The Journal of Chemical Physics

RESEARCH ARTICLE | DECEMBER 16 2024

DFT-based calculation of vibrational sum frequency generation spectral features of crystalline β-sheets in silk: Polarization and azimuth angle dependences

Jihyeong Ryu ; Sibing Chen; Juseok Choi ; Xing Chen ; Seong H. Kim

J. Chem. Phys. 161, 234901 (2024)
https://doi.org/10.1063/5.0236676

The Journal
of Chemical Physics

ARTICLE

pubs.aip.org/aip/jcp

DFT-based calculation of vibrational sum frequency generation spectral features of crystalline β-sheets in silk: Polarization and azimuth angle dependences

Cite as: J. Chem. Phys. **161**, 234901 (2024); doi: 10.1063/5.0236676
Submitted: 1 September 2024 • Accepted: 26 November 2024 •
Published Online: 16 December 2024

Jihyeong Ryu,[1] Sibing Chen,[2] Juseok Choi,[1] Xing Chen,[2] and Seong H. Kim[1,a)]

AFFILIATIONS

[1] Department of Chemical Engineering, Materials Research Institute, Pennsylvania State University, University Park, Pennsylvania 16802, USA

[2] Haihe Laboratory of Sustainable Chemical Transformations, Institute of Molecular Plus, School of Chemical Engineering and Technology, Tianjin University, Tianjin 300192, People's Republic of China

Note: This paper is part of the JCP Festschrift in Honor of Yuen-Ron Shen.
[a)] **Author to whom correspondence should be addressed:** shk10@psu.edu

ABSTRACT

Sum frequency generation (SFG) necessitates both noncentrosymmetry and coherence over multiple length scales. These requirements make vibrational SFG spectroscopy capable of probing structural information of noncentrosymmetric organic crystals interspersed in polymeric matrices and their three-dimensional spatial distributions within the matrices without spectral interferences from the amorphous components. However, this analysis is not as straightforward as simple vibrational spectroscopy or scattering experiments; it requires knowing the molecular hyperpolarizability of SFG-active vibrational modes and their interplay within the coherence length. This study demonstrates how density function theory (DFT) calculations can be used to construct the molecular hyperpolarizability of a model system and combine it with the SFG theory to predict the polarization and azimuth angle dependences of SFG intensities. A model system with short peptide chains mimicking β-sheet domains in *Bombyx mori* silk was chosen. SFG signals of the amide-I, II, III, and A bands and one of the CH deformation modes were simulated and compared with the experimental results and the predictions from the group theory. The SFG features of amide-I and A bands of antiparallel β-sheet could be explained with DFT-based theoretical calculations. Although vibrational coupling with neighboring groups breaks the symmetry of the D_2 point group, the group theory approach and DFT calculations gave similar results for the amide-I mode. The DFT calculation results for amide-II did not match with experimental data, which suggested vibrational coupling within a larger crystalline domain may dominate the SFG spectral features of these modes. This methodology can be applied to the structural analysis of other biopolymers.

I. INTRODUCTION

Sum frequency generation (SFG) vibrational spectroscopy has been widely used to probe the orientation and order of molecules at two-dimensional (2D) interfaces such as boundaries of gas/liquid, gas/solid, and liquid/liquid phases.[1–3] Being a nonlinear optical process, SFG requires a noncentrosymmetric arrangement of SFG-active mode in space. Since the interface of two bulk phases naturally produces the noncentrosymmetric environment, SFG can specifically probe molecular species present at the phase boundary without spectral interference of the same molecule in the bulk phase as long as the bulk phase is centrosymmetric or random.

Recently, SFG has also been demonstrated to be a powerful technique to probe three-dimensional (3D) structures of crystallites with noncentrosymmetric structures interspersed in amorphous matrices such as crystalline biopolymers in biological tissues.[4–10] In principle, conventional infrared (IR) and Raman spectroscopy can be used for the same purpose, but they often suffer from uncertainties in peak deconvolution, especially when the crystalline fraction inside the sample is low and the amorphous peaks overlap with the crystalline peaks.[11–15] In contrast, the amorphous phase cannot contribute to the SFG signal since it does not meet the noncentrosymmetry requirement;[4–6] thus, SFG can selectively detect the crystalline domains with noncentrosymmetric structures without interference from the amorphous matrix. In addition, the coherence requirement between the probe beams and the signal beam makes SFG sensitive to not only the nanoscale structure of the crystalline domains[5,6,16] but also the spatial arrangement and collective orientation of crystalline domains at the mesoscale.[17,18] Thus, SFG can provide the 3D structural information that other techniques may not be able to find.

Although the noncentrosymmetry and coherence requirements make SFG unique and powerful, at the same time those make SFG spectral interpretation inherently complicated, especially for 3D bulk materials.[19,20] This is because too many variables are involved in SFG analysis. Those include the hyperpolarizabilities of individual SFG-active domains in the molecular coordinate and their interplays in the laboratory coordinate in conjunction with experimental geometry (such as incidence plane and angle as well as the polarization of three probe beams—two input beams and one signal beam).[20] The parameters that are not considered in the 2D analysis, such as inter-particle distance, relative polarity, and orientational distributions along multiple axes, are involved in the 3D analysis.

Since the SFG-active mode must be both Raman-active and IR-active, one may think that comparing Raman and IR spectra of the analyte of interest can be helpful for the identification and assignment of peaks in SFG spectra. However, this approach does not work sufficiently well since it does not consider the interplay of the relative phases of vibrational modes within the SFG-coherence length.[19] For the meso-scale structural effect, one may consider preparing or obtaining a reference sample with precisely controlled or well-characterized geometry.[4,6,21] However, such samples are often unavailable, especially for crystalline biopolymer samples, since their crystal structure and crystallinity are often controlled during the synthesis in biological organisms.[21–23]

One could rely on the theoretical calculation to resolve such predicaments in SFG analysis of 3D structural analysis of bulk samples consisting of noncentrosymmetric crystalline domains interspersed in amorphous matrices. The theoretical calculation requires knowing the hyperpolarizability of individual SFG-active vibrational modes. For small molecules, hyperpolarizability can be obtained from the single-bond polarizability derivative (SBPD) model.[19] However, the SBPD model assumes the ideal molecular symmetry, which often breaks due to coupling with vibrational modes of neighboring functional groups.[24] For protein molecules, the peptide bond vibrations (such as amide-I mode) could also be modeled considering the group theory of the secondary structure of the protein.[25,26] However, again, this does not consider vibrational coupling among neighboring functional groups;[21,22] thus, its applicability or accuracy is somewhat limited.

A more comprehensive approach would be to use first-principles calculations to obtain the hyperpolarizability of the analyte at the molecular coordinate. Once that is obtained, the SFG intensity can be estimated or predicted by applying the Euler transformation to project the molecular hyperpolarizability onto the laboratory coordinate, from which coupling with the electric field of incident and outgoing lights can be calculated.[19,20,24] This approach has been demonstrated for SFG analysis of small molecules at 2D interfaces.[24,27,28] The same approach can be applied to the 3D bulk analysis, but the actual practice is more complicated than the 2D case since some of the assumptions or simplifications used in the 2D case are not applicable to the 3D case.

In this study, we demonstrate the theoretical calculation of SFG intensities using the Raman polarizability and IR transition dipole moment tensors obtained from density functional theory (DFT). The target system of interest is the β-sheet domain present in silk fibers produced by arthropods and insects. Elucidating its roles in the mechanical properties of silk fibers is of great interest in materials science and engineering.[29–31] We have chosen the silk fiber of *Bombyx mori* cocoon for this study because the crystalline structure of the β-sheet in the *B. mori* fiber is well characterized and available in the literature.[32–34] The repetitive amino acid sequence for the β-sheet produced by silkworms is poly(glycidylalanine).[35,36] This structure was modeled with a tetrameric peptide unit, Gly(G)–Ala(A)–Gly(G)–Ala(A); two units of the GAGA peptide were placed laterally, allowing inter-chain C=O· · ·H–N hydrogen bonding interactions in the antiparallel direction, following the antiparallel packing of peptide strands in the β-sheet.[33,37] The polarization and azimuth angle dependence of SFG signals were calculated and compared with the experimental SFG spectra of *B. mori* silk fiber. For comparison, the approach based on the molecular symmetry of the peptide group is also used for the amide-I modes.[25,26]

II. METHODS

A. SFG, ATR-IR, and Raman analyses

Degummed silkworm (*B. mori*) fiber was purchased from Living Dreams Yarn Co. and used as received. SFG experiments were performed using 800 nm (pulse width ~85 fs, 2 kHz repetition rate) from a Ti:Sapphire femtosecond laser system and tunable mid-IR (bandwidth = 120–150 cm^{-1}) from an optical parametric generation/amplification (OPG/OPA) system. A detailed description of the SFG system can be found elsewhere.[38] Both 800 nm and mid-IR incident beams were focused through a BaF_2 lens (focal point, $f = 15$ cm) on a silk fiber placed in the XY plane at an incident angle of 45° with respect to the surface normal. In this geometry, the laser incidence plane was the XZ plane, and the azimuth angle was the angle between the fiber axis and the X-axis. The SFG signal was detected in the reflection mode with volume-phase holographic grating and a CCD camera. Two separate gratings were used for the low wavenumber (1100–2000 cm^{-1}) and high wavenumber (2500–4000 cm^{-1}) regions. Multiple spots were analyzed in the 1000–1800 and 2700–3800 cm^{-1} spectral ranges with intervals of 100 cm^{-1} for the broadband IR setting. The notation of the three-letter polarization combination follows the order of SFG signal—narrowband 800 nm input beam—broadband IR input beam. For example, *psp* polarization indicates that SFG polarization is *p*, 800 nm polarization is *s*, and IR polarization is *p* with respect to the laser incidence plane.

To get statistically representative SFG spectra, individual spectra were normalized with the highest peak (i.e., 1262 cm^{-1} for the low wavenumber region and 3275 cm^{-1} for the high wavenumber region), then averaged and multiplied by the averaged intensity of the highest peak; in this way, the relative intensities among various spectral features in each spectral window were unaltered, allowing the azimuth angle dependence study without being influenced by local compositional and morphological variances of the fiber. The SFG spectra were collected with *ppp*, *psp*, *spp*, and *ssp* polarizations. The other four polarizations (*pps*, *pss*, *sps*, and *sss*) were not used, since the *psp*, *spp*, *ssp*, and *ppp* polarizations gave sufficient dataset to determine the orientation information of the crystalline β-sheet with respect to the fiber axis.

Fourier transform infrared (FT-IR) spectroscopy measurements were performed using a Bruker Vertex 80 spectrometer. The attenuated total reflectance (ATR) IR spectra of the fiber were collected with a diamond internal reflection element. The ATR-IR spectrum of the silk sample was analyzed from 400 to 4000 cm^{-1}, averaging over 400 scans. Raman analysis was carried out using a Horiba LabRAM HR system with a 644 nm excitation laser source and a 300 gr/nm grating.

B. Theoretical calculation of SFG intensity using tensors obtained from DFT

1. Model system of antiparallel β-sheet

Figure 1 illustrates two conformations of the antiparallel β-sheet structure of *B. mori* silk. In the molecular coordinate, the peptide chain is aligned along the *c* axis, the inter-chain C=O· · ·H–N hydrogen bonds are along the *b* axis, and the stacking of the sheets formed through the inter-chain van der Waals interactions is along the *a* axis. In 1955, Marsh *et al.*[39] proposed a polar conformation of the antiparallel β-sheet structure [Figs. 1(a)–1(c)] in which the β-strands are packed with all methyl groups of the alanine units pointing the same direction along the *a* axis. In 1999, Takahashi *et al.*[32] proposed an antipolar conformation in which the methyl groups in each chain are pointing in the same direction but the directions among adjacent chains are rotated by 180° alternatively [Figs. 1(d) and 1(e)]. The antipolar conformation results are discussed in the main text since it is energetically more stable, and the polar conformation results are presented in the supplementary material for comparison and discussed separately at the end.

2. DFT calculations of polarizability and dipole derivative tensors

The Raman polarizability and IR transition dipole moment tensors were acquired using density functional theory (DFT). All quantum chemical calculations were performed using the NWChem software package.[40] DFT calculations were conducted with the B3LYP/6-311++G(d,p) basis set. The molecular geometries were optimized with the default convergence criteria for maximum and root mean square (rms) forces and displacements to ensure the accuracy and stability of the optimized structures. The static polarizability was calculated using linear response theory, which involves applying a uniform electric field. The polarizability and dipole moment derivative tensors with respect to the mass-weighted normal mode ($\frac{\partial\alpha}{\partial Q_q}$ and $\frac{\partial\mu}{\partial Q_q}$, respectively) were obtained using the three-point numerical differentiation approach as follows:

$$\frac{\partial\alpha}{\partial Q_q} = \frac{\alpha(Q_q + \Delta Q_q) - \alpha(Q_q - \Delta Q_q)}{2s_Q\Delta Q_q} = \begin{pmatrix} \alpha'_{aa} & \alpha'_{ab} & \alpha'_{ac} \\ \alpha'_{ba} & \alpha'_{bb} & \alpha'_{bc} \\ \alpha'_{ca} & \alpha'_{cb} & \alpha'_{cc} \end{pmatrix}, \tag{1}$$

$$\frac{\partial\mu}{\partial Q_q} = \frac{\mu(Q_q + \Delta Q_q) - \mu(Q_q - \Delta Q_q)}{2s_Q\Delta Q_q} = \begin{pmatrix} \mu'_a \\ \mu'_b \\ \mu'_c \end{pmatrix}, \tag{2}$$

where s_Q is the mass-weighted step size. These calculations were performed at the equilibrium geometry distorted with the normal mode coordinates. The vibrational peak positions from the Hessian matrix analysis were adjusted using a scaling factor of 0.955 for the antipolar structure and 0.957 for the polar structure.

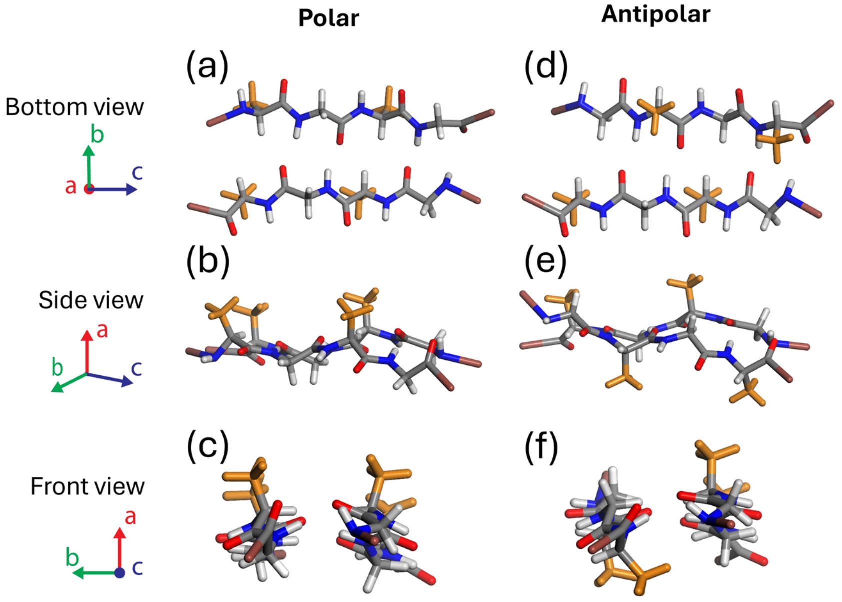

FIG. 1. Model of antiparallel β-sheet moiety constructed with two GAGA tetrameric peptide units in (a)–(c) polar conformation proposed by Marsh *et al.* in 1955,[39] and (d)–(f) antipolar conformation proposed by Takahashi *et al.* in 1999.[32] The C, O, N, and H atoms are shown in gray, red, blue, and white colors, and the methyl side group of alanine is in orange color. Both terminal groups are iodinated (dark brown) to eliminate the end group effects in the DFT calculation. The *a*, *b*, and *c* axes of the molecular coordinate is along the inter-sheet, inter-chain, and chain directions, respectively. The C=O· · ·H–N hydrogen bond is along the *b* axis within the *bc* plane.

The β-sheet structures were modeled with the models shown in Fig. 1 to run DFT calculations within the reasonable calculation limit. The chain end groups were replaced with iodine atoms for two reasons. Without this replacement, the hydrogen bonding interactions between the terminal NH_2 and COOH groups distorted the chain conformation in the middle, affecting the inter-chain hydrogen bonding interactions along the *b* axis. In addition, their vibrations were not the main focus of the study. By replacing the terminal hydrogen of primary amine and the hydroxyl group of carboxylic acid with heavy iodine atoms, we could mimic the large mass of other parts of the β-sheet structure to which the tetrameric unit is connected.[24] Note that we cannot model the effect of van der Waals stacking along the *a* axis since only two tetrameric peptides were used.

Using the polarizability and dipole derivative tensors obtained from DFT calculations, theoretical Raman and IR spectra of the antiparallel pair of the GAGA peptide unit (mimicking the antiparallel β-sheet) were calculated as follows:[41]

$$I_{Raman} = 45\overline{\alpha}^2 + 7\gamma^2, \tag{3}$$

$$I_{IR} = \left(\mu_a'\right)^2 + \left(\mu_b'\right)^2 + \left(\mu_c'\right)^2, \tag{4}$$

where $\overline{\alpha} = \frac{1}{3}\left(\alpha_{aa}' + \alpha_{bb}' + \alpha_{cc}'\right)$ and $\gamma^2 = \frac{1}{2}\Big[\left(\alpha_{aa}' - \alpha_{bb}'\right)^2 + \left(\alpha_{bb}' - \alpha_{cc}'\right)^2 + \left(\alpha_{cc}' - \alpha_{aa}'\right)^2 + 6\left(\alpha_{ab}'^{\,2} + \alpha_{ac}'^{\,2} + \alpha_{bc}'^{\,2}\right)\Big]$.

3. Numerical simulation of SFG intensities using tensors from DFT

Details of the SFG intensity simulation were published elsewhere,[18,20] and only the general procedure is described here. The hyperpolarizability ($\beta_{i'j'k'}$) of the vibrational mode of the antiparallel GAGA-peptide dimer was obtained through the tensor product of $\frac{\partial \alpha}{\partial Q_q}$ and $\frac{\partial \mu}{\partial Q_q}$,

$$\beta_{i'j'k'} = \frac{\partial \alpha}{\partial Q_q} \otimes \frac{\partial \mu}{\partial Q_q} = \begin{pmatrix} \begin{pmatrix} \alpha_{aa}'\mu_a' \\ \alpha_{aa}'\mu_b' \\ \alpha_{aa}'\mu_c' \end{pmatrix} & \begin{pmatrix} \alpha_{ab}'\mu_a' \\ \alpha_{ab}'\mu_b' \\ \alpha_{ab}'\mu_c' \end{pmatrix} & \begin{pmatrix} \alpha_{ac}'\mu_a' \\ \alpha_{ac}'\mu_b' \\ \alpha_{ac}'\mu_c' \end{pmatrix} \\ \begin{pmatrix} \alpha_{ba}'\mu_a' \\ \alpha_{ba}'\mu_b' \\ \alpha_{ba}'\mu_c' \end{pmatrix} & \begin{pmatrix} \alpha_{bb}'\mu_a' \\ \alpha_{bb}'\mu_b' \\ \alpha_{bb}'\mu_c' \end{pmatrix} & \begin{pmatrix} \alpha_{bc}'\mu_a' \\ \alpha_{bc}'\mu_b' \\ \alpha_{bc}'\mu_c' \end{pmatrix} \\ \begin{pmatrix} \alpha_{ca}'\mu_a' \\ \alpha_{ca}'\mu_b' \\ \alpha_{ca}'\mu_c' \end{pmatrix} & \begin{pmatrix} \alpha_{cb}'\mu_a' \\ \alpha_{cb}'\mu_b' \\ \alpha_{cb}'\mu_c' \end{pmatrix} & \begin{pmatrix} \alpha_{cc}'\mu_a' \\ \alpha_{cc}'\mu_b' \\ \alpha_{cc}'\mu_c' \end{pmatrix} \end{pmatrix}$$

$$= \begin{pmatrix} \begin{pmatrix} \beta_{aaa} \\ \beta_{aab} \\ \beta_{aac} \end{pmatrix} & \begin{pmatrix} \beta_{aba} \\ \beta_{abb} \\ \beta_{abc} \end{pmatrix} & \begin{pmatrix} \beta_{aca} \\ \beta_{acb} \\ \beta_{acc} \end{pmatrix} \\ \begin{pmatrix} \beta_{baa} \\ \beta_{bab} \\ \beta_{bac} \end{pmatrix} & \begin{pmatrix} \beta_{bba} \\ \beta_{bbb} \\ \beta_{bbc} \end{pmatrix} & \begin{pmatrix} \beta_{bca} \\ \beta_{bcb} \\ \beta_{bcc} \end{pmatrix} \\ \begin{pmatrix} \beta_{caa} \\ \beta_{cab} \\ \beta_{cac} \end{pmatrix} & \begin{pmatrix} \beta_{cba} \\ \beta_{cbb} \\ \beta_{cbc} \end{pmatrix} & \begin{pmatrix} \beta_{cca} \\ \beta_{ccb} \\ \beta_{ccc} \end{pmatrix} \end{pmatrix}. \tag{5}$$

This was projected to the laboratory coordinate using the Euler transformation matrix to obtain the second-order susceptibility, $\chi_{ijk}^{(2)}$, for the *q*th mode with a resonance frequency of ω_q and a damping factor of Γ_q,[20,42]

$$\chi_{ijk}^{(2)} \propto N \sum_{i',j',k'} \frac{\left\langle R_{ii'} R_{jj'} R_{kk'} \beta_{i'j'k'} \right\rangle}{\omega_{IR} - \omega_q + i\Gamma_q}, \tag{6}$$

where *N* is the number density of the SFG-active unit within the coherence length, $R_{ii'} R_{jj'} R_{kk'}$ is the products of Euler transformation matrices that convert the tensor in the molecular coordinate (i', j', and k' which can be any of *a*, *b*, and *c*) to the tensor in the laboratory coordinate (*i*, *j*, and *k* which can be any of *X*, *Y*, and *Z*) using azimuth angle (ϕ), tilt angle (θ), and twist angle (ψ). Then, the effective susceptibility, $\chi_{eff}^{(2)}$, was calculated to be[19,20]

$$\chi_{eff}^{(2)} = \sum_{i,j,k} C_{Fres} \hat{e}_{proj} \chi_{ijk}^{(2)}, \tag{7}$$

where C_{Fres} and $\hat{e}_{proj}$ terms describe the tensorial Fresnel factors and projected electric field vectors, respectively, of the SFG, 800 nm, and IR beams.

For a single unit in the coherence length (i.e., $N = 1$), the SFG intensity normalized with the IR and 800 nm beam intensities was calculated as follows:

$$\frac{I_{SFG}}{I_{800} I_{IR}} \propto \left|\chi_{eff}^{(2)}\right|^2. \tag{8}$$

For the case with multiple domains (i.e., $N \geq 2$) in 3D space (such as crystalline β-sheet domains in silk fiber), we must consider the random quasi-phase match condition, and the SFG intensity was calculated with the following equation:[18]

$$\frac{I_{SF}}{I_{800} I_{IR}} \propto \left| \sum_{i,j,k} C_{Fres} \hat{e}_{proj} \sum_{f=1}^{N-1} \chi_{ijk}^{(2)} \left(\frac{e^{-i\Delta k \cdot d} - 1}{\Delta k}\right) \left(e^{-i\Delta k (d+\Delta l) f}\right) \right|^2, \tag{9}$$

where *d* is the size of the individual domain, Δl is the inter-domain distance, and Δk is the phase mismatch among the SFG, 800 nm, and IR beams. The $\pi/\Delta k$ determines the coherence length (L_C). Here, the $\left(\frac{e^{-i\Delta k \cdot d} - 1}{\Delta k}\right)$ term represents the phase synchronization within the L_C, the $\left(e^{-i\Delta k (d+\Delta l) f}\right)$ term accounts for the random quasi-phase matching process,[18,43] and the *N* term in Eq. (6) is replaced with summation.

Numerical SFG simulations were performed using *Mathematica* software. The details of numerical SFG simulation are described elsewhere,[18,20] and one example is provided in the supplementary material. The refractive index was taken from the literature.[44] Since the fiber was placed in the XY plane and rotated around the Z axis, the tilt angle (θ) with respect to the Z axis was set to 90° and the azimuth angle (ϕ) was varied. The twist angle (ψ) was assumed to be random, i.e., the β-sheet domain can freely rotate about the *c* axis inside the silk fiber. This assumption allowed us to integrate Eq. (8) over ψ from 0° to 360° at given ϕ angles at a fixed tilt angle ($\theta = 90°$) for the single domain case (i.e., $N = 1$). The calculation was done for *ppp*, *psp*, *spp*, and *ssp* polarizations with ϕ varying from 0° to 360° with an increment of 1.5°.

For the multiple domain case, we estimated the crystalline β-sheet domain size (d) from the full-width-half-maximum (FWHM) of the (002) peak in x-ray diffraction using the Scherrer equation;[45] from this, d was set to be 3.5 nm. Then, the N and Δl values were estimated from the volume fraction of the crystalline β-sheet domains in the silk fiber. Assuming that the β-sheet domains were homogeneously distributed in the fiber, the volume fraction of 50% and the coherence length (L_C) of 260 nm corresponded to $N = 18$ and $\Delta l = 10.5$ nm.[45] It is known that the c axis of the β-sheet is aligned along the fiber axis with a certain angular distribution.[45] The standard deviation of the c axis alignment (σ_φ) with respect to the fiber axis was included as a variable in the calculation. Since the fiber surface was rounded, the incident angle varied locally within the laser irradiated area. This would be equivalent to a variance in the effective local tilt angle and was modeled by varying the tilt angle distribution (σ_θ). The net polarity of multiple domains differs from the polarity of a single domain depending on their polarity distribution within the L_C scale. The net polarity was defined with the directional excess (DE) in the following way:[17,18]

$$DE = \left|\frac{\text{Sum of the polarity of individual dimer}}{\text{Total number of dimers within } L_C}\right| \times 100(\%), \qquad (10)$$

where the polarity of the individual domain along the b axis (direction of C=O$\cdots$H–N) could be either +1 or −1 depending on its directionality. When all these variables (σ_φ, σ_θ, and DE) were considered simultaneously, Eq. (9) could not be integrated analytically. So, the calculation was done numerically by creating 2000 replicas with random probability functions with given constraints for σ_φ, σ_θ, and DE, calculating SFG intensities of these 2000 replicas using Eq. (9), and then averaging the calculation results, which can be summarized as follows:

$$I_{SF}^{\text{caln}} = \sum_{m=1}^{M_n} \frac{I_{SF}(m)}{I_{VIS} I_{IR}} / M_n, \qquad (11)$$

where m is the individual case with different σ_φ, σ_θ, and DE, and M_n was set to 2000. Taking the arithmetic average can be justified with the random quasi-phase matching principle.[18,43] The calculation with Eq. (11) was done for *ppp*, *psp*, *spp*, and *ssp* polarizations with ϕ varying from 0° to 360° with an increment of 1.5°.

III. RESULT AND DISCUSSION

A. SFG spectra of crystalline β-sheet domains in *B. mori* silk fiber

Figure 2 displays SFG spectra collected at three azimuth angles, $\phi = 0°$, 45°, and 90°, with *ppp*, *psp*, *spp*, and *ssp* polarization combinations.[45] Note that when fibers were rotated to acquire SFG spectra at different azimuth angles, the laser beam spot moved to different locations of the fibers. In addition, the distribution of incident angles of the 800 nm and IR beam with respect to the normal direction of the rounded fiber surface also changed as the sample was rotated. As a result, the absolute intensities of SFG spectra collected at different azimuth angles could not be compared directly. Furthermore, different gratings were used in the monochromator for the low and high wavenumber regions, so the relative intensities of the amide-I, II, and III modes vs the amide-A mode could not be compared directly. However, within a given spectral region, the relative intensities of spectral features collected with different polarizations at the same azimuth angle could still be compared because they were obtained without moving the sample or the beam spot—only the polarizations of the 800 nm incoming beam and the SFG signal were rotated.

In general, the polarization dependence of all vibrational peaks exhibits similar patterns. At $\phi = 0°$, the intensities of all peaks in the amide-I, II, and III band regions are slightly larger in the *spp* polarization than in the other three cases, i.e., $I_{spp} > I_{ppp} \approx I_{ssp} > I_{psp}$ [Fig. 2(a)]. The amide-A peak exhibits a slightly different pattern: $I_{spp} > I_{ppp} > I_{ssp} \approx I_{psp}$ [Fig. 2(a)]. At $\phi = 45°$, the relative intensity pattern is $I_{psp} \approx I_{ssp} \gg I_{spp} \approx I_{ppp}$ in both low and high wavenumber regions [Fig. 2(b)]. At $\phi = 90°$, the relative intensity pattern is $I_{psp} \gg I_{spp} \approx I_{ssp} \approx I_{ppp}$ in all wavenumber regions [Fig. 2(c)]. The relative intensities between the 1567 cm^{-1} peak (amide-II) and the 1262 cm^{-1} peak (amide-III) do not change significantly with the azimuth angle change, but the relative intensity of the amide-I band at 1632 cm^{-1} with respect to these two bands is changing dramatically, that is, $I_{1632} \approx I_{1567}$ at $\phi = 0°$, but $I_{1632} < I_{1567}$ at $\phi = 90°$.

B. IR and Raman spectra predicted by DFT calculation

Figure 3 compares the experimental Raman and IR spectra of *B. mori* (silkworm) cocoon fibers with the theoretical spectra predicted using the tensors from DFT calculations for the antipolar antiparallel GAGA-peptide dimer [Figs. 1(d)–1(f)]. Although the calculated peak positions do not match precisely with the experimental spectra, the overall trends are in reasonable agreement. We do not know the damping factors of individual vibrational modes, affecting peak intensities and widths, which will be a function of interactions with neighboring functional groups and the rigidity of the crystalline structure. Nonetheless, we can still compare the experimentally computed spectra and select which mode of the computation results corresponds to which peak in the experimental data based on the polarization dependence.[45] The rest of the discussion is based on the comparison of these selected modes with the experimental data.

C. Simulation of SFG intensity for amide-I band of a single domain

The amide-I modes consist of ~75% C=O stretch, ~15% CN stretch, and ~12% C$_\alpha$CN deformation.[46] Table I shows the normalized Raman polarizability tensor $\left(\frac{\partial \alpha_{i'j'}}{\partial Q_q}\right)$ and IR transition dipole moment tensor $\left(\frac{\partial \mu_{k'}}{\partial Q_q}\right)$ in matrix form for the amide-I modes estimated using the D$_2$ point group (to which the antiparallel β-sheet belongs) and DFT calculations of the antipolar antiparallel GAGA-peptide dimer. The D$_2$ point group calculation was done following the methods described by Nguyen *et al.*[25] and Ye *et al.*[26]

In the previous works,[25,26] the normal modes in the amide-I region were analyzed with the assumption that the peptide bond in the antiparallel β-sheet is perfectly aligned with the symmetry constraints of the D$_2$ point group, and they can be modeled as

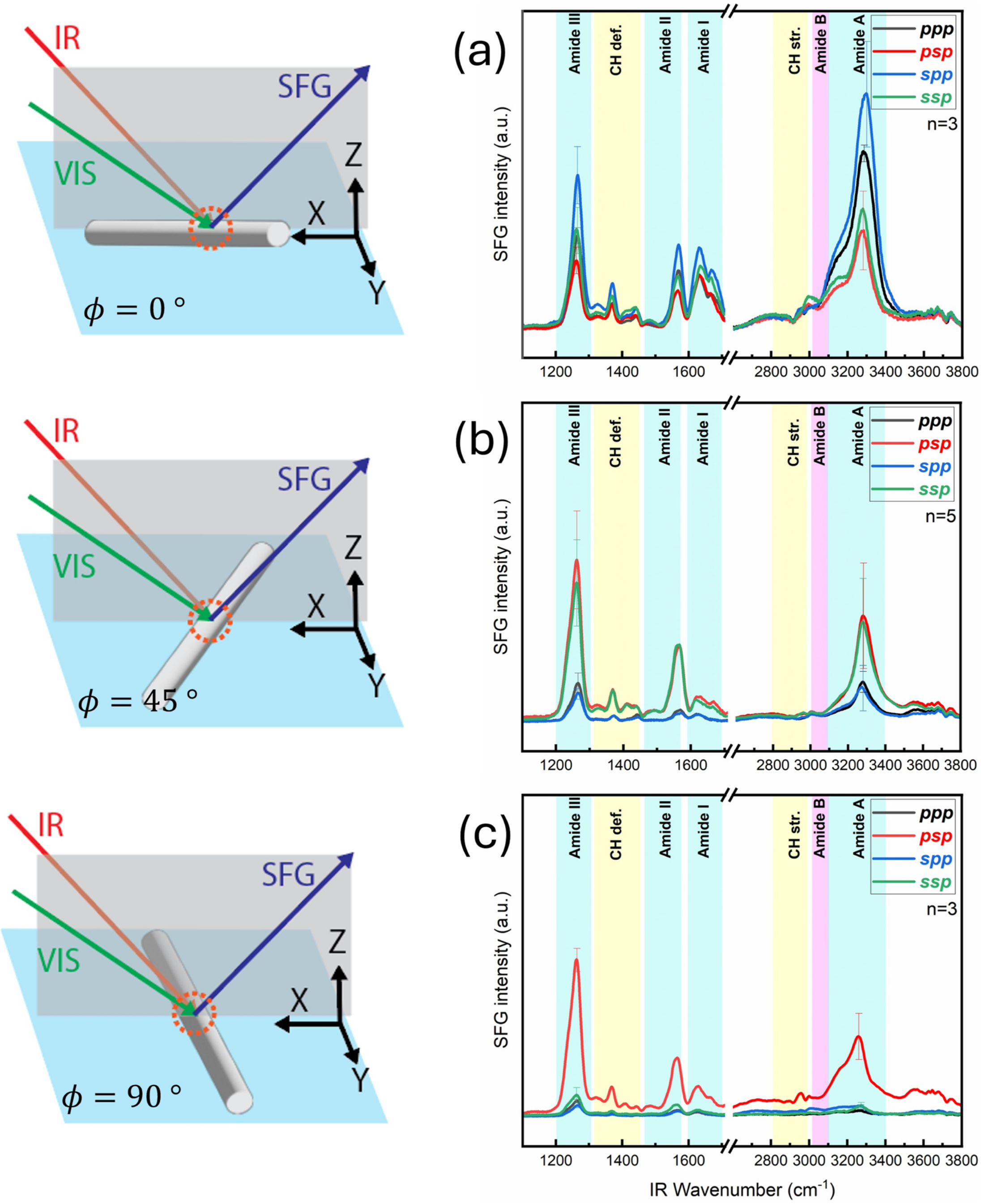

FIG. 2. SFG spectra, collected with four different polarization combinations, of degummed *B. mori* silk fiber placed within the XY plane of the lab coordinate with the azimuth angle of (a) $\phi = 0°$ (fiber axis along the X axis), (b) $\phi = 45°$ (diagonal direction in the XY plane), and (c) $\phi = 90°$ (fiber axis along the Y axis). The laser incidence plane is in the XZ plane. Note that the fibers have no directionality; thus, $\phi = 0°$ is equivalent to $\phi = 180°$, and $\phi = 90°$ is equivalent to $\phi = 270°$. The data shown here are the average of three measurements, and the standard deviations at selected peaks are shown with error bars. The spectra are replotted from Ref. 45.

an isolated vibrational mode without coupling with side-chain functional groups. Since the stretch directions of C=O in the adjacent peptide bonds can be in-phase (0) or out-of-phase (π) with respect to each other along the chain direction (*c* axis) and the inter-chain direction (*b* axis), there are four possible symmetries: A(0,0), B1(0,π), B2(π,0), and B3(π,π).[25,46–48] In the parentheses, the first character is the intra-chain phase, and the second is the inter-chain phase.

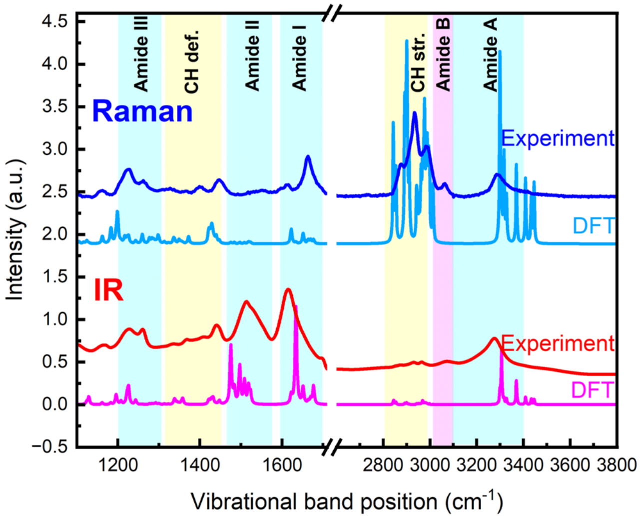

FIG. 3. Raman and ATR-IR spectra of degummed *B. mori* silk fiber and Raman and IR peaks calculated with the DFT-based tensors for the antipolar–antiparallel GAGA-peptide dimer. The peak positions of the DFT-calculated vibrational modes are adjusted with a scaling factor of 0.955, and the FWHM of the peak is set to 10 cm^{-1}.

The A symmetry of the D$_2$ point group has non-zero values for α'_{aa}, α'_{bb}, and α'_{cc}; and zero for all other components in the polarizability tensor; and zero in the dipole moment tensor (Table I). This mode is IR-inactive and, thus, SFG-inactive because all $\beta_{i'j'k'}$ are zero [Fig. 4(a)].

The B1 symmetry has non-zero elements at α'_{ab}, α'_{ba}, and μ'_c (Table I); so, it has non-zero hyperpolarizability components, i.e., $\beta_{abc} = \beta_{bac} \neq 0$, which go to the numerator terms in Eq. (6). However, the $\chi^{(2)}_{eff}$ term calculated with Eq. (7) is zero for all polarization combinations [Fig. 4(a)]. This is due to the symmetry cancellation in the laboratory coordinate when $\theta = 90°$ and ψ is freely rotatable. In this geometry, any modes with $\mu'_a = \mu'_b = 0$ and non-zero μ'_c (that is, its IR dipole is along the *c* axis) gives no SFG signal, regardless of the polarizability tensor, except the case where α'_{aa}, α'_{bb}, and α'_{cc} are non-zero, and all other components in the polarizability tensor are zero (see Fig. S1 in the supplementary material). Even in this exception case, the SFG polarizations sensitive to chiral contributions (i.e., *psp* and *spp*) are zero, and only *ppp* and *ssp* polarizations give an SFG signal when $\phi = 0°$.

The B2 mode, in which adjacent C=O groups stretch out-of-phase in the chain direction and in-phase in the inter-chain direction, is the main one that gives a strong signal in SFG [Fig. 4(a)]. It has two main components, β_{acb} and β_{cab}, in the hyperpolarizability tensor (Table I). In the *psp* polarization, which gives the strongest signal among the four polarizations, the β_{acb} component gives a dominant contribution at $\phi = 90°$ and 270°, and the β_{cab} component contributes weakly in the $\phi = 0°$ and 180° directions (Fig. S2). The *spp* polarization has a weak signal in the $\phi = 0°$ and 180° directions only. In contrast, the *ppp* and *ssp* polarizations have weak signals in the diagonal angles ($\phi = 45°$, 145°, 225°, and 315°; i.e., four lobes along the diagonal directions in the polar plot). This is in qualitative agreement with the experimental data shown in Fig. 2. However, we can see that at $\phi = 45°$, the D$_2$ point group predicts $I_{ssp} > I_{ppp} > I_{psp} > I_{spp}$, which is inconsistent with the experimental data in Fig. 2. This discrepancy is because the calculation result shown here was for a single domain and did not consider the angular distribution of multiple domains (see Sec. III E).[45]

For the B3 mode, the IR dipole moments along the *b* axis are canceled out, and only a small offset from the *bc* plane (i.e., along the *a* axis) remains (Table I).[48] Its IR intensity is small, and so is its SFG intensity [Fig. 4(a)]. Except for its intensity, the polarization and azimuth angle dependencies of the B3 mode are identical to those of the B2 mode (Fig. S3). Because the B3 mode intensity is so weak, it is difficult to reliably find the peak that can be assigned to this mode in

TABLE I. Raman polarizability tensor and IR dipole moment of the amide-I modes calculated using the D$_2$ point group of the antipolar antiparallel β-sheet and predicted with DFT calculations for the antipolar antiparallel GAGA-peptide dimer model. Tensor components are normalized with the largest value of the 1632–1634 cm^{-1} mode of each structure. The normalization factors are −9.08 and 3.80 for the polarizability and dipole moment tensors of the D$_2$ point group, respectively. For the antipolar antiparallel GAGA-peptide dimer, the normalization factors are 1.17 and 2.98 for the polarizability and dipole moment tensors, respectively. Bold numbers indicate the elements with the largest number in each vibrational mode.

D$_2$ point group	$\left(\frac{\partial\alpha_{i'j'}}{\partial Q_q}\right)\left(\frac{\partial\mu_{k'}}{\partial Q_q}\right)_{\text{peptide bond model}}$	$\left(\frac{\partial\alpha_{i'j'}}{\partial Q_q}\right)\left(\frac{\partial\mu_{k'}}{\partial Q_q}\right)_{\text{DFT}}$	Exp cm^{-1}	DFT cm^{-1}
A $v(0,0)$	$\begin{pmatrix} 1.02 & 0 & 0 \\ 0 & \mathbf{7.71} & 0 \\ 0 & 0 & 2.17 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$	$\begin{pmatrix} -0.22 & 0.08 & 0.37 \\ 0.08 & \mathbf{2.64} & -0.82 \\ 0.37 & -0.82 & 0.55 \end{pmatrix}\begin{pmatrix} -0.05 \\ \mathbf{-0.16} \\ 0.02 \end{pmatrix}$	1663	1664
B1 $v(0,\pi)$	$\begin{pmatrix} 0 & \mathbf{1.22} & 0 \\ \mathbf{1.22} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ 0 \\ \mathbf{-0.27} \end{pmatrix}$	$\begin{pmatrix} 0.04 & 0.49 & -0.35 \\ 0.49 & -0.76 & \mathbf{1.41} \\ -0.35 & \mathbf{1.41} & 0.65 \end{pmatrix}\begin{pmatrix} -0.09 \\ 0.25 \\ \mathbf{-0.35} \end{pmatrix}$	1698	1677
B2 $v(\pi,0)$	$\begin{pmatrix} 0 & 0 & \mathbf{-1} \\ 0 & 0 & 0 \\ \mathbf{-1} & 0 & 0 \end{pmatrix}\begin{pmatrix} 0 \\ \mathbf{1} \\ 0 \end{pmatrix}$	$\begin{pmatrix} 0.15 & -0.45 & \mathbf{1} \\ -0.46 & 0.65 & 0.28 \\ \mathbf{1} & 0.28 & 0.29 \end{pmatrix}\begin{pmatrix} -0.09 \\ \mathbf{1} \\ 0.10 \end{pmatrix}$	1632	1634
B3 $v(\pi,\pi)$	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \mathbf{-2.11} \\ 0 & \mathbf{-2.11} & 0 \end{pmatrix}\begin{pmatrix} \mathbf{0.17} \\ 0 \\ 0 \end{pmatrix}$	Not identified	1667	⋯

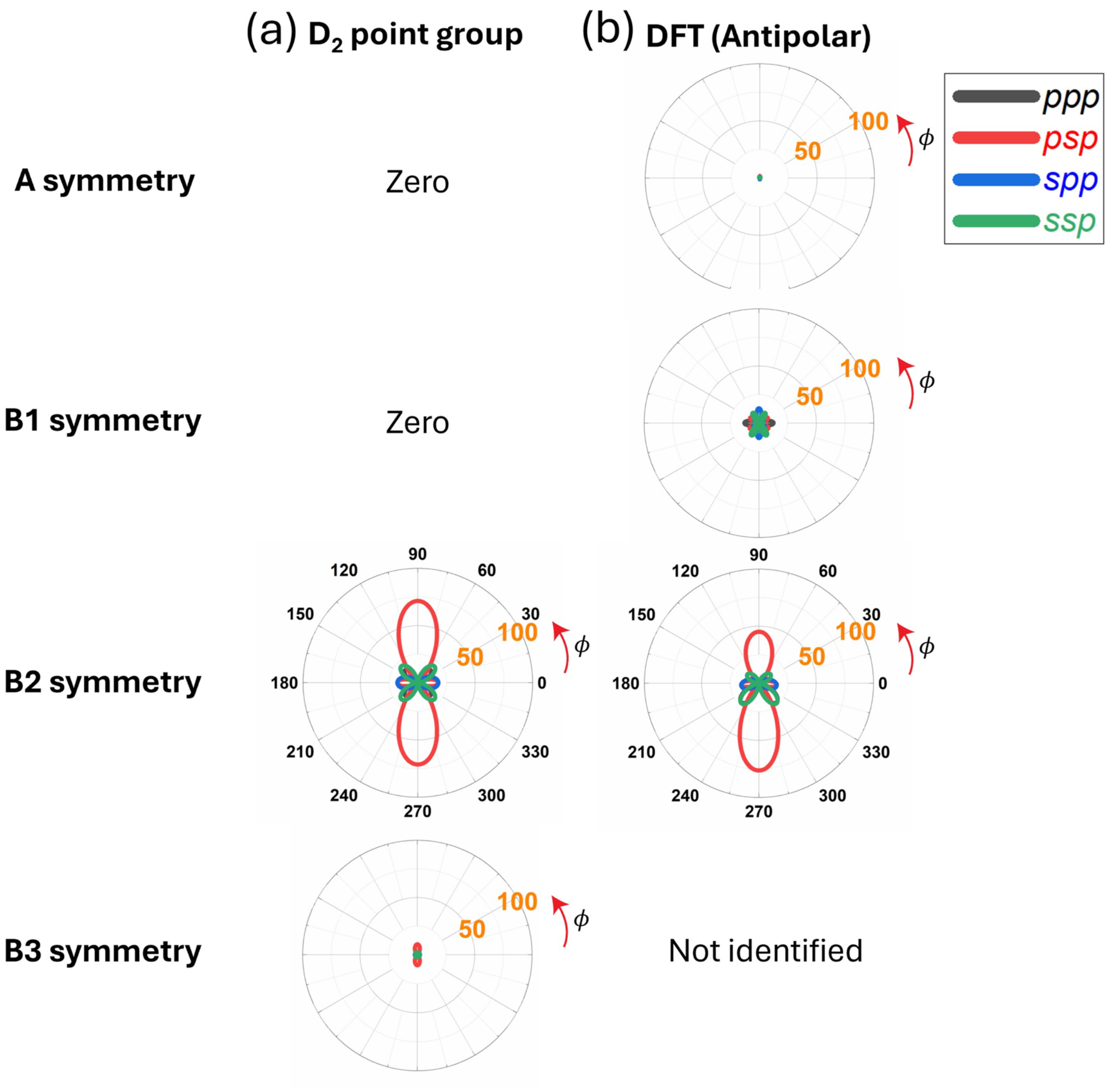

FIG. 4. Polarization and azimuth angle (ϕ) dependence of the SFG signals simulated for (a) amide-I modes with different symmetries within the D2 point group of the antiparallel β-sheet peptide bond and (b) corresponding modes from DFT calculations of a single domain of the antipolar antiparallel GAGA-peptide dimer. The orange numbers on the circles indicate the normalized intensity on an arbitrary scale.

IR spectra (Fig. 3). However, in the SFG spectra (Fig. 2), a small peak can be identified at 1667 cm^{-1}, which could tentatively be assigned to this mode.[45]

The D$_2$ point group analysis assumes that the amide-I mode is highly localized and not coupled with other side group vibrations, but in reality, the vibrational modes are coupled (see movies 1–3 of the supplementary material).[24] For that reason, no elements are zero in the tensors calculated with DFT (Table I). The 1664 cm^{-1} mode, which is closest to the A mode based on the C=O stretch directions, has non-zero values in the dipole derivative tensor and non-zero values in the off-diagonal components in the polarizability derivative tensor. So, no terms in the hyperpolarizability are zero, i.e., all $\beta_{i'j'k'} \neq 0$. However, the calculated $\chi^{(2)}_{eff}$ is nearly zero due to the symmetry cancellation in the laboratory coordinate. So, its SFG intensity is nearly zero [Fig. 4(b)]. Similarly, the 1677 cm^{-1} mode, which is closest to the B1 mode based on the C=O stretch directions, has a weak intensity when the SFG intensity is calculated using the tensors from DFT calculations [Fig. 4(b)].

The 1634 cm^{-1} mode, closest to the B2 mode, exhibits four lobes in the ϕ-polar plot for the *psp* polarization: two strong ones along ϕ = 90° and 270° but asymmetric, and two weak ones along ϕ = 0° and 180° [Fig. 4(b)]. The *spp*, *ppp*, and *ssp* polarization calculation results are similar to the ones calculated for the D$_2$ point group, but asymmetric along the ϕ = 90° and 270° directions. These

TABLE II. Raman polarizability and IR dipole moment tensors for amide-A modes from DFT calculations for the antipolar antiparallel GAGA-peptide dimer. The tensor components are normalized with the large value of the mode at 1634 cm^{-1}. Bold highlights are the tensor elements that have large effects on the SFG intensity.

	$\left(\frac{\partial\alpha_{i'j'}}{\partial Q_q}\right)\left(\frac{\partial\mu_{k'}}{\partial Q_q}\right)_{\mathrm{DFT}}$	Exp cm^{-1}	DFT cm^{-1}
Coupled NH stretching	$\begin{pmatrix} -0.58 & 1.56 & \mathbf{1.82} \\ 1.56 & 3.77 & 0.40 \\ \mathbf{1.82} & 0.40 & 1.64 \end{pmatrix}\begin{pmatrix} -0.11 \\ \mathbf{0.74} \\ 0 \end{pmatrix}$	3275	3306
Isolated NH stretching	$\begin{pmatrix} -0.46 & -0.45 & \mathbf{1.11} \\ -0.45 & -8.66 & \mathbf{-1.16} \\ \mathbf{1.11} & \mathbf{-1.16} & -3.33 \end{pmatrix}\begin{pmatrix} 0.02 \\ \mathbf{0.50} \\ 0.02 \end{pmatrix}$	⋯	3370

asymmetric features are due to the coupling with (or contributions from) the non-zero μ'_a component in the dipole derivative tensor (Fig. S4) and the non-zero α'_{bc} and α'_{cb} components in the polarizability derivative tensor (Fig. S5). The non-zero α'_{ab} and α'_{ba} components and non-zero α'_{aa}, α'_{bb}, and α'_{cc} components have no impact (Fig. S6).

In DFT calculations with the antipolar antiparallel GAGA-peptide dimer model, we could not see the normal mode that is close to the B3 symmetry. This is probably because the structural model used here does not have stacking along the *a* axis. It is known that the dipoles of the B3 symmetry are effectively canceled within the *bc* plane, leaving only a small component perpendicular to the plane.[48] Thus, the B3 mode would be sensitive to van der Waals interactions along the sheet stacking direction.

D. Simulation of SFG intensity for amide-A band of a single domain

The amide-A and B modes are from the Fermi resonance of the NH stretching modes and overtones of the amide-II modes.[46] The amide-B band is weaker than the amide-A band in both Raman and IR spectra (Fig. 3), and it is not discernible in the SFG spectra (Fig. 2). Thus, only the amide-A mode will be discussed here. DFT calculations do not provide overtone and Fermi resonance modes, but their symmetry should be the same (at least, similar) to the fundamental NH stretch mode because the symmetry is conserved in the Fermi resonance process. DFT calculations of the antipolar antiparallel GAGA-peptide dimer model showed many NH stretching modes; the 3306 cm^{-1} and 3370 cm^{-1} modes appear prominent in the IR spectrum (Fig. 3). The 3306 cm^{-1} mode is one of the collective stretches of multiple NH groups, and the 3370 cm^{-1} mode is the stretch of an individual NH group. Table II presents the Raman polarizability and IR dipole moment tensors of these two modes.

The coupled NH stretch mode at 3306 cm^{-1} has large values at the α'_{ac}, α'_{ca}, and μ'_c positions where large values are found in the tensor of the C=O stretching mode at 1634 cm^{-1}. Thus, the polarization and azimuth angle dependences of the SFG signal of the 3306 cm^{-1} mode [Fig. 5(a)] are similar to those of the 1634 cm^{-1} mode [Fig. 4(b)]. The isolated NH stretch mode at 3370 cm^{-1} has large α'_{bc} and α'_{cb} values, in addition to the α'_{ac} and α'_{ca} elements, in the polarizability derivative tensor, so the azimuth angle dependence of the SFG signal of this mode [Fig. 5(b)] is highly asymmetric along the ϕ = 90° and 270° directions [as shown in Fig. S5(b)]. The non-zero values in other elements in the Raman polarizability tensor have negligible effects (Fig. S6).

E. Simulation of SFG intensity for amide-I and A bands of multiple domains

The calculation results shown in Secs. III C and III D are for the single domain. In reality, SFG measures the signals from multiple domains within the effective probe volume. Thus, the relative polarity (which is quantified as DE in this work) as well as orientational distributions with respect to the fiber axis and surface (σ_φ and σ_θ) will affect the SFG intensities. From the literature, it is known that the *c* axis of the antiparallel β-sheet crystal is aligned along the silk fiber axis with a standard deviation of ~27°.[45] This could be modeled as a Gaussian distribution of the azimuth angle of crystallites (with a standard deviation of σ_φ) at a given azimuth angle (ϕ) of the fiber

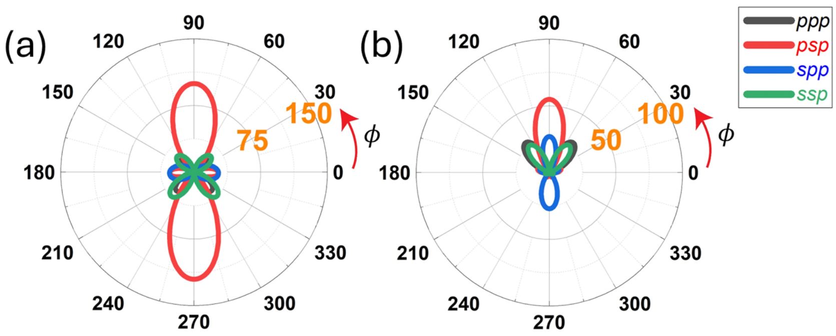

FIG. 5. Polarization and azimuth angle (ϕ) dependence of SFG signals calculated with DFT-based tensors of (a) concerted N–H stretches (coupled with adjacent N–H vibrations) at 3306 cm^{-1} and (b) one of the isolated N–H stretches at 3370 cm^{-1} of a single domain of the antipolar–antiparallel GAGA-peptide dimer. The orange numbers on the circles indicate the normalized intensity on an arbitrary scale for comparison with the 1634 cm^{-1} mode in Fig. 4(b).

with respect to the laboratory coordinate. In addition, the rounded surface of the silk fiber effectively changes the local incidence angle of the probe beam. This could be modeled with a Gaussian distribution with a certain standard deviation of the crystal tilt angle (σ_θ). The effects of DE, σ_φ, and σ_θ were modeled using Eq. (11).

Figure 6 compares the polarization and azimuth angle dependences of the SFG signal of 1634 cm^{-1} mode (amide-I) and 3306 cm^{-1} mode (amide-A) simulated with the tensors from DFT calculations for DE = 0% (statistically bidirectional along the b axis) and 100% (fully unidirectional along the b axis) for selected σ_φ and σ_θ distributions; more cases with smaller increments in σ_φ and σ_θ are shown in Figs. S7 and S8. First, we can see that the SFG intensities of the main modes in the amide-I and A regions are symmetric along the $\phi = 90°$ and 270° directions in the DE = 0% case, while they are slightly asymmetric [as is in the single domain case in Figs. 4(b) and 5(a)] in the DE = 100% case. Even the 3370 cm^{-1} mode, which is highly asymmetric in the single domain calculation [Fig. 5(b)], becomes symmetric along the $\phi = 90°$ and 270° directions when DE = 0% case (Fig. S9). Although the experimental intensities were somewhat fluctuating, we believe that the experimental data (Fig. 2) match better with the DE = 0% case. There is no thermodynamic driving force to enforce the polarity of the antiparallel β-sheets in the silk fiber in one direction when they are crystalized during the natural dehydration process of the fiber synthesis.[22,49]

Second, we can see that the degrees of azimuth angle dependence of all polarization signals decrease as σ_φ and σ_θ increase, which is somewhat expected. Without considering these angular distributions, the experimental SFG data collected at $\phi = 45°$ could not be explained (see Sec. III C). By comparing the relative intensities of the *psp*, *spp*, *ssp*, and *ppp* SFG intensities at three ϕ angles with the theoretically calculated intensities using the D_2 point group [Fig. S7(b)], we have calculated root-mean-square deviation (RMSD) between the experiment and SFG simulation results obtained at three azimuth angles. The details of RMSD calculations are presented in the supplementary material,[45] and the RMSD calculation results for each azimuth angle ($\phi = 0°, 45°, 90°$) and all three angles together are shown in Fig. S10.

When the experimentally observed polarization dependence of the amide-I band is compared with the calculation results obtained for the B2 symmetry of the D_2 point group, the most likely σ_φ and σ_θ are estimated to be ~27° and ~5°, respectively [Fig. S10(d)].[45] When the experimental data of the amide-I mode are compared with the theoretical trends calculated with the tensors from DFT calculations, the angular distributions are found to be slightly larger: $\sigma_\varphi \approx 35°$ and

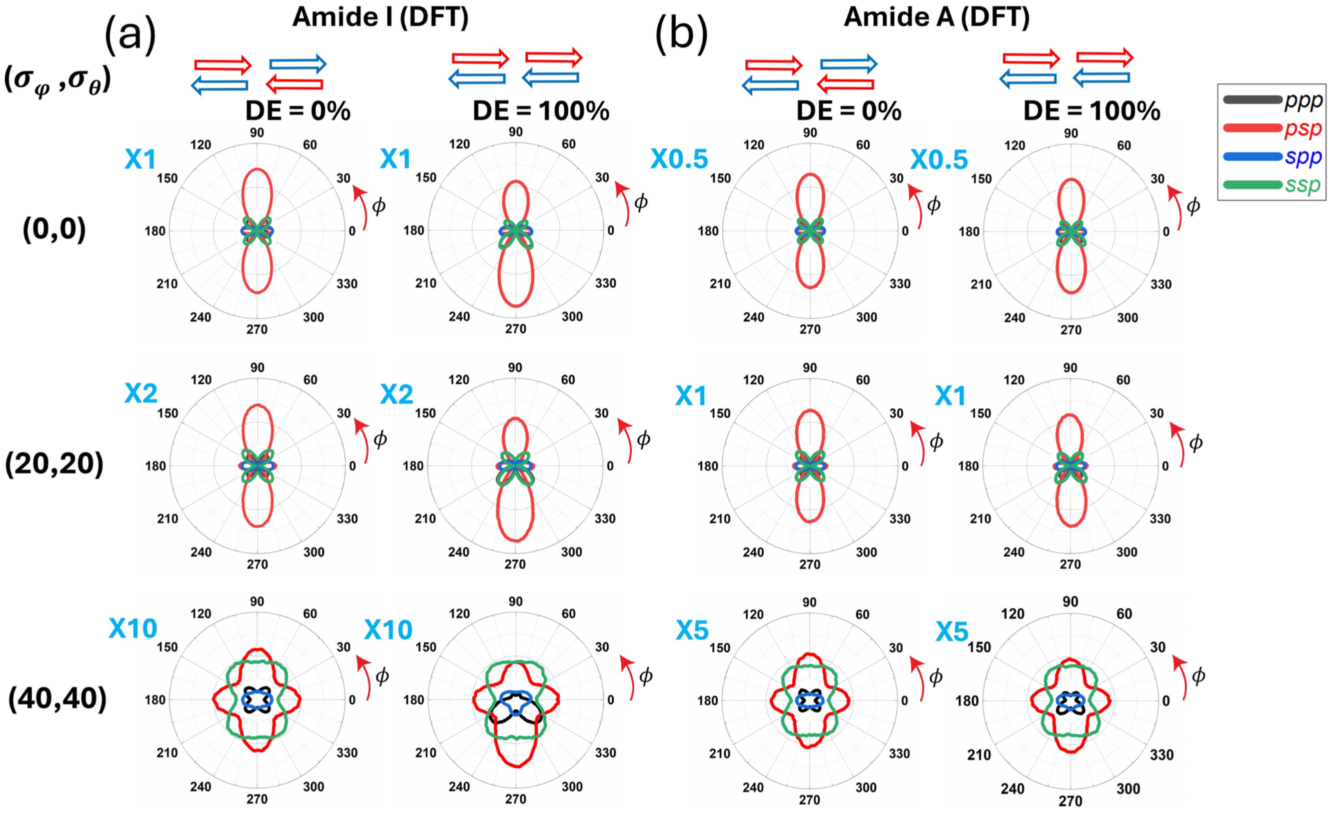

FIG. 6. Polarization and azimuth angle (ϕ) dependence of SFG signals, calculated using the tensors from DFT calculations, of (a) amide-I mode at 1634 cm^{-1} and (b) amide-A mode at 3306 cm^{-1} for bidirectional (DE = 0%) and unidirectional (DE = 100%) arrangements with the standard deviations of azimuth (σ_φ) and tilt angle (σ_θ) distributions marked in the left. More incremental charts are shown in Fig. S7. The intensities of ϕ-polar plots are normalized relative to the amide-I mode of the case with DE = 0% and $\sigma_\varphi = \sigma_\theta = 0°$.

TABLE III. Raman polarizability and IR dipole moment tensors for the dominant modes in the amide-II and III modes from DFT calculations of the antipolar antiparallel GAGA-peptide dimer. The tensor components are normalized with the largest values of the 1634 cm^{-1} mode.

	$\left(\frac{\partial \alpha_{i'j'}}{\partial Q_q}\right)\left(\frac{\partial \mu_{k'}}{\partial Q_q}\right)_{\mathrm{DFT}}$	Exp cm^{-1}	DFT cm^{-1}
Amide II mode	$\begin{pmatrix} -0.26 & 0.25 & 0.04 \\ 0.25 & 0.25 & 0.06 \\ 0.04 & 0.06 & 0.92 \end{pmatrix}\begin{pmatrix} -0.03 \\ -0.18 \\ -0.76 \end{pmatrix}$	1567	1475
Amide III mode	$\begin{pmatrix} 0.31 & 0.41 & 1.66 \\ 0.41 & -0.70 & -1.49 \\ 1.66 & -1.49 & -0.82 \end{pmatrix}\begin{pmatrix} 0.08 \\ -0.01 \\ 0.17 \end{pmatrix}$	1265	1207

$\sigma_\theta \approx 25°$ [Fig. S10(h)]. The results obtained with the amide-A mode show $\sigma_\varphi \approx 35°$ and $\sigma_\theta \approx 10°$ [Fig. S10(l)]. The self-consistency among the angular distribution analysis results for amide-I and amide-A suggests that the DFT-calculated tensors are quite reasonable for these modes. The reason that the calculation results are based on the D_2 point group might be due to the ignorance of vibrational couplings among neighboring groups, which causes deviation of these vibrational modes from the perfect B2 symmetry.

F. Simulation of SFG intensity for amide-II and III bands

Table III displays the Raman polarizability and IR dipole moment tensors for the dominant modes in the amide-II and III band regions for the antipolar antiparallel GAGA-peptide dimer. The amide-II modes primarily involve NH in-plane bending and CN stretching, with some contributions from H_α bending.[46,48] So, they have a relatively large μ'_c component and an intermediate μ'_b component in the IR tensor. They also have a relatively large α'_{cc} in the Raman tensor. The polarization and azimuth dependence of all modes in the amide-II region calculated for the single domain shows a large signal in the *ppp* polarization along the $\phi = 0°$ and 180° directions and weak signals in other polarizations (Fig. S11). Even if the multiple domains with DE = 0% are considered, the general trend does not change (Fig. S12). This trend does not match with the experimentally observed pattern in Fig. 2. This discrepancy must mean that the dimer model used in this study [Figs. 1(a)–1(c)] may not be big enough to simulate the amide-II modes (as in the case of the B3 symmetry mode in the amide-I region).

The amide-III modes encompass various vibrational displacements, including NH in-plane bending, CN stretching, H_α bending, and vibrations of other side groups at the C_α position.[46] The SFG intensities calculated for the single domain using the tensors shown in Table III show highly asymmetric patterns along the $\phi = 90°$ and 270° directions, and the *psp* peak is smaller than the *ssp* peak except at $\phi = 270°$ [Fig. 7(a)]. In the case of the multiple domains with DE = 0%, the *psp* peak becomes dominant and symmetric along the $\phi = 90°$ and 270° directions. Figure 7(b) shows the DE = 0% case with $\sigma_\varphi \approx 30°$ and $\sigma_\theta \approx 20°$. The overall trend is in qualitative agreement with the trend shown in Fig. 2. However, the calculated intensity ratio of the amide-III vs amide-I (I_{1207}/I_{1634}) is smaller than 1, while the experimental data show the opposite trend. Again, we speculate that this might be due to the limited size of the structural model used in this study.

G. Simulation of SFG intensity for CH deformation region

In the spectral region between amide-II and III bands, various types of deformations of the side groups at the C_α position (two hydrogens for glycine and the hydrogen and methyl group for alanine) are involved.[46] The tensor values for the 1357 cm^{-1} mode obtained from DFT calculations, which involve symmetric CH_3 bending coupled with CH_2 wagging, are shown in Table IV. Based on the relative magnitudes of components in the IR tensor, its dipole moment is pointing in the diagonal direction in the *ab* plane and slightly tilted along the *c* axis. The relative magnitudes of all components in the Raman tensor are comparable to each other (that is, no single component is excessively dominant).

The SFG intensities calculated using these tensor values are shown in Fig. 8. In the case of a single domain, the *spp* signal is symmetric along the $\phi = 90°$ and 270° directions, and all other polarizations are asymmetric in the ϕ-plot [Fig. 8(a)]. However, in the case of multiple domains with DE = 0%, the azimuth angle dependence

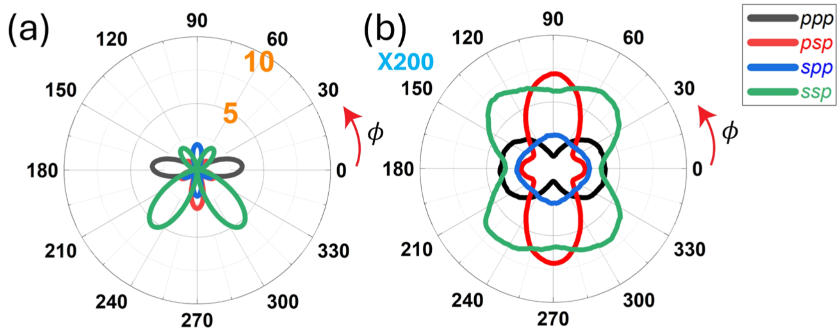

FIG. 7. Polarization and azimuth angle (ϕ) dependence of SFG signals calculated with the tensors obtained from DFT calculations for one of the amide-III modes at 1207 cm^{-1} for (a) a single domain and (b) multiple domains with DE = 0%, $\sigma_\varphi \approx 30°$ and $\sigma_\theta \approx 20°$. The orange numbers in (a) indicate the normalized intensity on an arbitrary scale for comparison with the 1634 cm^{-1} mode in Fig. 4(b). The multiplication factor in (b) shows the relative scale with respect to the 1634 cm^{-1} intensity for the DE = 0% and $\sigma_\varphi = \sigma_\theta = 0°$ case in Fig. 6(a).

TABLE IV. Raman polarizability and IR dipole moment tensors of the 1357 cm^{-1} mode in the CH deformation region calculated for the antipolar antiparallel GAGA-peptide dimer. The tensor components are normalized with the largest values of the 1634 cm^{-1} mode.

	$\left(\frac{\partial\alpha_{i'j'}}{\partial Q_q}\right)\left(\frac{\partial\mu_{k'}}{\partial Q_q}\right)_{\text{DFT}}$	Exp cm^{-1}	DFT cm^{-1}
CH$_3$ symmetric bending at alanine coupled with CH$_2$ wagging at glycine	$\begin{pmatrix} -0.27 & 0.13 & -0.86 \\ 0.13 & -0.44 & 0.36 \\ -0.86 & 0.36 & -0.28 \end{pmatrix}\begin{pmatrix} 0.13 \\ -0.15 \\ 0.08 \end{pmatrix}$	1370	1357

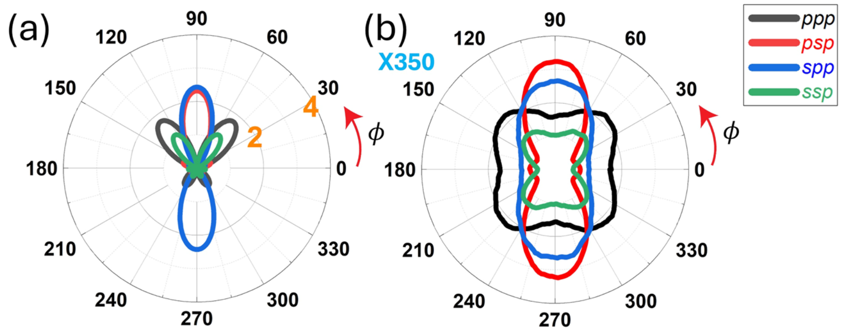

FIG. 8. Polarization and azimuth angle (ϕ) dependence of one of the CH$_3$ symmetric bending modes, coupled with the CH$_2$ wagging mode, at 1357 cm^{-1} for (a) a single domain and (b) multiple domains with DE = 0% and angular distributions ($\sigma_\varphi = 30°$ and $\sigma_\theta = 20°$). More incremental charts are shown in Fig. S13. The orange numbers in (a) indicate the normalized intensity on an arbitrary scale for comparison with the 1634 cm^{-1} mode in Fig. 4(b). The multiplication factor in (b) shows the relative scale with respect to the 1634 cm^{-1} intensity for the DE = 0% and $\sigma_\varphi = \sigma_\theta = 0°$ case in Fig. 6(a).

becomes symmetric and the *psp* signal becomes dominant in the $\phi = 90°$ and 270° directions (Fig. S13). Although this is qualitatively consistent with the trend observed in the experiment (Fig. 2), its intensity simulated from the DFT-based tensors is significantly smaller than the 1634 cm^{-1} mode as compared to the experimental data.

H. Comparison of antipolar vs polar conformations of antiparallel GAGA-peptide dimer

DFT calculations were also carried out for the polar antiparallel GAGA-peptide dimer [Figs. 1(a)–1(c)], which represents the structural model proposed by March *et al.* in 1955.[39] The DFT-based tensors and simulation results obtained for the polar conformation are shown in the supplementary material (Tables S1–S2 and Figs. S14–S16). The overall trends were similar to the ones observed for the antipolar antiparallel GAGA-peptide dimer case [Figs. 1(d)–1(f)]. The main difference between these two models was the degree of asymmetry along the $\phi = 90°$ and 270° directions in the ϕ-polar plot for the single domain and the multiple domains with DE = 100% cases. The comparison of the trends described for the antipolar conformation in the previous sections and those for the polar conformation in the supplementary material suggests that SFG would not be able to differentiate the antipolar vs polar conformations of the antiparallel β-sheet if DE = 0%. If DE = 100%, the azimuth angle dependent measurement of the *psp*-SFG signal of the amide-I and A modes would be able to distinguish the antipolar vs polar conformations. However, such polar alignments are statistically unlikely when the β-sheet crystallization naturally occurs during the dehydration process of the silk fiber.[22,49]

IV. CONCLUSION

This study demonstrates how DFT calculations can be used to predict the polarization and azimuth angle dependence of SFG signals of antiparallel β-sheet domains in *B. mori* silk fiber. The β-sheet was modeled with a dimer of two GAGA peptides paired in the antiparallel direction, and the terminal groups were replaced with iodine to avoid artifacts due to loose chain ends. The Raman polarizability and IR dipole moment tensors obtained from DFG calculations of this small structural model could explain the polarization and azimuth angle dependence of the amide-I and A bands of antiparallel β-sheet. The DFT calculation results for amide-II did not match with experimental data, which suggested vibrational coupling within a larger crystalline domain may dominate the SFG spectral features of these modes. The calculation results for amide-III and the CH deformation modes were in qualitative agreement with experimental data, although the predicted intensities were much smaller. The methodology demonstrated here can be applied to the structural analysis of other biopolymers.

SUPPLEMENTARY MATERIAL

The supplementary material contains the following: I. Effects of various non-zero tensor components on polarization and azimuth angle dependence of B2 and B3 symmetry modes of the D2 point group, II. Effects of chain alignment and tilt angle distribution of the amide-I mode close to B2 symmetry of the D2 point group, III. Polar plots of all amide-II modes predicted from DFT calculations of the antipolar antiparallel GAGA-peptide dimer model, IV. Polar plots of a CH deformation mode predicted from DFT calculations of the antipolar antiparallel GAGA-peptide dimer model, and V. Comparison of the DFT calculation results and predicted SFG intensities for the antipolar vs polar conformations of the antiparallel GAGA-peptide dimer.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (Grant No. NSF-2203635). S. Chen and X. Chen were supported by the Natural Science Foundation of Tianjin (Grant No. 23JCYBJC01660) for DFT calculations.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Jihyeong Ryu: Conceptualization (lead); Data curation (lead); Formal analysis (lead); Investigation (lead); Methodology (lead); Software (equal); Validation (lead); Visualization (lead); Writing – original draft (lead). **Sibing Chen**: Data curation (lead). **Juseok Choi**: Software (lead); Validation (supporting). **Xing Chen**: Data curation (equal); Supervision (equal). **Seong H. Kim**: Funding acquisition (lead); Project administration (lead); Supervision (lead); Validation (lead); Writing – review & editing (lead).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] Y.-R. Shen, *Principles of Nonlinear Optics* (Wiley-Interscience, New York, 1984).

[2] A. G. Lambert, P. B. Davies, and D. J. Neivandt, Appl. Spectrosc. Rev. **40**(2), 103–145 (2005).

[3] Y. Shen, J. Phys. Chem. C **116**(29), 15505–15509 (2012).

[4] A. L. Barnette, L. C. Bradley, B. D. Veres, E. P. Schreiner, Y. B. Park, J. Park, S. Park, and S. H. Kim, Biomacromolecules **12**(7), 2434–2439 (2011).

[5] Y. Ogawa, C. M. Lee, Y. Nishiyama, and S. H. Kim, Macromolecules **49**(18), 7025–7031 (2016).

[6] L. Kong, C. Lee, S. H. Kim, and G. R. Ziegler, J. Phys. Chem. B **118**(7), 1775–1783 (2014).

[7] Y. Zhang, J. Yu, X. Wang, D. M. Durachko, S. Zhang, and D. J. Cosgrove, Science **372**(6543), 706–711 (2021).

[8] L.-D. Koh, Y. Cheng, C.-P. Teng, Y.-W. Khin, X.-J. Loh, S.-Y. Tee, M. Low, E. Ye, H.-D. Yu, Y.-W. Zhang, and M. Y. Han, Prog. Polym. Sci. **46**, 86–110 (2015).

[9] P. Fratzl, *Collagen: Structure and Mechanics* (Springer, 2008), pp. 1–13.

[10] I. Chae, C. K. Jeong, Z. Ounaies, and S. H. Kim, ACS Appl. Bio Mater. **1**(4), 936–953 (2018).

[11] A. L. Barnette, C. Lee, L. C. Bradley, E. P. Schreiner, Y. B. Park, H. Shin, D. J. Cosgrove, S. Park, and S. H. Kim, Carbohydr. Polym. **89**(3), 802–809 (2012).

[12] S. H. Kim, C. M. Lee, and K. Kafle, Korean J. Chem. Eng. **30**, 2127–2141 (2013).

[13] T. Lefèvre, M.-E. Rousseau, and M. Pézolet, Biophys. J. **92**(8), 2885–2895 (2007).

[14] X. Hu, D. Kaplan, and P. Cebe, Macromolecules **39**(18), 6161–6170 (2006).

[15] D. Badillo-Sanchez, D. Chelazzi, R. Giorgi, A. Cincinelli, and P. Baglioni, Polym. Degrad. Stab. **157**, 53–62 (2018).

[16] C. M. Lee, A. Mittal, A. L. Barnette, K. Kafle, Y. B. Park, H. Shin, D. K. Johnson, S. Park, and S. H. Kim, Cellulose **20**, 991–1000 (2013).

[17] M. Makarem, Y. Nishiyama, X. Xin, D. M. Durachko, Y. Gu, D. J. Cosgrove, and S. H. Kim, J. Phys. Chem. B **124**(37), 8071–8081 (2020).

[18] J. Choi, M. Makarem, C. Lee, J. Lee, S. Kiemle, D. J. Cosgrove, and S. H. Kim, Sci. Rep. **13**(1), 22007 (2023).

[19] H.-F. Wang, W. Gan, R. Lu, Y. Rao, and B.-H. Wu, Int. Rev. Phys. Chem. **24**(2), 191–256 (2005).

[20] J. Choi, J. Lee, M. Makarem, S. Huang, and S. H. Kim, J. Phys. Chem. B **126**(35), 6629–6641 (2022).

[21] C. M. Lee, K. Kafle, Y. B. Park, and S. H. Kim, Phys. Chem. Chem. Phys. **16**(22), 10844–10853 (2014).

[22] W. Qiu, A. Patil, F. Hu, and X. Y. Liu, Small **15**(51), 1903948 (2019).

[23] S. Li, L. Bashline, Y. Zheng, X. Xin, S. Huang, Z. Kong, S. H. Kim, D. J. Cosgrove, and Y. Gu, Proc. Natl. Acad. Sci. U. S. A. **113**(40), 11348–11353 (2016).

[24] J. Choi, A. L. Kwansa, Y. G. Yingling, and S. H. Kim, J. Phys. Chem. B **127**(39), 8456–8467 (2023).

[25] K. T. Nguyen, J. T. King, and Z. Chen, J. Phys. Chem. B **114**(25), 8291–8300 (2010).

[26] S. Ye, F. Wei, H. Li, K. Tian, and Y. Luo, Adv. Protein Chem. Struct. Biol. **93**, 213–255 (2013).

[27] H. M. Chase, B. Rudshteyn, B. T. Psciuk, M. A. Upshur, B. F. Strick, R. J. Thomson, V. S. Batista, and F. M. Geiger, J. Phys. Chem. B **120**(8), 1919–1927 (2016).

[28] D. Xiao, L. Fu, J. Liu, V. S. Batista, and E. C. Yan, J. Mol. Biol. **421**(4-5), 537–547 (2012).

[29] J. M. Gosline, P. Guerette, C. Ortlepp, and K. Savage, J. Exp. Biol. **202**(23), 3295–3303 (1999).

[30] O. Tokareva, M. Jacobsen, M. Buehler, J. Wong, and D. L. Kaplan, Acta Biomater. **10**(4), 1612–1626 (2014).

[31] D. López Barreiro, J. Yeo, A. Tarakanova, F. J. Martin-Martinez, and M. J. Buehler, Macromol. Biosci. **19**(3), 1800253 (2019).

[32] Y. Takahashi, M. Gehoh, and K. Yuzuriha, Int. J. Biol. Macromol. **24**(2-3), 127–138 (1999).

[33] T. Asakura, T. Ohata, S. Kametani, K. Okushita, K. Yazawa, Y. Nishiyama, K. Nishimura, A. Aoki, F. Suzuki, H. Kaji *et al.*, Macromolecules **48**(1), 28–36 (2015).

[34] M. McGill, G. P. Holland, and D. L. Kaplan, Macromol. Rapid Commun. **40**(1), 1800390 (2019).

[35] L. F. Drummy, B. Farmer, and R. R. Naik, Soft Matter **3**(7), 877–882 (2007).

[36] C.-Z. Zhou, F. Confalonieri, N. Medina, Y. Zivanovic, C. Esnault, T. Yang, M. Jacquet, J. Janin, M. Duguet, and R. Perasso, Nucleic Acids Res. **28**(12), 2413–2419 (2000).

[37] T. Asakura, M. Okonogi, K. Horiguchi, A. Aoki, H. Saitô, D. P. Knight, and M. P. Williamson, Angew. Chem. **124**(5), 1238–1241 (2012).

[38] C. M. Lee, K. Kafle, S. Huang, and S. H. Kim, J. Phys. Chem. B **120**(1), 102–116 (2016).

[39] R. E. Marsh, R. B. Corey, and L. Pauling, Biochim. Biophys. Acta **16**, 1–34 (1955).

[40] A. Edoardo, E. Bylaska, W. D. Jong, N. Govind, K. Kowalski, T. Straatsma, M. Valiev, H. van Dam, Y. Alexeev, and J. Anchell, J. Chem. Phys. **152**(18), 182102 (2020).

[41] E. Smith and G. Dent, *Modern Raman Spectroscopy: A Practical Approach* (John Wiley & Sons, 2019).

[42] R. Superfine, J. Y. Huang, and Y. Shen, Chem. Phys. Lett. **172**(3–4), 303–306 (1990).

[43] M. Baudrier-Raybaut, R. Haidar, P. Kupecek, P. Lemasson, and E. Rosencher, Nature **432**(7015), 374–376 (2004).

[44] G. Perotto, Y. Zhang, D. Naskar, N. Patel, D. L. Kaplan, S. C. Kundu, and F. G. Omenetto, Appl. Phys. Lett. **111**(10) (2017).
[45] J. Ryu, J. Choi, J. Lee, and S. H. Kim, Biomacromolecules **25**, 7178 (2024).
[46] W. H. Moore and S. Krimm, Biopolymers **15**(12), 2465–2483 (1976).
[47] T. Miyazawa and E. Blout, J. Am. Chem. Soc. **83**(3), 712–719 (1961).
[48] A. Barth and C. Zscherp, Q. Rev. Biophys. **35**(4), 369–430 (2002).
[49] T. Asakura, K. Umemura, Y. Nakazawa, H. Hirose, J. Higham, and D. Knight, Biomacromolecules **8**(1), 175–181 (2007).