Label-Free Identification of Tumor Tissues by Coherent Nonlinear Vibrational Mode Imaging

Bin Yang[1], Jianyu Ren[1], Chun-Chieh Yu[1], Zixuan Wang[2], Minghui Cao,[3] Shizhen Emily Wang[3,4], Wei Xiong*[1,5,6]

[1]Department of Chemistry and Biochemistry, UC San Diego, La Jolla, CA, 92093, USA

[2]Department of Anesthesiology, School of Medicine, University of California, San Diego, La Jolla CA, 92093-0629, USA

[3]Department of Pathology, University of California, San Diego, La Jolla, CA, 92093, USA

[4]Moores Cancer Center, University of California, San Diego, La Jolla, CA, 92093, USA

[5]Materials Science and Engineering Program, UC San Diego, La Jolla, CA, 92093, USA

[6]Department of Electrical and Computer Engineering, UC San Diego, La Jolla, CA, 92093, USA

*Wei Xiong.

Email: w2xiong@ucsd.edu

Author Contributions: W.X. conceived the idea and initiated the project design after discussion with S.E.W. W.X. supervised the project. B.Y. made the experiment feasible and performed the spatially resolved VSFG spectra and VSFG imaging experiments. J.R. conducted the AFM measurements, and carried out the DFT calculations and VSFG theory simulations with support from C-C.Y. S.E.W. provided the mouse lung tissue samples and contributed to the discussions. M. Cao prepared the mouse tissue samples. Z.W. performed the mouse tissue cutting. B.Y. analyzed the experimental results, J.R. analyzed the theoretical results with discussion from C-C.Y. The manuscript was written by B.Y., J.R., C-C.Y. and W.X., with input from all authors.

Competing Interest Statement: The authors declare no competing interests.

Classification: Physical chemistry.

Keywords: Coherent 2nd order nonlinear optics, hyperspectral imaging, sum frequency generation, collagen, tumor identification.

This PDF file includes:

Main Text
Figures 1 to 6
Tables 1

Abstract

We present vibrational sum-frequency generation (VSFG) microscopy as a new label-free chemical imaging technique for tumor identification. This method combines the chemical-bond selectivity of vibrational spectroscopy with coherent interference of second-order coherent nonlinear optics. Using a fast line-scanning VSFG microscope, we obtained hyperspectral VSFG images of collagen I from both lung tissues bearing metastatic tumors and in tumor-free ones, which reveal drastic different spectral signatures: tumor samples exhibit large NH stretch (NH_S) and CH stretch (CH_S) versus the CH_2 symmetric stretch ($CH_{2,Ss}$), compared to healthy controls. We then identified two

key spectral signatures to distinguish metastatic tumor and tumor-free tissues: the intensity ratio of $NH_S/CH_{2,Ss}$ and $CH_S/CH_{2,Ss}$ modes. These signatures demonstrated high fidelity in distinguishing between tumors and normal tissue in both mouse and human lung samples. Theoretical modeling indicates that distinctive interferences of spectral peaks are sensitive to interfibrillar distances, at 130 of nanometer level. These findings suggest that collagen fibrils are more densely packed in tumors, corroborating the enhanced stiffness observed in tumor tissues. VSFG microscopy offers a highly selective, label-free chemical imaging characterization that preserves sample integrity, making it a valuable tool for oncology, pathology and fundamental biophysical research.

Significance Statement

Label-free identification of tumor tissues is essential for non-perturbative characterization, as it preserves sample integrity for further analysis. However, achieving this is challenging. In this study, we introduce vibrational sum-frequency generation (VSFG) microscopy as a novel, label-free chemical imaging technique for tumor identification, utilizing the coherent interference of nonlinear signals and the chemical specificity of vibrational modes. VSFG spectral signatures reveal distinct differences between malignant and non-malignant samples, offering reliable biomarkers for tumors. Benchmarking these spectral features demonstrates the sensitivity of this technique to tumors in both murine and human lung samples. Therefore, VSFG microscopy provides a highly selective, label-free imaging approach that complements existing modalities, making it a valuable tool for oncology, pathology, and biophysical research.

Main Text

Introduction

Biomedical imaging tests and biopsy imaging constitute a major pillar of clinical management for many human diseases, including cancer. They play a central role in diagnosis, staging, as well as in predicting and evaluating therapeutic response[1-3]. Traditional histopathology relies on tissue staining with dyes or probes (including antibodies) that label specific tissue, cellular, and molecular components.[4] Recently, advanced multimodal imaging approaches have been developed to achieve sensitive, multiplexed and high-performance biomarker detection, as a new promise for improved precision and accuracy of tumor detection and a better cancer care.[5-7] Despite their wide clinical and research applications, methods involving colorimetric or fluorescent probes to label and image tissue components often requires time-consuming tissue fixation and staining processes. These methods are thus limited in their multiplexing capabilities, and the detection depends on target (e.g. proteins nucleic acids and other biomolecules) abundance as well as the availability and specificity of probes[8]. In contrast, label-free imaging offers high chemical sensitivity and maintains sample integrity by directly probing chemical or structural changes in biomarkers, such as proteins, DNA/RNA and lipids in tumor tissues.[9-15] It also simplifies sample processing and is not constrained by the development of new detection probes. These advantages make label-free imaging a robust standalone detection method and a highly compatible module for integration into high-plex models of digital pathology.

Collagens, the most abundant proteins in the tumor microenvironment (TME), have been proposed as biomarkers for several types of cancer[16-22]. As major structural components of the extracellular matrix (ECM),[16] collagens are composed of a triple helix of polypeptide chains that further pack into fibrils and assemble into fibers (Figure S7). A higher collagen content is associated with increased mammographic density, which predicts a higher risk of breast cancer in women[23]. Pathological conditions such as fibrosis lead to collagen accumulation, resulting in variations in fibril density, packing configurations, and changes in fiber orientation and crimp, all of which can exert profound effects on tumor progression (Figure 1a). Different types of collagens in the stroma can either restrict or promote tumor growth[24]. Tumor formation and progression often result in alterations in collagen type, density, and architecture, which in turn affects a variety of cellular behaviors within the tumor milieu, thereby promoting tumor progression [25]. Studies using Atomic Force Microscopy

(AFM) have shown that the mechanical properties of collagens, i.e. stiffness, are significantly higher in the ECM of tumors compared to healthy tissues.[26,27] Considering these reported changes in mechanical properties, it is logical to infer that the associated structural changes at microscopic or mesoscopic levels can be detected by molecular spectroscopy – a primary method for monitoring molecular transformations. More importantly, molecular spectroscopy can be developed into hyperspectral imaging for label-free tumor detection. \

Currently, the label-free identification of tumor-related collagens using hyperspectral imaging is still in its early stages. Second harmonic generation (SHG) microscopy is a powerful modality for imaging fibrillar collagen in a diverse range of tissues[28-30]. SHG is a 2nd order nonlinear optical process generated by non-centrosymmetric systems, making it highly sensitive to collagen fibril and fiber structure.[31] This sensitivity arises from the coherent emission of the SHG signal, which interferes based on the distance, relative orientation and alignments of the SHG emitters, such as collagen fibrils and fibers. In contrast, many other biological species, featuring centrosymmetry, do not generate an SHG signal, rendering SHG highly sensitive for detecting collagens. However, it is difficult to infer collagen structures solely from the overall SHG intensity and to distinguish between collagens in healthy and tumor tissues, because SHG does not provide bond-specific information. On the other hand, various Raman microscopy techniques, such as stimulated Raman scattering and coherent anti-Stokes Raman scattering[9-12,32], as well as IR imaging[33-35], have yet to be directly applied for collagen-based imaging to differentiate tumor domains, despite their success in pre-clinical animal models. One reason is that, although these techniques offer molecular sensitivity by revealing chemical composition changes in biomarker molecules through spectral peak shifts[9,10], not all tumor growth leads to substantial chemical changes detectable by these shifts[16-19,36]. Furthermore, even when such changes occur, IR and Raman detect all molecules in the sample path, making the detection of these changes subtle due to significant spectral background from other intact species. More importantly, mesoscopic structure changes in collagen fibers cannot be detected by these scattering-based techniques due to their lack of coherent interference effects.

Herein, by combining vibrational spectroscopy and second order nonlinear optics, we demonstrate vibrational sum-frequency generation (VSFG) microscopy as a label-free imaging technique for tumor identification, leveraging its spectral sensitivity to collagen I fibril packing. VSFG uses two incident optical beams, one IR beam at ω_{IR} and another up-conversion beam at ω_{VIS}, generating a signal at the sum frequency $\omega_{SFG} = \omega_{IR} + \omega_{VIS}$. Because the IR beam resonantly excites molecular vibrational modes (Figure 1b), it is sensitive to specific molecular modes, similar to IR spectroscopy. The excited vibrational coherences then interact with the near IR beam nonresonantly to upconvert it to a virtual state that subsequently emit coherent visible VSFG signals. Thus, like SHG, VSFG is a 2nd order coherent nonlinear optical process that is sensitive to non-centrosymmetric structures due to symmetry and coherent interferences[37-45]. However, different from SHG, VSFG signals are vibrational modes specific, which gives it sensitivity to chemical bonds. For example, within the same collagen fibers, vibrational modes originating from different chemical bonds interfere distinctively within the coherent length, ca. ~100s of nm, depending on their relative orientation and arrangements in the collagen framework, similar to the pioneering VSFG imaging studies of cellulose fibers by Kim and coworkers.[46-48] We demonstrated that VSFG's unique combination of structural selectivity and chemical-bond sensitivity makes its spectra to exhibit sensitivity towards fibrils packing. As a result, VSFG can differentiate collagen in complex biological environments without signal contributions from other biological species.

Hyperspectral images of collagens were collected by a fast line-scanning VSFG microscopy with 1.6 micron spatial resolution[49,50] Within each pixel, a broadband VSFG spectrum encoding collagen structural information was recorded. We identified two key spectral signatures – the intensity ratio of $NH_S/CH_{2,Ss}$ and $CH_S/CH_{2,Ss}$ modes – to discern lung tissues bearing metastatic tumors and tumor-free lung tissues. These signatures were implemented to differentiate tumor from normal tissue in both mouse and human lung samples. Based on DFT calculations and VSFG theory, we explained this contrast by demonstrating that different spectral peaks depend on interfibrillar

spacings uniquely. The spectral observation suggested denser collagen packing in tumors, aligning with the observed enhanced stiffness in tumor tissues. Thus, building on the previous fundamental studies of collagen VSFG spectra and imaging [51-57], this work for the first time demonstrates the power of VSFG to identify collagen fibers in tumor tissues, offering a new diagnostic tool for oncology. Furthermore, the structural knowledge at the molecular scale gained by hyperspectral imaging offers valuable insights for fundamental biophysics research and pathology.

Results

Mechanical and IR imaging studies of mouse lung tissues

We first compared the stiffness of normal, tumor-free mouse lung tissues with those bearing metastatic tumors. The metastatic lung tumors were developed following tail-vein injection of MDA-MB-231 human breast cancer cells into female NOD-scid IL2Rgamma^{null} mice[58]. Using AFM, we accessed the mechanical properties of the collagen fibers. While tumor and control collagen fibers exhibited similar topographic maps (Figure 2a), force maps revealed differences in the elastic properties of the collagen-rich ECM between tumor tissues and control tissues. The elastic modulus of collagen in the control sample was measured to be ~200 MPa (Figure 2b), consistent with previous reported value[59]. In contrast, the elastic modulus for collagen in tumor tissues was significantly higher, at 1500 MPa. This substantial difference indicates that collagen-rich sites in tumors were up to 7.5-fold stiffer than in normal tissues.

The changes in stiffness could originate from microscopic structural changes, such as modifications in amino acid or peptide chemical structures, or mesoscopic alternations in fibril or fiber packing and alignment. To examine the former, we employed optical photothermal infrared (OPTIR) microscopy. OPTIR microscopy measures infrared absorption spectra at each site through photothermal expansion and scans the samples to attain hyperspectral images[60,61]. For both healthy and tumor tissues, their spectral profiles were remarkably similar, characterized by pronounced CH vibrational modes (2800-3000 cm^{-1}) and a broadband peak spanning the NH (3200-3500 cm^{-1}) regions (Figure 2c). Consequently, the intensity map of CH and NH vibrational modes for both tumor and control samples resembled each other (Figure 2d). Furthermore, all pixels in the image showed an IR signal even if weak (e.g., spot 5 in Figure 2c). This is because all molecules (e.g. proteins, lipids.) in the tissue contributed to IR absorption as long as they possess the corresponding chemical bonds. This uniformity in the spectral characteristics and similar spatial distribution of CH and NH modes suggested no dramatic chemical differences between healthy and tumor tissues, such as bond formation or breaking. These results also confirm that IR imaging alone lacks the capability to differentiate between tumor and healthy tissues based on chemical composition.

VSFG spectra analysis and tumor identification based on hyperspectral imaging of mouse lung tissues

Because OPTIR results showed dominant vibrational signals of the CH and NH modes, we measured spatially resolved VSFG spectra in these spectral regions. The VSFG spectra of control and tumor samples displayed substantial differences (Figure 3a,b). In the healthy sample, the VSFG spectra showed two major peaks near 2970 and 3050 cm^{-1} respectively, overlaid on a broad feature. In contrast, the tumor tissues exhibited a large peak near 3350 cm^{-1}. For the control sample, the dominating peaks located at 2970 cm^{-1}, 3050 cm^{-1} and 3310 cm^{-1} were assigned to CH_2 symmetric stretch ($CH_{2,\ Ss}$), CH_2 asymmetric stretch ($CH_{2,As}$) and amide A band (associated with the NH-stretch, NH_S), respectively, according to literature [52-55] and our calculation (Figure S6). A side peak at around 2905 cm^{-1} is assigned to CH stretch and a broadband peak at 3150 cm^{-1} might result from non-resonant signal that is not the focus of this study.

In contrast, for the tumor sample, the relative $CH_{2,Ss}$ was suppressed, while CH_S became apparent and NH_S dominated (Figure 3b). We also noted the NH_S peak in tumor sample red-shifted to 3330 cm^{-1}. The NH_S peak is reported to be located at 3200-3400 cm^{-1} in literature[62], and it tends to

redshift as the NH group of the peptide participates in or experiences strengthened hydrogen bonds[63]. To eliminate the influence of water vapor absorption on the NH_S region, all experiments were performed inside of an environmental sample cell to maintain the relative humidity, while with the entire IR beam path under dry air purging. As discussed in Figure S4, OH stretch was concluded to have a negligible effect on the studied spectra region (2800-3400 cm^{-1}).

Overall, the VSFG spectra in Figure 3a,b suggested that the relative intensity of NH_S peak was much higher than the $CH_{2,Ss}$ in the tumor sample, while these two peaks had comparable intensities in the healthy sample. Thus, the peak intensity ratio between NH_S and $CH_{2,Ss}$ stretches could be used as a spectral signature to identify tumor tissues. To determine if this signature was statistically significant, we randomly measured VSFG spectra at 44 different spots in control samples and 17 spots in tumor samples. The $NH_S/CH_{2,Ss}$ ratio for control sample is 2.0 ± 0.7, while the ratio for tumor sample is 17 ± 6 (Figure 3c). These results confirm that statistically, the $NHs/CH_{2,Ss}$ ratio is dramatically different between these two samples.

We then performed hyperspectral image at $CH_{2,Ss}$ (IR centered at 2941 cm^{-1}) and NH_S (3333 cm^{-1}) region (Figure 3e), by setting our IR laser at these specific frequencies, to explore the potential of differentiating tumor tissues using VSFG image with specific vibrational modes. Unlike OPTIR, the VSFG images only show strong intensities in specific regions, because VSFG is sensitive to non-centrosymmetric collagen structures[51-57, 64-66], while other centrosymmetric TMEs did not appear. This selectivity eliminated unwanted background signal and allowed us to track changes in collagen tissue signals. Most importantly, the VSFG images revealed drastic differences between tumor and control tissues. For the control sample, VSFG images showed that the overall $CH_{2,Ss}$ and NH_S had similar intensity and spatial mapping (all images were displaced using the same color bar for comparison, Figure 3e, bottom). Conversely, the tumor sample presented a much higher NH_S intensity mapping than $CH_{2,Ss}$ (Figure 3e, top). The observed differences in VSFG image were consistent with those from the VSFG spectra (Figure 3a and b).

We then calculated the image intensity ratio from a specific area of the control sample image, which was used as the training data sets (highlighted in Figure 4c, details of analysis see SI, and Figure S5). The control sample showed an average $NH_S/CH_{2,Ss}$ ratio of 2.5, with a range from the 10th to the 90th percentile of 1.5 to 3.4. In contrast, the average $NH_S/CH_{2,Ss}$ ratio obtained from the entire tumor sample image was much higher ~ 56, with a range from the 10th to the 90th percentile of 6.3 to 127 (Table 1). The substantial difference indicates that a threshold based on the $NHs/CH_{2,Ss}$ intensity ratio can be established to directly identify tumor domains in the hyperspectral images.

From the $NH_S/CH_{2,Ss}$ ratio distribution histogram (Figure 4a), the control sample exhibited a Gaussian-like distribution, while the tumor sample overlaps slightly with the control but extends to much higher values. This overlap could be explained by the presence of healthy domains within the tumor sample, which is reasonable because not all the collagen tissues are fully degraded in tumors. Based on the histogram, we chose the upper bound of the 99.7% confidence interval of the $NH_S/CH_{2,Ss}$ ratio for control samples (i.e. the mean plus 3 times of the standard deviation) as the threshold for differentiating between tumor and control domains, leading to a boundary ratio of 8.5. This method ensured that nearly all testing area of control samples fall below the threshold while maintaining high sensitivity for detecting tumor regions.

Using this threshold, tumor samples exhibited a spatially dependent $NH/CH_{2,Ss}$ ratio, with most areas above the threshold (yellow-color domains in Figure 4c). In contrast, the $NH/CH_{2,Ss}$ ratio in control (healthy) samples consistently remained well below the threshold (Figure 4c). Remarkably, the area with $NH_S/CH_{2,Ss}$ overlaps with the tumor tissue area identified by the hematoxylin and eosin (H&E) staining (Figure 4e, f), validating that the $NH_S/CH_{2,Ss}$ ratio is a reliable spectral identifier for tumor tissue. (Note: the samples measured by VSFG microscopy and the stained samples are two consecutive slices from the same tissue sample, so their tumor areas are directly correlated).

Overall, the pronounced contrast in $NH_S/CH_{2,Ss}$ ratio between healthy and tumor tissues serves as a robust signature for tumor identification.

However, using the $NH/CH_{2,Ss}$ ratio has a practical limitation. In many VSFG systems (including ours), the two spectral regions are far apart, requiring scanning the mid-IR wavelength, which lengthens the imaging process. Thus, it is beneficial to find signatures with peak frequencies close to each other, allowing for the hyperspectral VSFG images without laser frequency scanning. Interestingly, as the NH_S increases its intensity in tumor samples, the CH_S stretch signal also becomes stronger relative to the $CH_{2,Ss}$ stretch, despite the overall spectral feature being weak (Figure 3a,b). This observation is further confirmed by plotting the $CH_S/CH_{2,Ss}$ ratio extracted the VSFG spectra; tumor samples exhibit a large $CH_S/CH_{2,Ss}$ ratio, while control samples display negligible ratios (Figure 3d).

Applying the same analysis to the $CH_S/CH_{2,Ss}$ ratio of the training image (shown in Figure 4b), the control sample has an average ratio of 0.07, with a range from the 10th to the 90th percentile of 0.04 to 0.11. For the tumor sample, the ratio is 1.13, with a range from the 10th to the 90th percentile of 0.11 to 2.52 (Table 1). Clearly, the $CH_S/CH_{2,\ Ss}$ intensity ratio is 10 times larger for the tumor samples than the healthy control, allowing us to establish a threshold of 0.22 (the upper bound of 99.7% confidence level of the control sample) to differentiate tumor from healthy tissues.

When applying this threshold value of 0.22 for the $CH/CH_{2\ Ss}$ ratio to differentiate tumor and healthy tissues, we found all test area of healthy tissue showed a ratio below the threshold, while the tumor tissues displayed a large domain above the threshold (yellow coded in Figure 4c) and a few small domains below it. The above threshold areas for the $CH_S/CH_{2,Ss}$ ratio largely coincided with those identified by $NH_S/CH_{2,Ss}$ ratio, suggesting that $CH_S/CH_{2,Ss}$ had a similar ability to identify collagen associated with tumor tissues. It was also noteworthy that there was good agreement between the areas with $NH_S/CH_{2,Ss}$ and $CH/CH_{2,Ss}$ ratios below their respective thresholds (orange in Figure 4c). By inspecting the spectra of a few specific spots, we observed that areas with a high NH_s peak always exhibited a visible CH_s feature (Figure 4d), which was absent in the control sample (Figure S13). This consistency in the spatial mapping further indicated that the collagen fibers of a small region in the tumor tissue maintain their normal structures.

Application to human tissues

The results from mouse lung tissues indicated VSFG microscopy could directly image tumor domain based on the $NH_S/CH_{2,Ss}$ and $CH_S/CH_{2,Ss}$ ratio mapping. To examine whether this observation can be extended to tissues from other mammals, we performed hyperspectral imaging on human lung tissues (obtained from Biochain Inc.). The hyperspectral VSFG images of $CH_{2,Ss}$, CH_S, and NH_S vibrational mode in Figure 5a showed that the NH_S signal was much stronger than the $CH_{2,Ss}$ signal in tumor tissues, whereas they were comparable in control samples – mirroring the observation in the mouse model.

Statistical analysis (Figure 5b) revealed significant differences in both the $NH_S/CH_{2,Ss}$ and $CH_S/CH_{2,Ss}$ ratios between tumor tissues and control samples. As summarized in Table 1, the mean values, as well as 10th and 90th percentile of both $NH_S/CH_{2,Ss}$ and $CH_S/CH_{2,Ss}$ were much higher in the tumor sample than in control sample. These substantial differences indicate that the $NH_S/CH_{2,Ss}$ and $CH_S/CH_{2,Ss}$ ratio could also be used to differentiate the human tumor tissue from the healthy ones.

Now, we applied the threshold obtained from the mouse sample to differentiate the tumor and healthy domains in human tissues. The ratio mappings (Figure 5c) indicate that all the area of healthy tissue exhibit $CH_S/CH_{2,Ss}$ and $NH_S/CH_{2,Ss}$ ratio below the threshold of 0.22, and 8.5, respectively. In contrast, the tumor tissues exhibit spatially dependent ratio, with most regions above the threshold (yellow coded in Figure 5c) and only a few exception areas. Similar to the mouse tissue, the areas identified as tumor tissues using $CH_S/CH_{2,Ss}$ and $NH_S/CH_{2,Ss}$ ratios largely

correspond with each other. Notably, similar results for tumor identification were obtained when applying a new threshold derived from statistical analyzing an area of the healthy human lung tissue. (see SI, Figure S14). Thus, this suggest that the spectral peak ratio of VSFG reflects fundamental structural changes in collagens, making the method applicable across different host species.

Comparison to other optical methods

Lastly, we examined whether similar results could be obtained using SHG images (Figure 5a) and VSFG spectroscopy. The overall intensity of SHG was lower than that of VSFG, likely because SHG involved only non-resonant excitations, unlike VSFG. Furthermore, SHG imaging only revealed morphological and intensity differences between the samples, which alone are insufficient to differentiate tumor tissues from healthy ones.

For VSFG spectroscopy (without imaging), although the spectra of the tumor and control samples display a similar line shape to those obtained using VSFG microscopy, the signal-to-noise ratio was extremely low under the same data acquisition conditions, making quantitative analysis impossible (Figure S15). This issue could be attributed to the ensemble-averaging of many fiber domains, which destructively interfere with each other due to the random orientations[37].

Discussion

While VSFG shows marked spectral differences between collagen tissues bearing metastatic tumors and healthy controls, underpinning the origins of these differences is crucial. One potential reason is that the $CH_{2,Ss}$ modes and NH_S (and CH_S) modes have different orientations, which can lead to the observed intensity change due to the relative orientation between the sample and laser beam polarization. If this is the case, we expect the relative peak intensity to vary dramatically with different polarization combination of VSFG pulse sequence changes[46,67]. However, we observed that all spectra from the same sample species displayed consistent relative peak intensities, regardless of the polarization schemes used (Figure S16 and S17). This consistency suggests that the changes in peak ratios changes are not due to sample orientations.

This leaves the origin of the peak ratio change to be structural, either on the molecular level, or on the mesoscopic assembly level. We first consider whether large chemical modifications occur[17-19]. If this occurred, they would affect the vibrational modes of amide bonds. However, OPTIR hyperspectral imaging did not reveal dramatic spectral changes in the N-H modes (Amide A) between the tumor and control tissues, making substantial chemical structure changes unobservable.

Next, we evaluate how collagen packing might influence the observed VSFG spectra. We consider several possibilities: (i) relative fibril alignments, (ii) interfibrillar distances, and (iii) fibril diameters. We can rule out the first option because if fibril alignments change from ordered to random, the symmetry of collagen samples would be lost, leading to a decrease of the overall VSFG signal. The similar spectral peak ratio regardless of the polarization schemes further confirm our interpretation (Figure S16 and S17).

To explore options (ii) and (iii), we applied the established nonlinear optical theory of VSFG to determine how these factors influence the coherent signal[46,67]. The dependence of the VSFG intensity on the diameter (d) of uniaxially aligned collagen fibrils and the distance between collagen fibrils (Δl) can be modeled via[46]

$$I_{VSFG} \propto I_{VIS} I_{IR} \left|\chi^{(2)}_{IJK}\right|^2 \quad (1)$$

$$\chi^{(2)}_{IJK} = \sum_{IJK} < R_{Ii} R_{Jj} R_{Kk} > \sum_{f=0}^{n-1} \beta_a^{fbl} \frac{e^{-i\Delta k d} - 1}{\Delta k} \left(e^{-i\Delta k (d+\Delta l) f}\right) \; n((-1)^f (cos\alpha) + cos\gamma sin\alpha) \quad (2)$$

Here, I_{VSFG}, I_{VIS}, and I_{IR} represent the intensities of the VSFG signal, the 1030 nm up-conversion beam, and the resonant IR beam, respectively. $\chi^{(2)}_{IJK}$ is the effective second-order susceptibility

tensor of the VSFG-active vibrational mode of the fibril, corresponding to the specific polarization combinations of VSFG, 1030 nm, and IR beams. R_{Ii}, R_{Jj} and R_{Kk} are the Euler matrices transforming the fiber frame to lab frame. β_a^{fbl} are the hyperpolarizability of the fibril frame. The phase mismatch is $\Delta k = k_{IR} + k_{vis} - k_{VSFG}$, a is the angle of the transition dipole vector with respect to the c axis, γ is the angle of the transition dipole vector with respect to the b axis (illustrated in Figure 6a), and n is the number of fibrils within the coherence length (L_c), $n = L_c/(d + \Delta l)$, where the coherent length $L_c = \frac{2}{\Delta k}$, d is the diameter of the collagen fibril and Δl is the effective spacing between the fibrils. $d + \Delta l$ can be regarded as the interfibrillar distance, (Figure S18), which influence the coherent sum of the VSFG signals. The $(e^{-i\Delta k(d+\Delta l)f})$ term represents the quasi-random phase-matching process[46] by the neighboring collagen fibrils. The $\frac{e^{-i\Delta kd}-1}{\Delta k}$ term accounts for the phase matching synchronization of the VSFG signal. The $(-1)^f(cos\alpha) + cos\gamma sin\alpha)$ term represents the bidirectional packing of the collagen fibrils.[67]

To simulate the VSFG intensity, we first determined the effective second-order susceptibility tensor of fibril, $\chi_{IJK}^{(2)}$. We compute the hyperpolarizability for each amino acid in the collagen I structure using Density Functional Theory (DFT) calculations with B3LYP/6-31g (d) level.[68-71] By summing the hyperpolarizability tensors of all amino acid with their appropriate orientations, we obtain second order susceptibilities of the fibril. This methodology aligns with the vibrational mode calculation of biological macromolecules in previous study[55]. The calculated angles of the vibrational modes with respect to the triple helix direction (c-axis) are: CH_S (60°), $CH_{2,Ss}$ (45°), and NH_S (80°), as shown in Figure 6a. We subsequently calculated the fiber 2nd order susceptibility by summing contributions of individual fibrils, considering the interfibrillar distance and fibril diameter, as described in equation (2). Further details can be found in the Supporting Information (SI).

Figure 6 b and c illustrate the impact of collagen fibril distance on the ratios of CH_S to CH_{2Ss} mode (left plot) and NH_S stretch to CH_{2Ss} mode (right plot) for various collagen fibril diameters. Both the CH_S/CH_{2Ss} and NH_S/CH_{2Ss} intensity decrease with increasing fibril distance over a certain range, regardless of collagen fibril diameters. Notably, the NH_S intensities become smaller than the $CH_{2,Ss}$ intensity as the fibril distance decreases. This result suggests that different vibrational mode interferes distinctively due to varying effective second-order susceptibilities. This is consistent with the fact that NH_S and CH_S modes have similar orientation while the $CH_{2,Ss}$ orients differently in the fibril frame. The oscillatory features observed in Figure 6b and c are caused by two effects: (i) As the inter-fibrillar distance Δl increases, the quasi-random phase matching term $(e^{-i\Delta k(d+\Delta l)f})$ undergoes periodic modulation.[67] (ii) In the bidirectional packing, the odd-numbered collagen fibrils correspond to intensity maxima, while even-numbered fibrils correspond to minima. This is due to the symmetry cancellation effect of transition dipoles oriented in opposite directions. The interplay between these two mechanisms results in an oscillation behavior (see SI).

The simulation results suggested a decrease of interfibrillar effective spacings within a certain range, which corresponds to an increase in the packing density. There are two possibilities that lead to this effect a) reducing interfibrillar distance, or b) enlarged fibril diameters. (as illustrated in Figure S18),. This insight remarkably agrees with our mechanical measurements from AFM shown in Figure 1, indicating that the tumor-related collagen is stiffer than healthy collagen. Since the elastic modulus is positively related to density, these two results now corroborate with each other – the collagen tissues in the tumor environments engendered denser packing, leading to enhanced stiffness. These preliminary structural insights warrant further quantitative theoretical analysis and investigation using other advanced structural biology techniques, such as cryo- transmission electron microscopy in the future[73].

Conclusion

In this study, we demonstrate that VSFG imaging uniquely combines the best of both worlds from coherent SHG and IR absorption spectroscopy. By using vibrational modes as the coherent light

emitting source, VSFG allows different vibrational modes within the sample interfere in distinct way, producing unique spectral lineshapes that can differentiate between collagen in tumor and healthy tissues. This approach provides a new label-free method for detecting tumor tissues. With the integration of fiber optics and endoscope techniques[74], this method has the potential to become a new non-invasive in vivo diagnostic tool.

By combining VSFG imaging with theoretical simulations, we corroborate that the spectral variations are due to the collagen density change, which result in the effective spacing between adjacent fibrils being smaller than the coherence length. These findings are consistent with the heightened stiffness in the tumor collagen samples, offering structural insights into the widely observed mechanical changes in tumor collagen. While the structural analysis presented here is preliminary due to the complexity of the sample and the multivariate nature of VSFG signals, the potential of VSFG as a structural analysis tool for complex samples can be further established. Future efforts to benchmark its spectral-structure relationships using by advanced structural biology tools such as cryo-EM, along with theoretical calculations, will be essential for advancing this method.

Materials and Methods

Sample preparation

The mouse metastatic lung tumors were developed following tail-vein injection of MDA-MB-231 human breast cancer cells into female NOD-scid IL2Rgammanull mice. The human lung tumor and control tissues were purchased from BioChain company. Both the mouse and human tumor/ healthy control samples were prepared by placing 2 drops of optimal cutting temperature (OCT) compound into a plastic cryomold. The filled cryomold was placed on top of aluminum foil resting on dry ice for rapid freezing. The specimens of collagen fibrils were cut into 20-μm thick sections using a Leica CM 1950 cryostat (Leica, Germany). The slices were laid on a glass coverslip (VWR International) and immersed in phosphate-buffered saline (pH 7.4, sterile-filtered, Sigma-Aldrich) with sodium azide (extra pure, Sigma-Aldrich) at 4 °C. Before the experiment, the coverslip was cleaned with deionized (DI) water in an ultrasonic bath for 5 min, followed by rinsing with ethanol and drying using a stream of nitrogen.

VSFG microscopy experiments

The VSFG microscopy used a pulsed laser system (Light Conversion, Carbide) centered at 1025nm ± 5nm (Figure S3). The laser operates at 40W, 200kHz (200μJ/pulse) with a pulse width of ~290 fs. The output of the seed laser was guided into a commercial optical parametric amplifier (OPA, LightConversion) to generate a mid-infrared (MIR) beam, and the frequency can be continuously tuned. For the up-conversion near IR (NIR) beam, we passed the residual 1025 nm beam from OPA through a Fabry-Perot etalon to produce a spectrally narrowed up-conversion beam with an FWHM of ~4.75 cm^{-1}. The polarization of the 1025 nm pulse and MIR beam were controlled with their corresponding λ/2 waveplates. Both up-conversion and MIR beams were spatially overlapped at a customized dichroic mirror (DM in Figure S3) that is transmissive to MIR and reflective to NIR. The overlapped beams were directed into an inverted microscope with an integrated 325Hz single-axis resonant beam scanner. The two spatially overlapped beams were focused onto the sample with a purely reflective 20× Schwarzschild objective (Figure S3). The VSFG signal generated by the sample was collected with an infinity, 20× refractive microscope objective (Zeiss, Fluar 0.75 NA, working distance = 0.6 mm) (Figure S3), which was guided through a linear polarizer and then through a telecentric tube lens system composed of two f = 60mm focal lenses (TL1 and TL2 in Figure S3). The magnified image from the tube lenses was formed at the entrance slit of the monochromator and the spatially/frequency resolved data was detected on a two-dimensional CCD detector (CCD in Figure S3). To switch to SHG imaging, we can block the IR beam and rotate the grating of the spectrograph to 501.5 nm to image the SHG signal.

Theory calculation and simulation

A computational approach to calculate the SFG intensity of various vibrational modes of Type I collagen is achieved in a four-step process (see Figure S7): (1) We performed the density functional theory (DFT) calculation of amino acids. (2) The amino acids were transformed to the coordinates frame of the collagen fibril. (3) The distance and diameter between collagen fibrils were considered using the numerical simulation to calculate the second order susceptibility. (4) Through the Euler transformation, the collagen fiber was transformed into the Lab frame to account for the experimental conditions. The calculation and simulation of each step is provided in the SI.

A detailed description of Materials and Methods is provided in SI Appendix.

Acknowledgments

B. Yang, J. Ren, C-C. Yu are supported by the National Institutes of Health, National Institute of General Medical Sciences, Grant 1R35GM138092-01 awarded to W. Xiong. M. Cao is supported by National Institutes of Health R01CA266486, awarded to S.E.Wang. We thank Dr. Z. Wu and Dr. J. C. Wagner for their efforts on the early stage of this research, Dr. Fraley for her inspiring discussion at the initial stage, Prof. W. Campana for providing facilities for mouse tissue cutting, and Dr. C.M. Pavez and Prof. Grassian for the access of the OPT-IR measurements. J.Ren and W. Xiong also want to acknowledge Prof. Groenhof and Dr. Morozov for their insightful discussion of collagen simulation using GROMACS. Figure 1, Figure 6a and Figure S7 were created by Biorender.com.

References

1. Abramczyk, H. & Brozek-Pluska, B. Raman imaging in biochemical and biomedical applications. Diagnosis and treatment of breast cancer. *Chem. Rev.* **113,** 5766–5781 (2013).
2. Hong, G., Antaris, A. L. & Dai, H. Near-infrared fluorophores for biomedical imaging. *Nat. Biomed. Eng.* **1,** 0010 (2017).
3. Lobato-Delgado, B., Priego-Torres, B. & Sanchez-Morillo, D. Combining molecular, imaging, and clinical data analysis for predicting cancer prognosis. *Cancers* **14,** 3215 (2022).
4. Alturkistani, H. A., Tashkandi, F. M. & Mohammedsaleh, Z. M. Histological stains: a literature review and case study. *Glob. J. Health Sci.* **8,** 72 (2016).
5. Liu, H. W. et al. Recent progresses in small-molecule enzymatic fluorescent probes for cancer imaging. *Chem. Soc. Rev.* **47,** 7140–7180 (2018).
6. Wang, F., Qu, L., Ren, F., Baghdasaryan, A., Jiang, Y., Hsu, R., Liang, P., Li, J., Zhu, G., Ma, Z. and Dai, H., High-precision tumor resection down to few-cell level guided by NIR-IIb molecular fluorescence imaging. *Proc. Natl. Acad. Sci.*,119, p.e2123111119 (2022).
7. Umezawa, K., Yoshida, M., Kamiya, M., Yamasoba, T. & Urano, Y. Rational design of reversible fluorescent probes for live-cell imaging and quantification of fast glutathione dynamics. *Nat. Chem.* **9,** 279–286 (2017).
8. Gao, M., Yu, F., Lv, C., Choo, J. & Chen, L. Fluorescent chemical probes for accurate tumor diagnosis and targeting therapy. *Chem. Soc. Rev.* **46,** 2237–2271 (2017).
9. Freudiger, C. W. et al. Label-free biomedical imaging with high sensitivity by stimulated Raman scattering microscopy. *Science* **322,** 1857–1861 (2008).
10. Ji, M. et al. Rapid, label-free detection of brain tumors with stimulated Raman scattering microscopy. *Sci. Transl. Med.* **5,** 201ra119 (2013).
11. Hollon, T. & Orringer, D. A. Label-free brain tumor imaging using Raman-based methods. *J. Neurooncol.* **151,** 393–402 (2021).
12. Fantuzzi, E. M. et al. Wide-field coherent anti-Stokes Raman scattering microscopy using random illuminations. *Nat. Photonics* **17,** 1097–1104 (2023).
13. Shaked, N. T., Boppart, S. A., Wang, L. V. & Popp, J. Label-free biomedical optical imaging. *Nat. Photonics* **17,** 1031–1041 (2023).

14. Lu, F.K., Basu, S., Igras, V., Hoang, M.P., Ji, M., Fu, D., Holtom, G.R., Neel, V.A., Freudiger, C.W., Fisher, D.E. and Xie, X.S., Label-free DNA imaging in vivo with stimulated Raman scattering microscopy. *Proc. Natl. Acad. Sci.*, *112*,11624-11629 (2015).
15. Ji, M. et al. Detection of human brain tumor infiltration with quantitative stimulated Raman scattering microscopy. *Sci. Transl. Med.* **7,** 309ra163 (2015).
16. Su, H. & Karin, M. Collagen architecture and signaling orchestrate cancer development. *Trends Cancer* **9,** 764–773 (2023).
17. Fang, M., Yuan, J., Peng, C. & Li, Y. Collagen as a double-edged sword in tumor progression. *Tumor Biol.* **35,** 2871–2882 (2014).
18. Egeblad, M., Rasch, M. G. & Weaver, V. M. Dynamic interplay between the collagen scaffold and tumor evolution. *Curr. Opin. Cell Biol.* **22,** 697–706 (2010).
19. Wohlgemuth, R. P., Brashear, S. E. & Smith, L. R. Alignment, cross-linking, and beyond: a collagen architect's guide to the skeletal muscle extracellular matrix. *Am. J. Physiol. Cell Physiol.* **325,** C1017–C1030 (2023).
20. Kobayashi, T., Gabazza, E. C., Taguchi, O., Risteli, J., Risteli, L., Kobayashi, H., Yasui, H. *et al.* Type I collagen metabolites as tumor markers in patients with lung carcinoma. *Cancer* **85**, 1951–1957 (1999).
21. Xi, G., Guo, W., Kang, D., Ma, J., Fu, F., Qiu, L., Zheng, L. *et al.* Large-scale tumor-associated collagen signatures identify high-risk breast cancer patients. *Theranostics* **11**, 3229 (2021).
22. Nyström, H., Naredi, P., Hafström, L., & Sund, M. Type IV collagen as a tumour marker for colorectal liver metastases. *Eur. J. Surg. Oncol.* **37**, 611–617 (2011).
23. Boyd, N. F., Lockwood, G. A., Byng, J. W., Tritchler, D. L. & Yaffe, M. J. Mammographic densities and breast cancer risk. *Cancer Epidemiol. Biomarkers Prev.* **7,** 1133–1144 (1998).
24. Lu, P., Weaver, V. M. & Werb, Z. The extracellular matrix: a dynamic niche in cancer progression. *J. Cell Biol.* **196,** 395–406 (2012).
25. Nadiarnykh, O., LaComb, R. B., Brewer, M. A. & Campagnola, P. J. Alterations of the extracellular matrix in ovarian cancer studied by Second Harmonic Generation imaging microscopy. *BMC Cancer* **10,** 94 (2010).
26. Levental, K. R. et al. Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell* **139,** 891–906 (2009).
27. Riegler, J. et al. Tumor elastography and its association with collagen and the tumor microenvironment. *Clin. Cancer Res.* **24,** 4455–4467 (2018).
28. Cicchi, R. et al. From molecular structure to tissue architecture: collagen organization probed by SHG microscopy. *J. Biophotonics* **6,** 129–142 (2013).
29. Chen, X., Nadiarnykh, O., Plotnikov, S. & Campagnola, P. J. Second harmonic generation microscopy for quantitative analysis of collagen fibrillar structure. *Nat. Protoc.* **7,** 654–669 (2012).
30. Williams, R. M., Zipfel, W. R. & Webb, W. W. Interpreting second-harmonic generation images of collagen I fibrils. *Biophys. J.* **88**, 1377–1386 (2005).
31. Tuer, A. E. et al. Nonlinear optical properties of type I collagen fibers studied by polarization dependent second harmonic generation microscopy. *J. Phys. Chem. B* **115,** 12759–12769 (2011).
32. Shin, K. S. et al. Intraoperative assessment of skull base tumors using stimulated Raman scattering microscopy. *Sci. Rep.* **9,** 20392 (2019).
33. Synytsya, A. et al. Evaluation of IR and Raman spectroscopic markers of human collagens: Insides for indicating colorectal carcinogenesis. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **296,** 122664 (2023).
34. Petibois, C. & Déléris, G. Chemical mapping of tumor progression by FT-IR imaging: towards molecular histopathology. *Trends Biotechnol.* **24,** 455-462 (2006).
35. Noreen, R. et al. FTIR spectro-imaging of collagen scaffold formation during glioma tumor development. *Anal. Bioanal. Chem.* **405,** 8729-8736 (2013).
36. Spadea, A. et al. Analysis of fixed and live single cells using optical photothermal infrared with concomitant Raman spectroscopy. *Anal. Chem.* **93,** 3938-3950 (2021).

37. Wang, H., Gao, T. & Xiong, W. Self-phase-stabilized heterodyne vibrational sum frequency generation microscopy. *ACS Photonics* **4,** 1839-1845 (2017).
38. Wang, H. & Xiong, W. Vibrational sum-frequency generation hyperspectral microscopy for molecular self-assembled systems. *Annu. Rev. Phys. Chem.* **72,** 279-306 (2021).
39. Wagner, J. C. et al. Water capture mechanisms at zeolitic imidazolate framework interfaces. *J. Am. Chem. Soc.* **143,** 21189-21194 (2021).
40. Cimatu, K. & Baldelli, S. Sum frequency generation microscopy of microcontact-printed mixed self-assembled monolayers. *J. Phys. Chem. B* **110,** 1807-1813 (2006).
41. Fellows, A. P. et al. Spiral packing and chiral selectivity in model membranes probed by phase-resolved sum-frequency generation microscopy. *Nat. Commun.* **15,** 3161 (2024).
42. Stiopkin, I. V., Weeraman, C., Pieniazek, P. A., Shalhout, F. Y., Skinner, J. L., & Benderskii, A. V. Hydrogen bonding at the water surface revealed by isotopic dilution spectroscopy. *Nature* **474**, 192 – 195 (2011).
43. Farah, Y. R. & Krummel, A. T. The pH-dependent orientation of N3 dye on a gold substrate is revealed using heterodyne-detected vibrational sum frequency generation spectroscopy. *J. Chem. Phys.* **154**, 124702 (2021).
44. Hsieh, C.-S., Okuno, M., Hunger, J., Backus, E. H. G., Nagata, Y., & Bonn, M. Aqueous heterogeneity at the air/water interface revealed by 2D-HD-SFG spectroscopy. *Angew. Chem. Int. Ed.* **53**, 8146–8149 (2014).
45. Cotton, D. E. & Roberts, S. T. Sensitivity of sum frequency generation experimental conditions to thin film interference effects. *J. Chem. Phys.* **154**, 114704 (2021).
46. Makarem, M. et al. Distinguishing mesoscale polar order (unidirectional vs bidirectional) of cellulose microfibrils in plant cell walls using sum frequency generation spectroscopy. *J. Phys. Chem. B* **124,** 8071-8081 (2020).
47. Lee, C. M., Mittal, A., Barnette, A. L., Kafle, K., Park, Y. B., Shin, H. & Kim, S. H. Cellulose polymorphism study with sum-frequency-generation (SFG) vibration spectroscopy: identification of exocyclic CH_2OH conformation and chain orientation. *Cellulose* **20**, 991 – 1000 (2013).
48. Lee, C. M., Chen, X., Weiss, P. A., Jensen, L. & Kim, S. H. Quantum mechanical calculations of vibrational sum-frequency-generation (SFG) spectra of cellulose: dependence of the CH and OH peak intensity on the polarity of cellulose chains within the SFG coherence domain. *J. Phys. Chem. Lett.* **8**, 55–60 (2017).
49. Wagner, J. C., Yang, B., Wu, Z. & Xiong, W. Multimodal nonlinear hyperspectral chemical imaging using line-scanning vibrational sum-frequency generation microscopy. *J. Visualized Exp.* **202**, e61103 (2023).
50. Wagner, J. C., Wu, Z., Wang, H. & Xiong, W. Imaging orientation of a single molecular hierarchical self-assembled sheet: the combined power of vibrational sum frequency generation microscopy and neural network. *J. Phys. Chem. B* **126**, 7192–7201 (2022).
51. Han, Y., Hsu, J., Ge, N. H. & Potma, E. O. Polarization-sensitive sum-frequency generation microscopy of collagen fibers. *J. Phys. Chem. B* **119**, 3356–3365 (2015).
52. Han, Y. et al. Mapping molecular orientation with phase-sensitive vibrationally resonant sum-frequency generation microscopy. *J. Phys. Chem. B* **117**, 6149–6156 (2013).
53. Maekawa, H., Kumar, S. K., Mukherjee, S. S. & Ge, N. H. Phase-sensitive vibrationally resonant sum-frequency generation microscopy in multiplex configuration at 80 MHz repetition rate. *J. Phys. Chem. B* **125**, 9507–9516 (2021).
54. Rocha-Mendoza, I. et al. Sum frequency vibrational spectroscopy: the molecular origins of the optical second-order nonlinearity of collagen. *Biophys. J.* **93**, 4433–4444 (2007).
55. Johansson, P. K. & Koelsch, P. Vibrational sum-frequency scattering for detailed studies of collagen fibers in aqueous environments. *J. Am. Chem. Soc.* **136**, 13598–13601 (2014).
56. Johansson, P. K. & Castner, D. G. Vibrational sum-frequency scattering as a sensitive approach to detect structural changes in collagen fibers treated with surfactants. *Langmuir* **35**, 7848–7857 (2019).
57. Luna Palacios, Y. Y. et al. Spectroscopic analysis of the sum-frequency response of the carbon–hydrogen stretching modes in collagen type I. *J. Chem. Phys.* **160**, 185101 (2024).

58. Cao, M. et al. Cancer-cell-secreted extracellular vesicles suppress insulin secretion through miR-122 to impair systemic glucose homeostasis and contribute to tumor growth. *Nat. Cell Biol.* **24**, 954–967 (2022).
59. Layton, B. E. & Sastry, A. M. Equal and local-load-sharing micromechanical models for collagens: quantitative comparisons in response of non-diabetic and diabetic rat tissue. *Acta Biomater.* **2**, 595–607 (2006).
60. Ahn, T., Jueckstock, M., Mandair, G. S., Henderson, J., Sinder, B. P., Kozloff, K. M. & Banaszak Holl, M. M. Matrix/mineral ratio and domain size variation with bone tissue age: a photothermal infrared study. *J. Struct. Biol.* **214**, 107878 (2022).
61. Mankar, R. et al. Polarization sensitive photothermal mid-infrared spectroscopic imaging of human bone marrow tissue. *Appl. Spectrosc.* **76**, 508–518 (2022).
62. Synytsya, A. et al. Evaluation of IR and Raman spectroscopic markers of human collagens: insights for indicating colorectal carcinogenesis. *Spectrochim. Acta A* **296**, 122664 (2023).
63. Krishnan, R. S. & Krishnan, K. Influence of the hydrogen bond on the NH stretching frequencies in amino acids. *Proc. Indian Acad. Sci. A* **60**, 11–19 (1964).
64. Reiser, K. M., McCourt, A. B., Yankelevich, D. R. & Knoesen, A. Structural origins of chiral second-order optical nonlinearity in collagen: amide I band. *Biophys. J.* **103**, 2177–2186 (2012).
65. Hanninen, A., Shu, M. W. & Potma, E. O. Hyperspectral imaging with laser-scanning sum-frequency generation microscopy. *Biomed. Opt. Express* **8**, 4230–4242 (2017).
66. Howell, C., Diesner, M.-O., Grunze, M. & Koelsch, P. Probing the extracellular matrix with sum-frequency-generation spectroscopy. *Langmuir* **24**, 13819–13821 (2008).
67. Baudrier-Raybaut, M., Haïdar, R., Kupecek, P., Lemasson, P. & Rosencher, E. Random quasi-phase-matching in bulk polycrystalline isotropic nonlinear materials. *Nature* **432**, 374-376 (2004).
68. Becke, A. D. Density-functional thermochemistry III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).
69. Lee, C., Yang, W. & Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37**, 785–789 (1988).
70. Miehlich, B., Savin, A., Stoll, H. & Preuss, H. Results obtained with the correlation energy density functionals of Becke and Lee, Yang and Parr. *Chem. Phys. Lett.* **157**, 200–206 (1989).
71. Hehre, W. J., Ditchfield, R. & Pople, J. A. Self-consistent molecular orbital methods. XII. Further extensions of Gaussian–type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.* **56**, 2257–2261 (1972).
72. Siadat, S. M., Silverman, A. A., DiMarzio, C. A. & Ruberti, J. W. Measuring collagen fibril diameter with differential interference contrast microscopy. *J. Struct. Biol.* **213**, 107697 (2021).
73. Sharp, T. H. et al. Cryo-transmission electron microscopy structure of a gigadalton peptide fiber of de novo design. *Proc. Natl. Acad. Sci.* **109**, 13266–13271 (2012).
74. Kaur, M., Lane, P. M. & Menon, C. Scanning and actuation techniques for cantilever-based fiber optic endoscopic scanners—a review. *Sensors* **21**, 251 (2021).

Figures and Tables

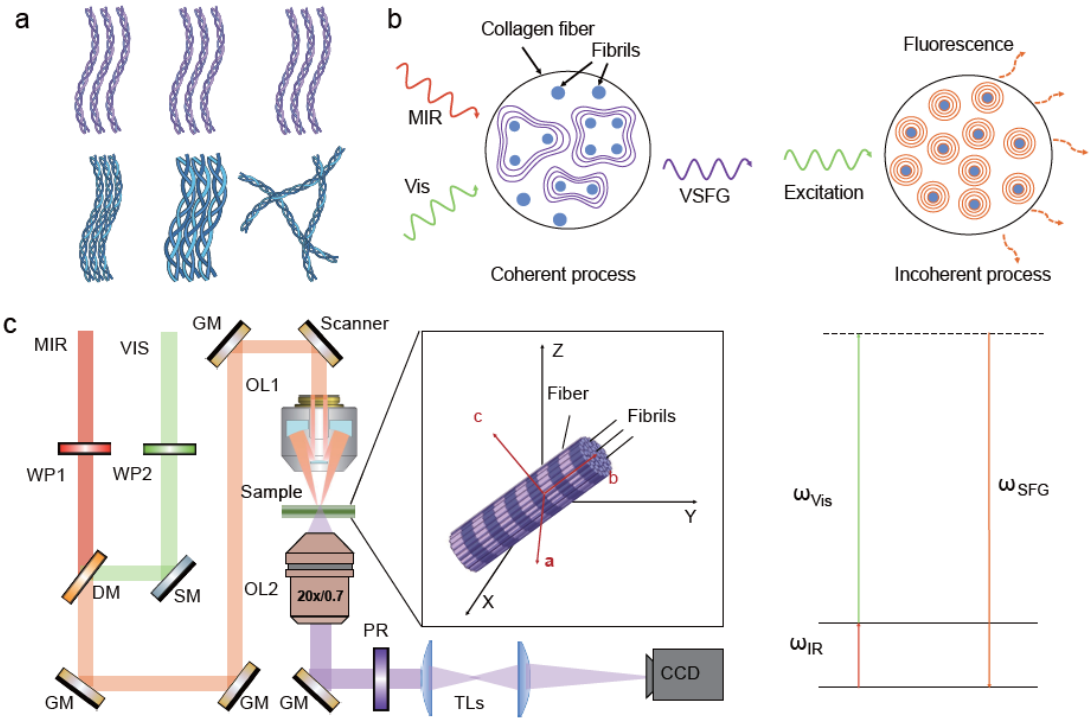

Figure 1. VSFG microscope to image collagen tissues. (a) Schematic diagram of collagen structure changes in tumors. The top figures indicate the normal collagen structure, while the bottom figures illustrate the possible structural changes caused by a tumor. From left to right, the columns represent changes in distance, diameter, and orientation, respectively. (b) Difference in optical signals toward collagen arrangement: VSFG, as a coherent nonlinear optical signal, can exhibit sensitivity to fibril/fiber packing; in contrast, incoherent optical signal, such as fluorescence, is insensitive to these arrangements. (c) Schematic diagram of VSFG microscopy and the corresponding fiber and lab frame (inset). Jablonski diagram of VSFG (right). WP: wave plate; DM: dichroic mirror; GM: gold mirror; SM: silver mirror; OL: objective; PR: Polarizer; TLs: Tube Lens.

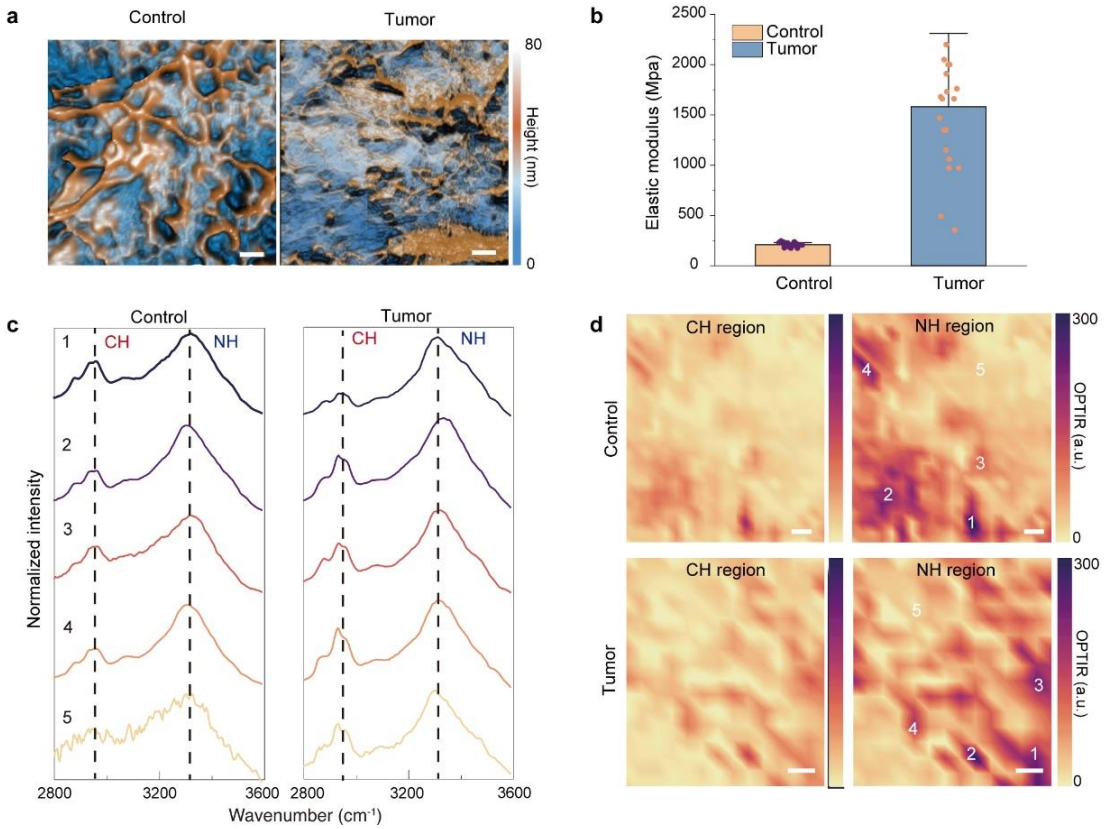

Figure 2. Mechanical and IR mapping of collagen tissues. (a) AFM shows the collagen morphology of control and tumor sample. The scale bar for control is 0.5 μm, while the scale bar for tumor is 2 μm. Vertical color bars represent height in 80 nm. (b) Statistic of the elastic modulus of lung control and tumor. Bars represent the means, and error bars are standard deviations. (c) Spectra of different spots of control and tumor tissue. (d) OPTIR image of CH and NH region of control and tumor tissue. Scale bars, 2 μm.

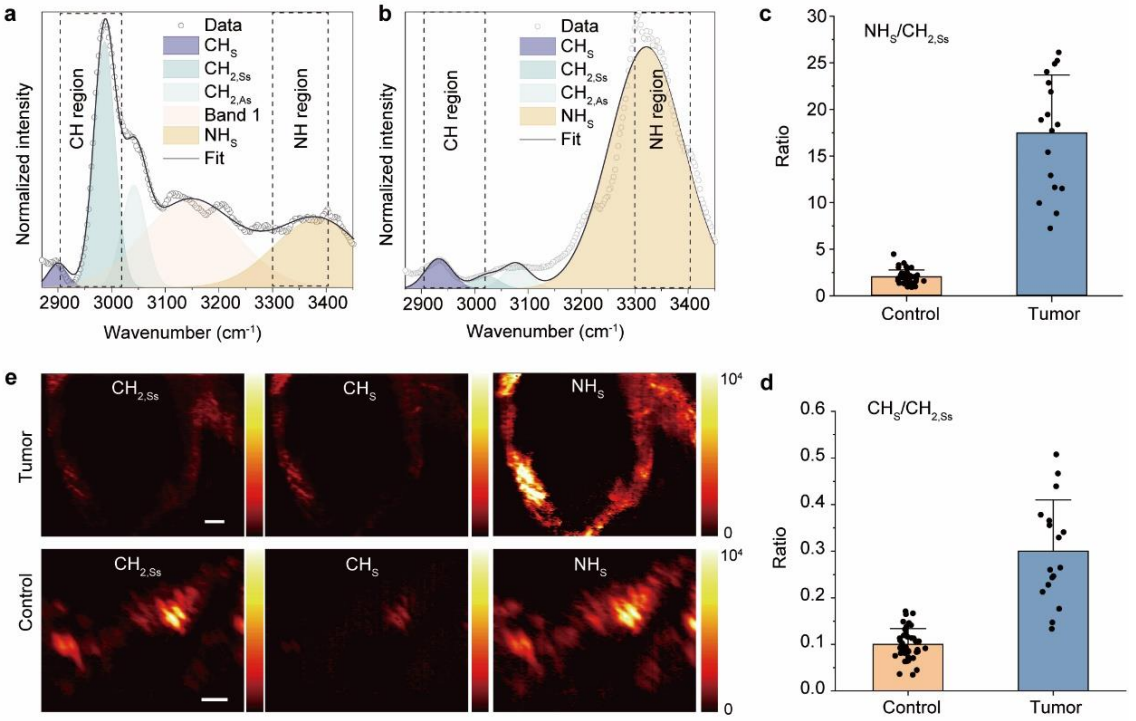

Figure 3. Representative VSFG spectra of control (a) and tumor (b) tissue. Statistic results show (c) the $NH_S/CH_{2,Ss}$ and (d) $CH_S/CH_{2,Ss}$ based on the spatially resolved VSFG spectra at control and tumor tissues. Bars represent the means, and error bars are standard deviations. (e). VSFG image of $CH_{2,Ss}$, CH_S, and NH_S mode of control and tumor tissue. The scale bar is 20 μm.

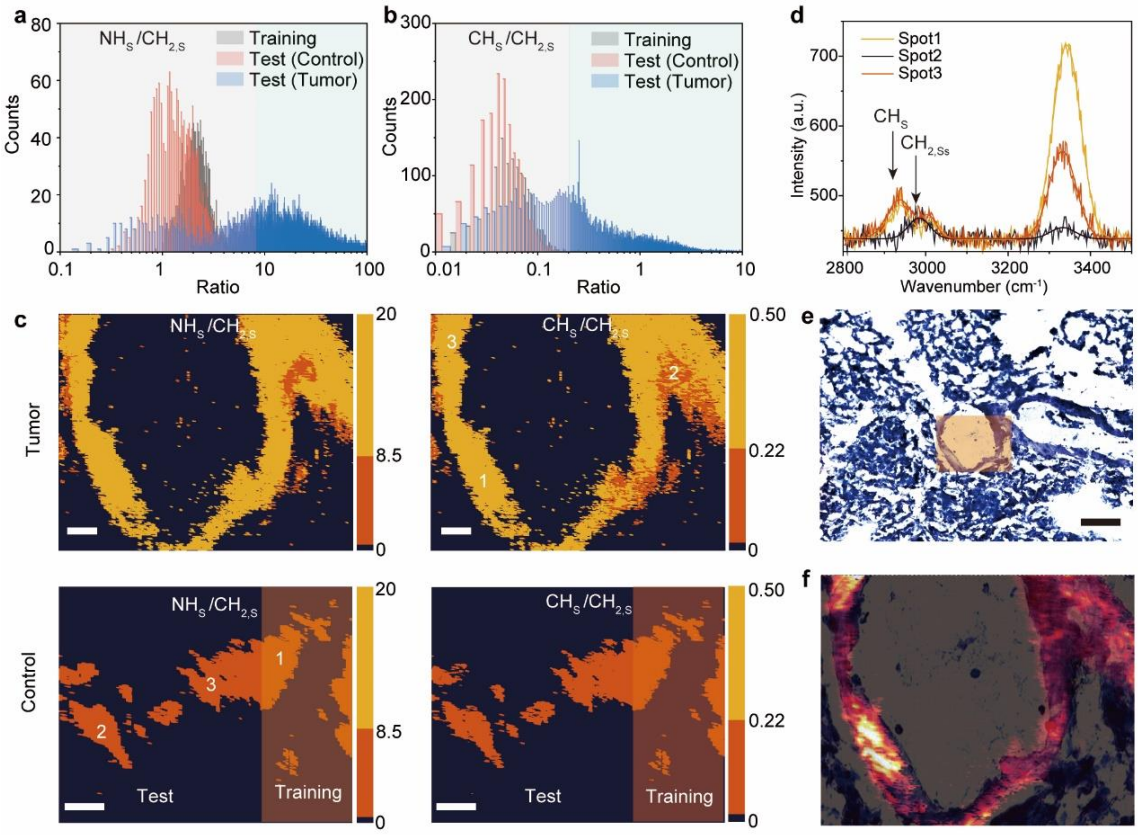

Figure 4. VSFG peak ratio statistics and the corresponding hyperspectral images. Statistic results of (a) $NH_S/CH_{2,Ss}$ and (b) $CH_S/CH_{2,Ss}$ of the mouse control and tumor tissue. The gray, red and green bars show the ratio distribution of the training area of control samples, the test area of control samples, and the image of tumor sample. Note the training and test data of the control samples yield similar distributions and therefore overlap with each other in this plot. (c) $NH_S/CH_{2,Ss}$ and $CH_S/CH_{2,Ss}$ ratio mapping of the mouse control and tumor tissue. The scale bar is 20 μm. (d) Spatially resolved VSFG spectra show the CH and NH region of tumor sample. (e) Bright field image of tumor tissue after H&E stain. It is observed from the H&E staining image that the detected collagen is surrounded by tumor cells. The scale bar is 100 μm (f)An overlaid image of the VSFG image and the H&E image of tumor tissue, taken from the orange-shaded region of (e).

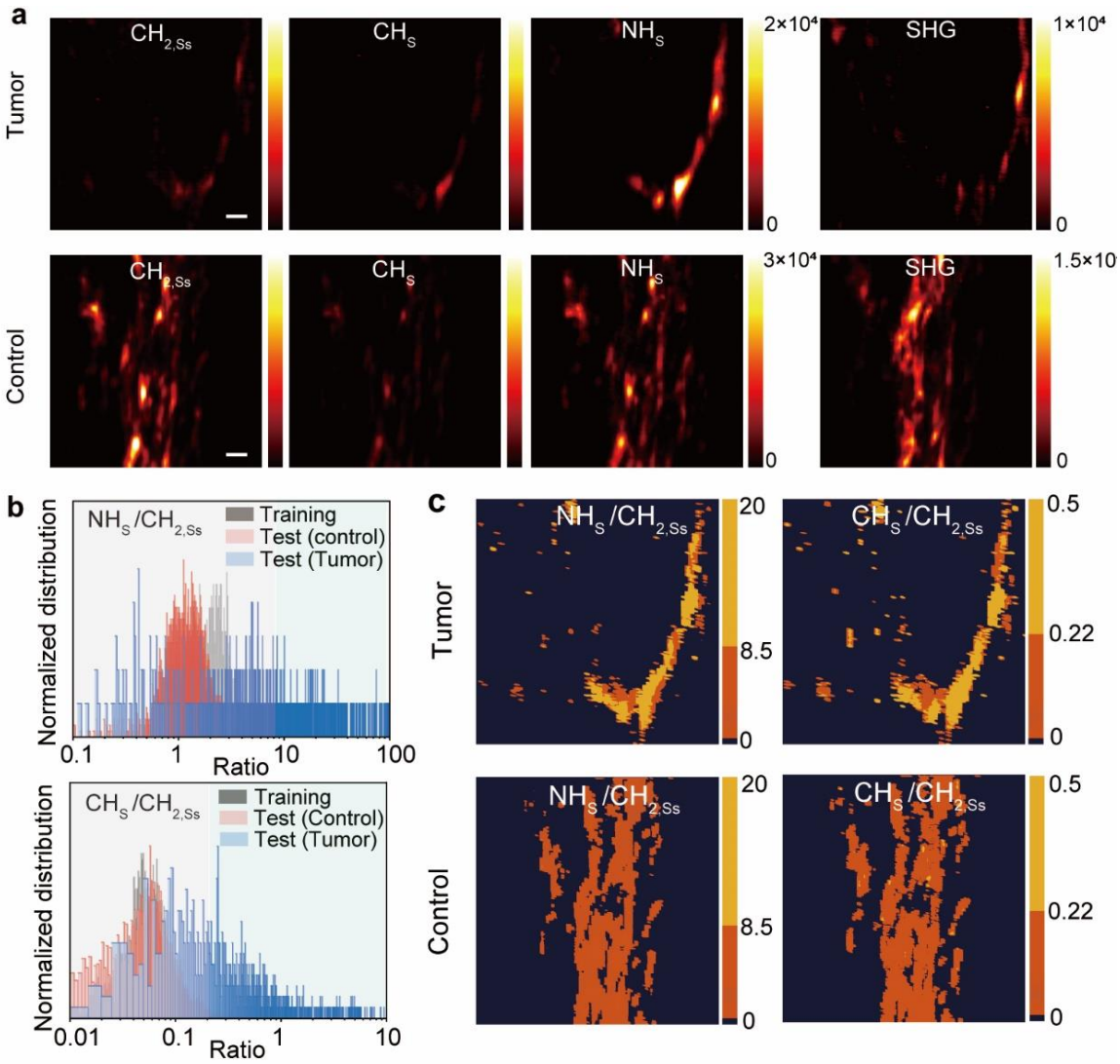

Figure 5. VSFG images of human lung samples. (a) VSFG image of $CH_{2,Ss}$, CH_S, NH_S mode and SHG image of human control and tumor tissue. The scale bar is 15 μm (b) Statistic results $NH_S/CH_{2,Ss}$ and $CH_S/CH_{2,Ss}$ of the human control and tumor tissue. (c) $NH_S/CH_{2,Ss}$ and $CH_S/CH_{2,Ss}$ ratio mapping of the human control and tumor tissue.

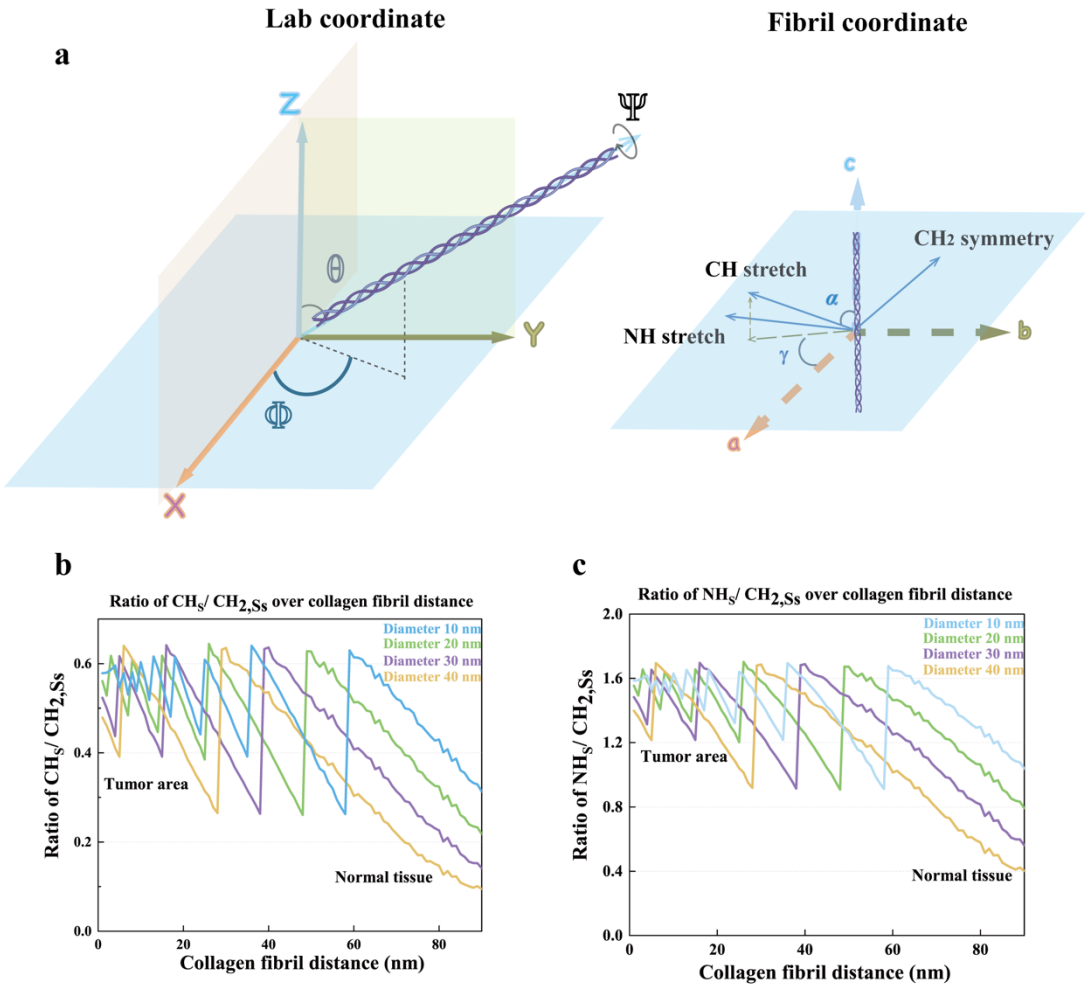

Figure 6. Theoretical modeling to understand the original of the VSFG peak intensity changes. The selection of collagen fibril diameters (10, 20, 30 and 40 nm) was guided by employing differential interference contrast (DIC) microscopy enhanced with extraordinary image shift (EIS) techniques. [72] Theses studies have revealed that collagen fibrils can exist in diameters below 40 nm. The collagen fibril distance range (0-100 nm) was carefully chosen to align with the coherence length (L_c), calculated to be around 130 nm. (a) Orientation of vibrational modes relative to collagen fibril direction, and effect of collagen fibril distance and diameter on the ratios of (b) $CH_S/CH_{2,Ss}$and (c) $NH_S/CH_{2,Ss}$ modes.

Table 1. Statistic results of mouse and human lung sample.

Sample		Ratio	Mean	10 th	90 th
Mouse	Control	$NH_S/CH_{2,Ss}$	2.5	1.5	3.4
		$CH_S/CH_{2,Ss}$	0.07	0.04	0.11
	Tumor	$NH_S/CH_{2,Ss}$	56	6.3	127
		$CH_S/CH_{2,Ss}$	1.13	0.11	2.52
Human	Control	$NH_S/CH_{2,Ss}$	1.9	0.8	3.5
		$CH_S/CH_{2,Ss}$	0.08	0.02	0.16
	Tumor	$NH_S/CH_{2,Ss}$	20	1.2	44
		$CH_S/CH_{2,Ss}$	0.88	0.09	2.3