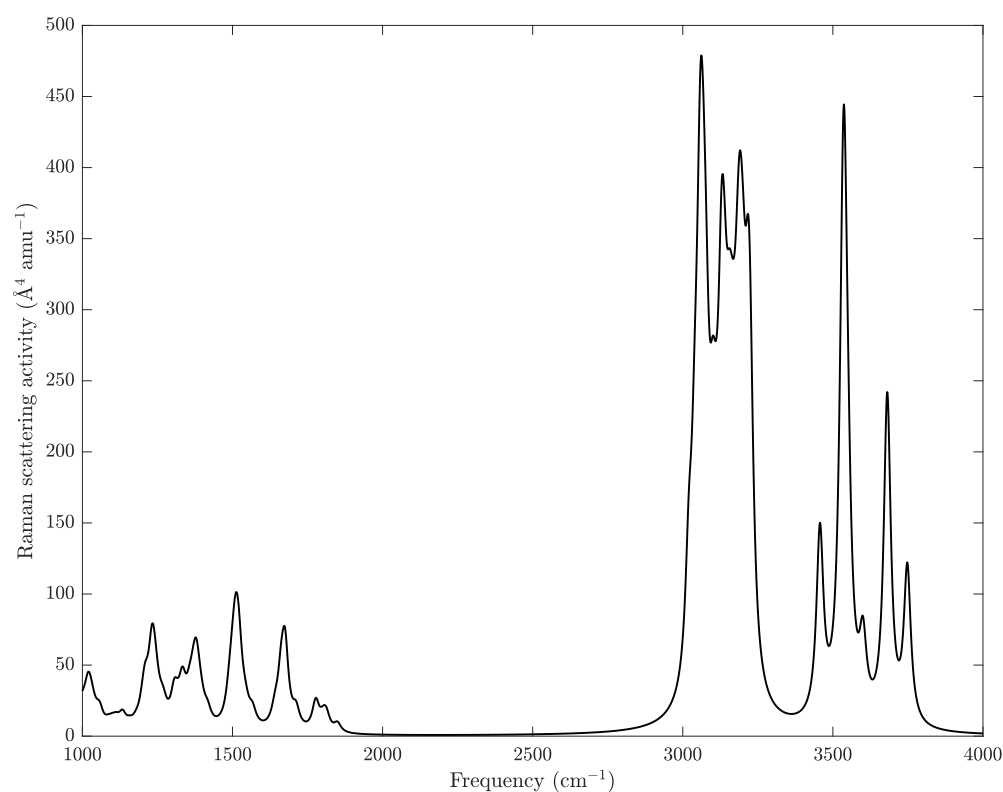
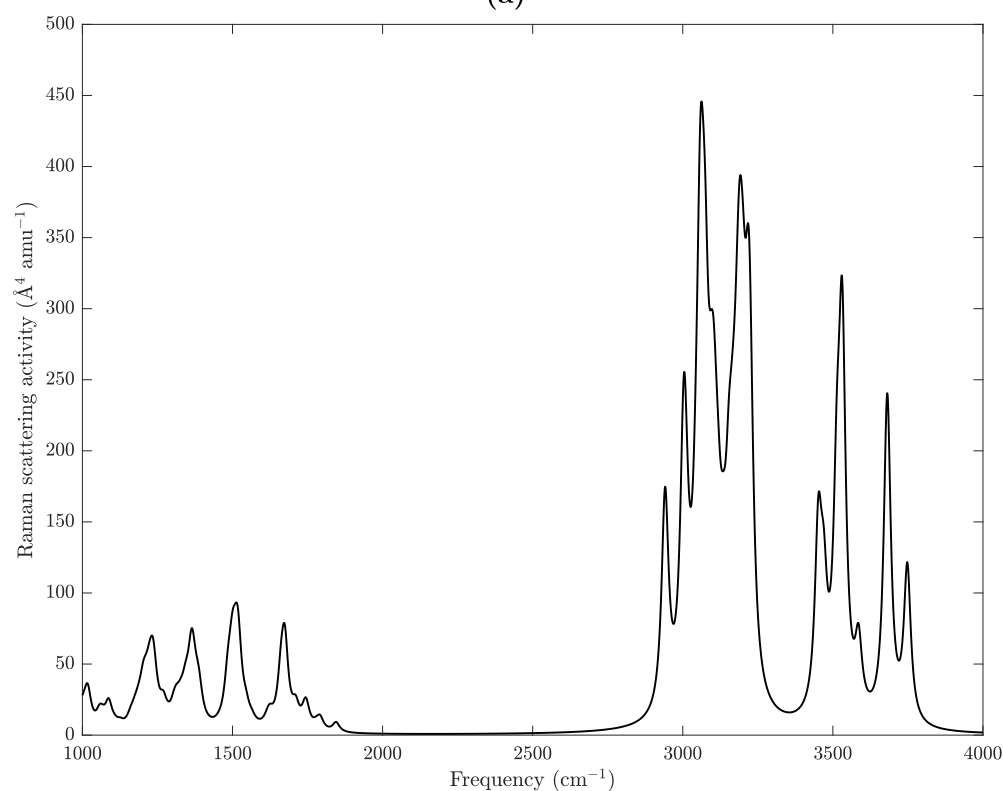


Figure 6. Infrared (IR) spectra of enkephalin (1PLW) with a full-width at half maximum (FWHM) of 14 cm⁻¹, obtained after applying 16 dihedral angle constraints as determined by the DOC analysis. Subfigure (a) displays the IR spectrum derived from the unprojected Hessian matrix (raw spectrum), illustrating the initial spectral characteristics. Subfigure (b) presents the IR spectrum from the projected Hessian matrix using eq 3, following the DOC-enhanced protocol depicted in Figure 5, to demonstrate alterations in spectral features due to the projection.



(a)



(b)

Figure 7. Raman scattering spectra of enkephalin (1PLW) with a full-width at half maximum (FWHM) of 14 cm^{-1} , obtained after implementing 16 dihedral angle constraints as identified by the DOC analysis. Subfigure (a) shows the Raman spectrum derived from the unprojected Hessian matrix (raw spectrum), capturing the original spectral characteristics. Subfigure (b) features the Raman spectrum obtained from the Hessian matrix projected using eq 3, in accordance with the DOC-enhanced protocol illustrated in Figure 5. Notably, spectral changes are minimal due to the high intensity peaks being in the high-frequency region, where the normal modes exhibit minimal coupling with dihedral angles.

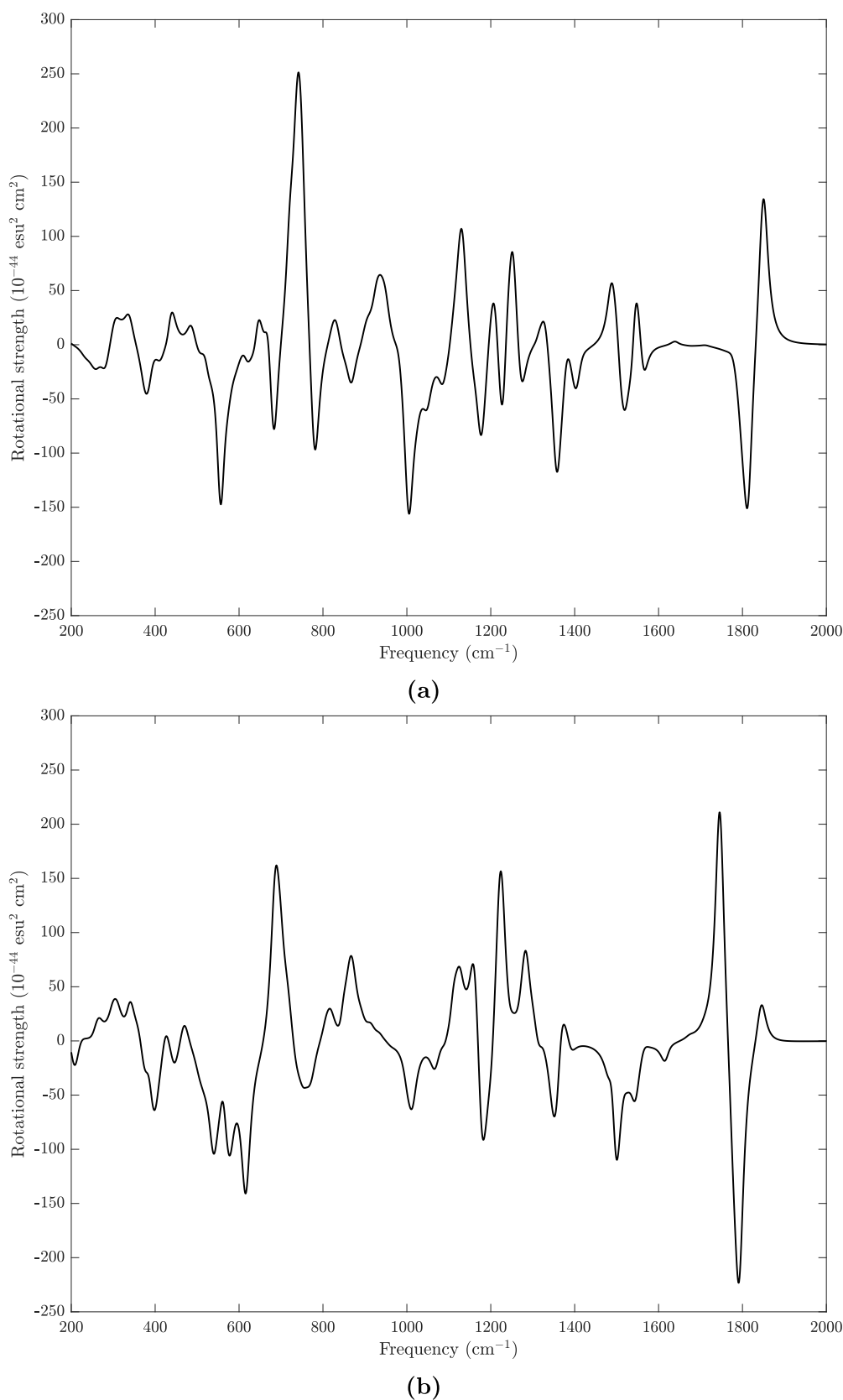


Figure 8. Vibrational circular dichroism (VCD) spectra of enkephalin (1PLW) with a full-width at half maximum (FWHM) of 14 cm⁻¹, obtained after implementing 16 dihedral angle constraints as identified by the DOC analysis. Subfigure (a) displays the VCD spectrum derived from the unprojected Hessian matrix (raw spectrum), showcasing the original spectral characteristics. Subfigure (b) features the VCD spectrum from the projected Hessian matrix using eq 3, following the DOC-enhanced protocol depicted in Figure 5. This projection may result in shifts or alterations in spectral features, reflecting the influence of constrained dihedral angles on the VCD spectrum.

molecule is then reoptimized with these new constraints, and the resultant Hessian is projected using these constraints to generate the desired spectrum. This process, illustrated as a flowchart in Figure 5, may undergo further iterations to reduce coupling between normal modes and constraints. However, we found that additional iterations typically result in only minor changes to structures and spectra. Therefore, our protocol provides a balanced approach to efficiently determine the most appropriate constraints while considering computational demands.

Consider the DOC analysis applied to the partially optimized structure of 1PLW from Section 4.1, which led to a new set of constraints. After reoptimizing the molecule with these enhanced constraints, we obtained corresponding raw and projected IR spectra, as displayed in Figure 6a,b. Notably, the absorption peaks in the new projected spectrum (Figure 6b) are less dispersed compared to those in the earlier projected spectrum (Figure 3b). This reduced spread is attributed to the new set of internal coordinates being less coupled with the high-intensity normal modes. Additionally, a Raman spectrum was generated for this geometry. As shown in the Raman scattering spectra (Figure 7), all high-intensity modes are in the high-frequency region, where the coupling of dihedral angles with normal modes is minimal, leading to negligible spectral changes after projection. The vibrational circular dichroism (VCD) spectra (Figure 8) further illustrate an interesting aspect where peak signs can vary, adding an extra layer of complexity. Consequently, the projection process can cause shifts and sign changes in spectral peaks. In the case of 1PLW, the raw spectrum shows a negative couplet for amide I, which transforms into a (+,−,+) band pattern after projection.

5. CONCLUSIONS

This work aims to rigorously calculate the vibrational structures of peptides while preserving their native conformations. Given that peptide molecules tend to alter their conformations during optimization when removed from their real biochemical environment, specific geometric constraints, such as torsion angles, must be imposed on the polypeptide. Our protocol projects out unphysical vibrations along coordinates that are constrained and not fully optimized. The diagonalization of the projected force constant matrix then results in a new set of harmonic frequencies and corresponding normal modes, which do not include contributions from the constrained internal coordinates.

In summary, our protocol applies dihedral angle constraints during the optimization of molecules. The force constant matrix of the partially relaxed geometry is then projected using appropriate projection operators, resulting in a projected force constant matrix. This process enables us to generate a mathematically rigorous vibrational spectrum. While our projection method preserves all peaks associated with amide modes, it also leads to reduced overall intensity and broader peak distribution. This effect arises from the coupling between the internal coordinate constraints and normal coordinates. To mitigate this coupling interference, we have introduced a protocol that yields a less perturbed vibrational spectrum while still being consistent with the structural constraints.

While our results are mathematically rigorous, validating this method against experimental data remains essential. Peptides studied experimentally tend to be large and complex, often existing in solution, posing significant challenges for computational modeling. For such systems, fragment-based methods

may be necessary to achieve more realistic simulations. Crucially, our protocol is designed for seamless integration with a variety of fragment-based methods. Its primary function is to process the Hessian matrix with defined constraints, rendering it adaptable and agnostic to specific computational methods or theoretical levels.

It is crucial to acknowledge that while the projection operator model offers a rigorous method for simulating vibrational spectra of systems optimized with constraints, careful application of our protocol is advised. For instance, it may be prudent to explore the fewest torsional constraints needed to preserve the experimental conformation during optimization. Additionally, for complex systems in constrained environments, the actual physical limitations experienced by the molecule may not align perfectly with the torsional constraints imposed in our computational model. Imposed dihedral constraints that do not accurately reflect the molecular environment or are based on unrealistic assumptions could limit the physical accuracy of the model's predictions. Despite these considerations, our approach establishes a robust mathematical framework for analyzing the vibrational spectra of large molecules under environmental constraints.

APPENDIX

Note: The equations in this section are adapted from Wilson, E.; Decius, J.; Cross, P. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*; Dover Books on Chemistry; Dover Publications, 1980.

Projection Operators for Vibrational Analysis

Any vector $\mathbf{v}$ in the $3N$ -dimensional vector space R^{3N} can be written as

$$\mathbf{v} = \mathbf{v}^S + \mathbf{v}^{S'} \quad (\text{A1})$$

where $\mathbf{v}^S$ is the projection of $\mathbf{v}$ on the subspace S of R^{3N} and $\mathbf{v}^{S'}$ represents the projection of $\mathbf{v}$ on the complementary space (of S), S' . The corresponding linear projection operators $\mathbf{P}^S$ and $\mathbf{P}^{S'}$ on the subspaces S and S' , respectively, are defined such that

$$\begin{aligned} \mathbf{v}^S &= \mathbf{P}^S \mathbf{v} \\ \mathbf{v}^{S'} &= \mathbf{P}^{S'} \mathbf{v} \end{aligned} \quad (\text{A2})$$

Some properties associated with these projection operators are as follows

$$(\mathbf{P}^S)^2 = \mathbf{P}^S \quad (\text{A3})$$

$$\mathbf{P}^S + \mathbf{P}^{S'} = \mathbf{I} \quad (\text{A4})$$

$$\mathbf{P}^S \mathbf{P}^{S'} = \mathbf{0} \quad (\text{A5})$$

If we define a subspace S_1 spanned by k orthonormal vectors (or constraints) $\mathbf{e}_i$, $i = 1, \dots, k$, then the corresponding projection operator can be written as

$$\mathbf{P}^{S_1} = \sum_{i=1}^k \mathbf{e}_i \mathbf{e}_i^T \quad (\text{A6})$$

The force constant matrix ($\mathbf{f}_1^S$) and the gradient vector ($\mathbf{g}_1^S$) in the subspace S_1 (and its complementary space S'_1) can then be obtained as follows

$$\begin{aligned} \mathbf{f}_1^S &= \mathbf{P}^{S_1} \mathbf{f} \mathbf{P}^{S_1} \\ \mathbf{f}_1^{S'_1} &= (\mathbf{I} - \mathbf{P}^{S_1}) \mathbf{f} (\mathbf{I} - \mathbf{P}^{S_1}) \end{aligned} \quad (\text{A7})$$

$$\mathbf{g}^{S_1} = \mathbf{P}^{S_1} \mathbf{g}$$

$$\mathbf{g}^{S'_1} = (\mathbf{I} - \mathbf{P}^{S_1}) \mathbf{g} \quad (\text{A8})$$

where $(\mathbf{I} - \mathbf{P}^{S_1}) = \mathbf{P}^{S'_1}$ represents the projection operator for the complementary space S'_1 following eq A4.

To project out translational and rotational coordinates from the force constant matrix ($\mathbf{f}$), it is necessary to construct corresponding projection operators. The normal coordinates for translational and rotational modes are derived from the six Eckart conditions. These conditions facilitate the formation of vectors (in the basis of mass-weighted Cartesian coordinates) that encompass the subspace of translational and rotational motion. Therefore, the vectors $\mathbf{t}_x, \mathbf{t}_y, \mathbf{t}_z$ (for translation) and $\mathbf{r}_x, \mathbf{r}_y, \mathbf{r}_z$ (for rotation) that span this subspace can be expressed as follows:

$$\mathbf{t}_x = \begin{pmatrix} \sqrt{m_1} \\ 0 \\ 0 \\ \sqrt{m_2} \\ 0 \\ 0 \\ \vdots \\ \sqrt{m_N} \\ 0 \\ 0 \end{pmatrix}, \mathbf{t}_y = \begin{pmatrix} 0 \\ \sqrt{m_1} \\ 0 \\ 0 \\ \sqrt{m_2} \\ 0 \\ \vdots \\ 0 \\ \sqrt{m_N} \\ 0 \end{pmatrix}, \mathbf{t}_z = \begin{pmatrix} 0 \\ 0 \\ \sqrt{m_1} \\ 0 \\ 0 \\ \sqrt{m_2} \\ \vdots \\ 0 \\ 0 \\ \sqrt{m_N} \end{pmatrix} \quad (\text{A9})$$

$$\mathbf{r}_x = \begin{pmatrix} 0 \\ -\sqrt{m_1} z_1^e \\ \sqrt{m_1} y_1^e \\ 0 \\ -\sqrt{m_2} z_2^e \\ \sqrt{m_2} y_2^e \\ \vdots \\ 0 \\ -\sqrt{m_N} z_N^e \\ \sqrt{m_N} y_N^e \end{pmatrix}, \mathbf{r}_y = \begin{pmatrix} \sqrt{m_1} z_1^e \\ 0 \\ -\sqrt{m_1} x_1^e \\ \sqrt{m_2} z_2^e \\ 0 \\ -\sqrt{m_2} x_2^e \\ \vdots \\ \sqrt{m_N} z_N^e \\ 0 \\ -\sqrt{m_N} x_N^e \end{pmatrix}, \mathbf{r}_z = \begin{pmatrix} -\sqrt{m_1} y_1^e \\ \sqrt{m_1} x_1^e \\ 0 \\ -\sqrt{m_2} y_2^e \\ \sqrt{m_2} x_2^e \\ 0 \\ \vdots \\ -\sqrt{m_N} y_N^e \\ \sqrt{m_N} x_N^e \\ 0 \end{pmatrix} \quad (\text{A10})$$

where m_i and (x_i^e, y_i^e, z_i^e) ($i = 1, \dots, N$) represent the mass and position (at the reference configuration) of the i th atom, respectively. The corresponding projection operator ($\mathbf{P}^{RT}$) can then be obtained following eq A6 as

$$\mathbf{P}^{RT} = (\mathbf{t}_x \mathbf{t}_x^T + \mathbf{t}_y \mathbf{t}_y^T + \mathbf{t}_z \mathbf{t}_z^T + \mathbf{r}_x \mathbf{r}_x^T + \mathbf{r}_y \mathbf{r}_y^T + \mathbf{r}_z \mathbf{r}_z^T) \quad (\text{A11})$$

The six vectors ($\mathbf{t}_x, \mathbf{t}_y, \mathbf{t}_z, \mathbf{r}_x, \mathbf{r}_y, \mathbf{r}_z$) spanning the translational and rotational space must be orthonormalized (using QR factorization) before calculating their projection operator ($\mathbf{P}^{RT}$). They are already orthogonal if the center of mass of the molecule lies at the origin.

Using eq A7, the force constant matrix can be projected onto a vector space devoid of translational and rotational coordinates

$$\mathbf{f}^{RT'} = (\mathbf{I} - \mathbf{P}^{RT}) \mathbf{f} (\mathbf{I} - \mathbf{P}^{RT}) \quad (\text{A12})$$

The resultant hessian matrix ($\mathbf{f}^{RT'}$) can then be diagonalized to obtain the $3N - 6$ vibrational modes of the molecule.

Wilson vectors

The elements of the $\mathbf{B}$ -matrix are derived by first representing an internal coordinate in terms of the Cartesian coordinates of the atoms, and then determining a small variation in these coordinates (to first order) through differentiation of their Cartesian expressions. For instance, to construct the Wilson vector (a column in the $\mathbf{B}$ matrix) for the bond stretching coordinate, we consider the square of the distance between two atoms, a and b , expressed in terms of their Cartesian coordinates.

$$r_{ab}^2 = (x_b - x_a)^2 + (y_b - y_a)^2 + (z_b - z_a)^2 \quad (\text{A13})$$

Differentiating eq A13, a small change along the direction connecting a and b can be obtained

$$r_{ab} \Delta r_{ab} = (x_b - x_a)(\Delta x_b - \Delta x_a) + (x_b - x_a)(\Delta y_b - \Delta y_a) + (y_b - y_a)(\Delta z_b - \Delta z_a) \quad (\text{A14})$$

Rearranging eq A14, we obtain a linear relationship between the distance variation parameter and the Cartesian displacement coordinates

$$\Delta r_{ab} = \mathbf{b}_a^{ab} \cdot \mathbf{x}_a + \mathbf{b}_b^{ab} \cdot \mathbf{x}_b = (-\hat{\mathbf{e}}^{ab}) \cdot \mathbf{x}_a + \hat{\mathbf{e}}^{ab} \cdot \mathbf{x}_b \quad (\text{A15})$$

where $\mathbf{b}_a^{ab}$ and $\mathbf{b}_b^{ab}$ are the elements of the corresponding Wilson vector and $\hat{\mathbf{e}}$ represents a unit vector directed from atom a to atom b . These quantities are related to each other in the following way

$$\begin{aligned} \mathbf{b}_b^{ab} &= -\mathbf{b}_a^{ab} \\ &= \hat{\mathbf{e}}^{ab} \\ &= \frac{(x_b - x_a)}{r_{ab}} \hat{\mathbf{i}} + \frac{(y_b - y_a)}{r_{ab}} \hat{\mathbf{j}} + \frac{(z_b - z_a)}{r_{ab}} \hat{\mathbf{k}} \end{aligned} \quad (\text{A16})$$

Also, the Cartesian displacement vectors for atoms a and b are

$$\mathbf{x}_a = \Delta x_a \hat{\mathbf{i}} + \Delta y_a \hat{\mathbf{j}} + \Delta z_a \hat{\mathbf{k}}, \mathbf{x}_b = \Delta x_b \hat{\mathbf{i}} + \Delta y_b \hat{\mathbf{j}} + \Delta z_b \hat{\mathbf{k}} \quad (\text{A17})$$

Equation A15 is nothing but the representation of the internal displacement coordinate in the basis of Cartesian displacement coordinates; hence, the Wilson vector associated with the bond stretching coordinate between atoms a and b can be written as

$$\mathbf{b}^{ab} = \begin{pmatrix} 0 \\ \vdots \\ b_{ax}^{ab} \\ b_{ay}^{ab} \\ b_{az}^{ab} \\ 0 \\ \vdots \\ b_{bx}^{ab} \\ b_{by}^{ab} \\ b_{bz}^{ab} \\ 0 \\ \vdots \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ -e_x^{ab} \\ -e_y^{ab} \\ -e_z^{ab} \\ 0 \\ \vdots \\ e_x^{ab} \\ e_y^{ab} \\ e_z^{ab} \\ 0 \\ \vdots \end{pmatrix} \quad (\text{A18})$$

Wilson vectors for any other curvilinear internal coordinate can be obtained in a similar fashion. For a bond angle (ϕ) between atoms a , b , and c (Figure A1), the corresponding Wilson vector can be constructed using the following vectors

$$\begin{aligned} \mathbf{b}_b^{abc} &= \frac{\hat{\mathbf{e}}^{ab} \cos \phi - \hat{\mathbf{e}}^{ac}}{r_{ab} \sin \phi} \\ \mathbf{b}_c^{abc} &= \frac{\hat{\mathbf{e}}^{ac} \cos \phi - \hat{\mathbf{e}}^{ab}}{r_{ac} \sin \phi} \\ \mathbf{b}_a^{abc} &= -\mathbf{b}_b^{abc} - \mathbf{b}_c^{abc} \end{aligned} \quad (\text{A19})$$

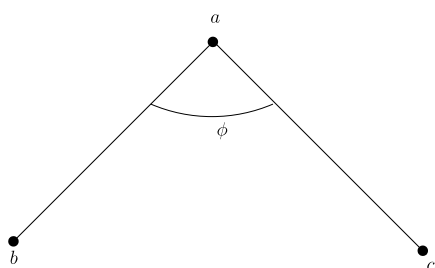
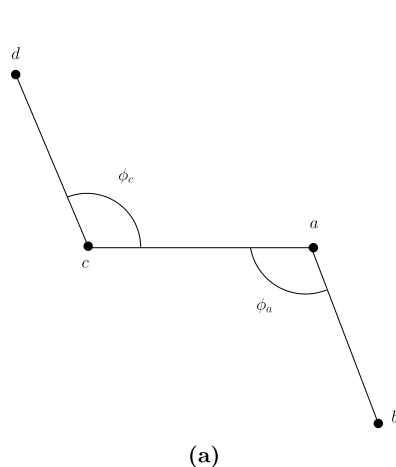


Figure A1. Bond angle between atoms a , b , and c .



and the Wilson vector associated with a dihedral angle (τ) formed by atoms a , b , c , and d (Figure A2) has the following components

$$\begin{aligned} \mathbf{b}_b^{abcd} &= -\frac{\hat{\mathbf{e}}^{ba} \times \hat{\mathbf{e}}^{ac}}{r_{ba} \sin 2\phi_a} \\ \mathbf{b}_d^{abcd} &= -\frac{\hat{\mathbf{e}}^{dc} \times \hat{\mathbf{e}}^{ca}}{r_{dc} \sin 2\phi_c} \\ \mathbf{b}_c^{abcd} &= \frac{(r_{ac} - r_{dc} \cos 2\phi_c) \hat{\mathbf{e}}^{dc} \times \hat{\mathbf{e}}^{ca}}{r_{ac} r_{dc} \sin \phi_c \sin \phi_c} \\ &\quad - \frac{\cos \phi_a \hat{\mathbf{e}}^{ba} \times \hat{\mathbf{e}}^{ac}}{r_{ac} \sin \phi_a \sin \phi_a} \\ \mathbf{b}_a^{abcd} &= \frac{(r_{ca} - r_{cd} \cos \phi_a) \hat{\mathbf{e}}^{ba} \times \hat{\mathbf{e}}^{ac}}{r_{ca} r_{ba} \sin \phi_a \sin \phi_a} - \frac{\cos \phi_c \hat{\mathbf{e}}^{dc} \times \hat{\mathbf{e}}^{ca}}{r_{ca} \sin \phi_c \sin \phi_c} \end{aligned} \quad (\text{A20})$$

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jpca.3c04395>.

A figure depicting the fully optimized structure of 1PLW, table summarizing the Ramachandran angles for the original and fully optimized 1PLW molecule, overlapping IR, Raman, and VCD spectra for 1PLW, and Cartesian coordinates for both the starting and the fully optimized structures of 1PLW (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Ankur K. Gupta – Department of Chemistry, Indiana University, Bloomington, Indiana 47405, United States; orcid.org/0000-0002-3128-9535; Email: ankur@lbl.gov

Author

Krishnan Raghavachari – Department of Chemistry, Indiana University, Bloomington, Indiana 47405, United States; orcid.org/0000-0003-3275-1426

Complete contact information is available at:

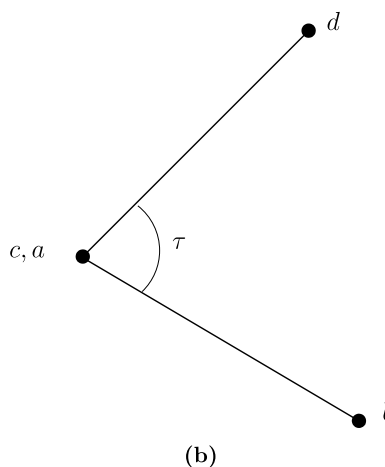


Figure A2. Dihedral angle (τ) formed by atoms a , b , c , and d .

<https://pubs.acs.org/10.1021/acs.jpca.3c04395>

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation grant CHE-2102583 at Indiana University.

REFERENCES

- (1) Wilson, E.; Decius, J.; Cross, P. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*; Dover Books on Chemistry; Dover Publications, 1980.
- (2) Kessler, J.; Kapitán, J.; Bouř, P. First-Principles Predictions of Vibrational Raman Optical Activity of Globular Proteins. *J. Phys. Chem. Lett.* **2015**, *6*, 3314–3319.
- (3) Bouř, P.; Sopkov, J.; Bednrov, L.; Maloň, P.; Keiderling, T. A. Transfer of molecular property tensors in cartesian coordinates: A new algorithm for simulation of vibrational spectra. *J. Comput. Chem.* **1997**, *18*, 646–659.
- (4) Kessler, J.; Andrushchenko, V.; Kapitán, J.; Bouř, P. Insight into vibrational circular dichroism of proteins by density functional modeling. *Phys. Chem. Chem. Phys.* **2018**, *20*, 4926–4935.
- (5) Kubelka, J.; Bouř, P.; Keiderling, T. A. *Biological and Biomedical Infrared Spectroscopy*; Barth, A., Haris, P. L., Eds.; IOS Press: Amsterdam, The Netherlands, 2009; pp 178–223.
- (6) Petrovic, A. G.; Polavarapu, P. L.; Mahalakshmi, R.; Balaram, P. Characterization of folded conformations in a tetrapeptide containing two tryptophan residues by vibrational circular dichroism. *Chirality* **2009**, *21*, E76–E85.
- (7) Lu, D.-H.; Zhao, M.; Truhlar, D. G. Projection operator method for geometry optimization with constraints. *J. Comput. Chem.* **1991**, *12*, 376–384.
- (8) Lu, D.; Truhlar, D. G. Bond-distance and bond-angle constraints in reaction-path dynamics calculations. *J. Chem. Phys.* **1993**, *99*, 2723–2738.
- (9) Wilson, E. B. A Method of Obtaining the Expanded Secular Equation for the Vibration Frequencies of a Molecule. *J. Chem. Phys.* **1939**, *7*, 1047–1052.
- (10) Wilson, E. B. Some Mathematical Methods for the Study of Molecular Vibrations. *J. Chem. Phys.* **1941**, *9*, 76–84.
- (11) Jensen, F. *Introduction to Computational Chemistry*; Wiley, 2016.
- (12) Porezag, D.; Pederson, M. R. Infrared intensities and Raman-scattering activities within density-functional theory. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 7830–7836.
- (13) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H. et al. *Gaussian 16 Revision A.03*; Gaussian Inc: Wallingford CT, 2016.
- (14) Raveh, B.; London, N.; Schueler-Furman, O. Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 2029–2040.
- (15) Marcotte, I.; Separovic, F.; Auger, M.; Gagné, S. M. A Multidimensional ¹H NMR Investigation of the Conformation of Methionine-Enkephalin in Fast-Tumbling Bicelles. *Biophys. J.* **2004**, *86*, 1587–1600.