



A robust collagen scoring method for human liver fibrosis by second harmonic microscopy.

Thomas Guilbert, Christophe Odin, Yann Le Grand, Luc Gailhouste, Bruno Turlin, Frédéric Ezan, Yoann Désille, Georges Baffet, Dominique Guyader

► To cite this version:

Thomas Guilbert, Christophe Odin, Yann Le Grand, Luc Gailhouste, Bruno Turlin, et al.. A robust collagen scoring method for human liver fibrosis by second harmonic microscopy.. Optics Express, 2010, 18 (25), pp.25794-807. 10.1364/OE.18.025794 . hal-00684532

HAL Id: hal-00684532

<https://hal.science/hal-00684532>

Submitted on 1 Oct 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A robust collagen scoring method for human liver fibrosis by second harmonic microscopy

Thomas Guilbert,¹ Christophe Odin,^{1,*} Yann Le Grand,^{1,2} Luc Gailhouste,³ Bruno Turlin,⁴ Frédéric Ezan,³ Yoann Désille,⁵ Georges Baffet,³ and Dominique Guyader⁵

¹ Institut of Physics of Rennes IPR/UMR CNRS 6251, University of Rennes I, Campus de Beaulieu, Bat 11A, 35042 Rennes Cedex, France

² Laboratoire de Spectrométrie et Optique Laser, EA 938, University of Bretagne Occidentale, 29238 Brest, France

³ EA eRAIC 4427, IFR 140, University of Rennes I, 35033 Rennes, France

⁴ Department of Pathology, Hôpital Pontchaillou, 35033 Rennes, France

⁵ Liver Department Hôpital Pontchaillou, 35033 Rennes, France

*christophe.odin@univ-rennes1.fr

Abstract: Second Harmonic Generation (SHG) microscopy offers the opportunity to image collagen of type I without staining. We recently showed that a simple scoring method, based on SHG images of histological human liver biopsies, correlates well with the Metavir assessment of fibrosis level (Gailhouste et al., J. Hepatol., 2010). In this article, we present a detailed study of this new scoring method with two different objective lenses. By using measurements of the objectives point spread functions and of the photomultiplier gain, and a simple model of the SHG intensity, we show that our scoring method, applied to human liver biopsies, is robust to the objective's numerical aperture (NA) for low NA, the choice of the reference sample and laser power, and the spatial sampling rate. The simplicity and robustness of our collagen scoring method may open new opportunities in the quantification of collagen content in different organs, which is of main importance in providing diagnostic information and evaluation of therapeutic efficiency.

© 2010 Optical Society of America

OCIS codes: (180.4315) Nonlinear microscopy; (170.3880) Medical and biological imaging; (170.4580) Optical diagnostics for medicine.

References and links

1. S. L. Friedman, "Liver fibrosis - from bench to bedside," J. Hepatol. **38**, 38-53 (2003).
2. R. Bataller and D. A. Brenner, "Liver fibrosis," J. Clin. Invest. **115**, 209-218 (2005).
3. N. H. Afdhal and D. Nunes, "Evaluation of liver fibrosis: a concise review," Am. J. Gastroenterol. **99**, 1160-1174 (2004).
4. G. L. Davis, J. E. Albright, S. F. Cook, and D. M. Rosenberg, "Projecting future complications of chronic hepatitis C in the United States," Liver Transpl. **9**, 331-338 (2003).
5. F. Oberti, E. Valsesia, C. Pilette, M. C. Rousselet, P. Bedossa, C. Aub, Y. Gallois, H. Rifflet, M. Y. Maga, D. Penneau-Fontbonne, and P. Cals, "Noninvasive diagnosis of hepatic fibrosis or cirrhosis," Gastroenterology **113**, 1609-1616 (1997).
6. N. H. Afdhal, A. P. Keaveny, S. B. Cohen, D. P. Nunes, N. Maldonado, M. O'Brien, and P. J. Stone, "Urinary assays for desmosine and hydroxylsypyrindoline in the detection of cirrhosis," J. Hepatol. **27**, 993-1002 (1997).
7. S. C. Faria, K. Ganesan, I. Mwangi, M. Shiehmoza, B. Viamonte, S. Mazhar, M. Peterson, Y. Kono, C. Santillan, G. Casola, and C. B. Sirlin, "MR Imaging of Liver Fibrosis: Current State of the Art," Radiographics **29**, 1615-1635 (2009).

8. C. Aube, F. Oberti, N. Korali, M.-A. Namour, D. Loisel, M. C. Rousselet, P. Bedossa, H. Rifflet, M. Y. Maga, C. Penneau, C. Caron, and P. Cals, "Ultrasonographic diagnosis of hepatic fibrosis or cirrhosis," *J. Hepatol.* **30**, 472-478 (1999).
9. H. A. Saleh and A. H. Abu-Rashed, "Liver Biopsy Remains the Gold Standard for Evaluation of Chronic Hepatitis and Fibrosis," *J. Gastrointest. Liver Dis.* **16**, 425-426 (2007).
10. P. Bedossa, D. Dargre, and V. Paradis, "Sampling variability of liver fibrosis in chronic hepatitis C," *J. Hepatol.* **38**, 1356-1358 (2003).
11. P. J. Scheuer, "Classification of chronic viral hepatitis : a need for reassessment," *J. Hepatol.* **13**, 372-374 (1991).
12. K. Ishak, A. Baptista, L. Bianchi, F. Callea, J. De Groote, F. Gudat, H. Denk, V. Desmet, G. Korb, and R. N. MacSween, "Histological grading and staging of chronic hepatitis," *J. Hepatol.* **22**, 696-699 (1995).
13. The French METAVIR Cooperative Study Group, "Intraobserver, interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C," *J. Hepatol.* **20**, 15-20 (1995).
14. M. Masseroli, T. Caballero, F. O'Valle, R. M. Del Moral, A. Perez-Milena, and R. G. Del Moral, "Automatic quantification of liver fibrosis design and validation of a new image analysis method: comparison with semi-quantitative indexes of fibrosis," *J. Hepatol.* **32**, 453-464 (2000).
15. A. Y. Hui, C.-T. Liew, M. Y.-Y. Go, A. M.-L. Chim, H. L.-Y. Chan, N. W.-Y. Leung, and J. J.-Y. Sung, "Quantitative assessment of fibrosis in liver biopsies from patients with chronic hepatitis B," *Liver Intl.* **24**, 611-618 (2004).
16. M. A. Friedenberg, L. Miller, C. Y. Chung, F. Fleszler, F. L. Banson, R. Thomas, K. P. Swartz, and F. K. Friedenberg, "Simplified method of hepatic fibrosis quantification: design of a new morphometric analysis application," *Liver Intl.* **25**, 1156-1161 (2005).
17. K. König (2000), "Multiphoton microscopy in life sciences," *J. Microsc.* **200**, 83-104.
18. A. Zoumi, A. Yeh, and B. J. Tromberg, "Imaging cells and extracellular matrix in vivo by using second-harmonic generation and two-photon excited fluorescence," *Proc. Natl. Acad. Sci. U.S.A.* **20**, 11014-11019 (2002).
19. P. J. Campagnola and L. M. Loew, "Second-harmonic imaging microscopy for visualizing biomolecular arrays in cells, tissues and organisms," *Nat. Biotechn.* **21**, 1356-1360 (2003).
20. W.R. Zipfel, R. M. Williams, and W. W. Webb, "Nonlinear magic : multiphoton microscopy in the biosciences," *Nat. Biotechn.* **21**, 1369-1377 (2003).
21. Zipfel, W. R., R. M. Williams, R. Christie, A. Y. Nikitin, B. T. Hyman, and W. W. Webb, "Live tissue intrinsic emission microscopy using multiphoton-excited native fluorescence and second harmonic generation," *Proc. Natl. Acad. Sci. USA.* **100**, 7075-7080 (2003).
22. S. Roth and I. Freund, "Second harmonic generation in collagen," *J. Chem. Phys.* **70**, 1637-1643 (1979).
23. S. W. Chu, S. Y. Chen, G. W. Chern, T. H. Tsai, Y. C. Chen, B. L. Lin, and C.K. Sun, "Studies of $\chi(2)/\chi(3)$ tensors in submicron-scaled bio-tissues by polarization harmonics optical microscopy," *Biophys. J.* **86**, 3914-3922 (2004).
24. R. M. Williams, W.R. Zipfel, and W. W. Webb, "Interpreting Second-Harmonic Generation Images of Collagen I Fibrils," *Biophys. J.* **88**, 1377-1386 (2005).
25. S. V. Plotnikov, A. C. Millard, P. J. Campagnola, and W. A. Mohler, "Characterization of the myosin-based source for second-harmonic generation from muscle sarcomeres," *Biophys. J.* **90**, 693-703 (2006).
26. C. Odin, Y. Le Grand, A. Renault, L. Gailhouse, and G. Baffet, "Orientation fields of nonlinear biological fibrils by second harmonic generation microscopy," *J. Microscopy* **229**, 32-38 (2008).
27. C. Odin, T. Guilbert, A. Alkilani, O. Boryskina, V. Fleury, and Y. Le Grand, "Collagen and myosin characterization by orientation field second harmonic microscopy," *Opt. Express* **16**, 16151-16165 (2008).
28. C. Odin, C. Heichette, D. Chretien, and Y. Le Grand, "Second harmonic microscopy of axonemes," *Opt. Express* **17**, 9235-9240 (2009).
29. S. Psilodimitrakopoulos, D. Artigas, G. Soria, I. Amat-Roldan, A. Planas, and P. Loza-Alvarez, "Quantitative discrimination between endogenous SHG sources in mammalian tissue, based on their polarization response," *Opt. Express* **17**, 10168-10176 (2009).
30. S. Psilodimitrakopoulos, V. Petegnief, G. Soria, I. Amat-Roldan, D. Artigas, A. Planas, and P. Loza-Alvarez, "Estimation of the effective orientation of the SHG source in primary cortical neurons," *Opt. Express* **17**, 14418-14425 (2009).
31. M. Strupler, A. Pena, M. Hernest, P. Tharaux, J. Martin, E. Beaurepaire, and M. Schanne-Klein, "Second harmonic imaging and scoring of collagen in fibrotic tissues," *Opt. Express* **15**, 4054-4065 (2007).
32. A.-M. Pena, A. Fabre, D. Dbarre, J. Marchal-Somme, B. Crestani, J.-L. Martin, E. Beaurepaire, and M.-C. Schanne-Klein, "Three-dimensional investigation and scoring of extracellular matrix remodeling during lung fibrosis using multiphoton microscopy," *Microsc. Res. Tech.* **70**, 162-170 (2007).
33. X. Han, R. M. Burke, M. L. Zettel, P. Tang, and E. B. Brown, "Second harmonic properties of tumor collagen: determining the structural relationship between reactive stroma and healthy stroma," *Opt. Express* **16**, 1846-1859 (2008). <http://www.opticsinfobase.org/abstract.cfm?URI=oe-16-3-1846>
34. D. C. S. Tai, N. Tan, S. Xu, C. H. Kang, S. M. Chia, C. L. Cheng, A. Wee, C. Li Wei, A. M. Raja, G. Xiao, S. Chang, J. C. Rajapakse, P. T. C. So, H.-H. Tang, C. Shing Chen, and H. Yu, "Fibro-C-Index: comprehensive

- sive, morphology-based quantification of liver fibrosis using second harmonic generation and two-photon microscopy," *J. Biomed. Opt.* **14**(4), 044013-1-10 (2009), DOI:10.1117/1.3183811
35. L. Gailhouste, Y. Le Grand, C. Odin, D. Guyader, B. Turlin, F. Ezan, Y. Dsille, T. Guilbert, A. Bessard, C. Frmin, N. Theret, and G. Baffet, "Fibrillar Collagen Scoring by Second Harmonic Microscopy as a New Tool in the Assessment of Liver Fibrosis," *J. of Hepatology* **52**, 398-406 (2010).
 36. P. Bedossa, "Harmony in liver fibrosis," *J. of Hepatology* **52**, 313-314 (2010).
 37. J. Mertz and L. Moreaux, "Second-harmonic generation by focused excitation of inhomogeneously distributed scatterers," *Optics Com.* **196**, 325-330 (2001).
 38. L. I-K Lin, "A concordance correlation coefficient to evaluate reproducibility," *Biometrics* **45**, 255-268 (1989).
 39. L. I-K Lin, "A note on the concordance correlation coefficient," *Biometrics* **56**, 324-325 (2000).
 40. H. Bao, A. Boussioutas, R. Jeremy, S. Russell, and M. Gu, "Second harmonic generation imaging via nonlinear endomicroscopy," *Opt. Express* **18**, 1255-1260 (2010).

1. Introduction

Chronic liver diseases are generally associated with extra-cellular matrix (ECM) over-production. Human liver fibrosis is characterized by an accumulation of fibrillar collagen as a result of wound healing processes in response to hepatic injury like chronic alcohol intake or hepatitis B and C [1–4]. It may evolve to cirrhosis and cancer with high morbidity and mortality rates. However liver fibrosis is reversible, and accurate diagnosis is vital to the management of patients. Noninvasive diagnosis such as blood and urinary tests [5,6], magnetic resonance imaging [7] and ultrasonic elastography [8] provide first indication to the disease prominence, and liver needle biopsy is still considered as the gold standard method for a precise evaluation of hepatic fibrosis [9], despite sampling variability [10]. However the various numerical scoring systems like Scheuer [11], Ishak [12] or Metavir [13] scores based on histological examination of ECM-stained biopsies by trained pathologists remain qualitative or semi-quantitative, and are subject to intra and interobserver variations. Automatized analysis systems related on image segmentation such as FibroQuant [14], Bioquant Nova Prime [15] and FibroXact [16] provide a continuum in fibrosis quantification but the results are affected by change in staining with protocols, time and illumination, and lack of internal reference. Taking advantage of the inherent localization of nonlinear excitation at the objective focal volume, multiphoton excitation fluorescence and multi-harmonic microscopies provide intrinsic optical sectioning and in-depth imaging, while dramatically reducing out-of-focus photobleaching and phototoxicity [17–21]. Moreover, isotype I and III fibrillar collagens, the main components involved in liver fibrosis, can be selectively imaged within tissues using second harmonic generation (SHG) microscopy, avoiding staining drawbacks. SHG microscopy has revealed its high potential for imaging and analyzing fibrillar collagen in unstained tissues [22–30]. SHG microscopy also becomes an interesting approach in the assessment of diseases that induce a modification of collagen content and/or structure in the ECM, in particular fibroproliferative diseases [31–35].

We recently used a new collagen scoring method in assessing human liver fibrosis [35] (see also the editorial [36]). The aim of our scoring method was to propose a simple, robust and rapid method of evaluation of fibrosis extension that could be implemented and used in a straightforward manner in every biomedical laboratory. In order to limit the number of images to be taken to sample the whole biopsy volume, it is necessary to use large field of view and low resolution microscope lenses, ie low NA lenses. For instance, the PSF volume of a 10x-NA0.25 lens is ~400 times larger than that of a 60x-W-NA1.2 lens, thus the experimental imaging time and memory space required to probe a given tissue volume with the latter lens is typically 400 larger. Such a large amount of information to resolve the ultrastructure of the collagen deposits in fibrotic tissues is usually not relevant from the diagnosis point of view, which requires to sum up the fibrotic collagen amount with a single figure. Moreover, it is well known that in the case of high NA objectives, complicated effects occur, due to phase matching effects (Gouy phase shift, non paraxial rays,... [24, 37]). Thus, this simple scoring procedure, that we develop in the

present paper, is based on using low NA and large field of view lenses, to average SHG signals over many collagen clusters. In fact, we will show in the following that using low-NA lenses permits to compare the scores between different setups.

In addition, our scoring procedure relies on using a reference sample to set the overall gain of the imaging system so as to obtain a given mean SHG intensity on nonfibrotic areas. All the samples with different fibrosis grades were then imaged with this calibrated experimental setup (in particular laser excitation power, PMT gain,...). The corresponding SHG images were binarized by using an intensity threshold t . Practically, the score was obtained by transforming the intensity image to a binary image by replacing each pixel intensity-value I by 1 if $I > t$ or 0 if $I < t$, and by calculating the mean number of pixels with a value of 1 in the resulting binary image. Thus, the score $\bar{S}(t)$ can be written as

$$\bar{S}(t) = \frac{1}{N^2} \sum_{i,j=1}^N H(I_{ij} - t) \quad (1)$$

where I_{ij} is the intensity measured at pixel ij , N^2 is the total number of pixels of the image, t is the intensity threshold, and $H(x)$ is the Heaviside function ($H(x) = 1$ when $x > 0$ and 0 if $x \leq 0$).

With our choice of the reference SHG intensity, almost all the SHG images showed saturated areas, typically in regions of high collagen content, whereas nonsaturated regions ($I < I_{sat}$) typically corresponded to nonfibrotic regions of human liver. In fact, the images are almost already binarized by the acquisition procedure. This is illustrated in Fig. 1, for a F1-Metavir graded patient (fibrosis without septa). In Fig. 1(a), the PMT gain was set to a value g_1 leading to almost no saturation of the imaging system, as exemplified in the intensity histogram of Fig. 1(d). Note that this histogram is almost structureless, with no intensity peaks, since it is a monotonic decreasing function of the intensity. To mimic saturation, the same image was multiplied by 4.7, and all pixels values above the saturation intensity were set to this value I_{sat} . This image Fig. 1(b) has the same contrast as a real saturated image obtained from another acquisition with the PMT gain set to $4.7 \times g_1$, Fig. 1(c). The histogram of Fig. 1(c) superimposes almost perfectly with the unsaturated image's histogram, as well as the histogram of an image acquired with a smaller PMT gain set $2.15 \times g_1$, indicating that the saturation procedure does not modify the low intensity histogram values.

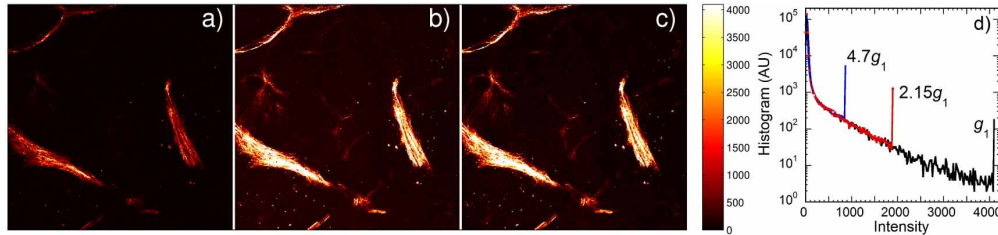


Fig. 1. (Color inline) SHG images acquired with a 4x-objective of a F1 biopsies. (a) PMT gain g_1 was set to obtain an image with almost no saturated pixels ($3.5 \times 3.5 \text{ mm}^2$); (b) Same image as in a), but multiplied by 4.7 and shown with a scale limited to the 12-bits range [0 4095]; (c) Experimental image with a PMT gain set to $4.7 \times g_1$; (d) Comparison of image intensity histograms for three PMT gains g_1 , $2.15 \times g_1$, and $4.7 \times g_1$.

However, in our experience, it is very difficult to find an experimental setup that leads to unsaturated images. As shown in Fig. 1(d), the absence of intensity peaks in the intensity histogram prevents from obtaining informations on fibrotic or nonfibrotic regions directly from the

shape of the intensity histogram. Moreover, we observed that the PMT gain should be varied within a range of 1 to 20 to obtain unsaturated images over all histological sections from non fibrotic to cirrhotic livers. This large extend of gains, and the difficulty to set a PMT gain by trial and errors to obtain a non saturated image is a critical point, because the intensity threshold needed to calculate the score from Eq. (1) should be modified for each experimental setup, as soon as the intensity level changes, to perform the segmentation of fibrotic regions from nonfibrotic regions with the same accuracy. Therefore, the difficulty of finding an intensity reference level, needed to be able to compare different scores, prompted us to develop the new method of scoring described in the beginning of the introduction. Although it represents a loss of information on regions of high collagen content, it is the cost to be paid to be able to set a reference intensity allowing direct comparison between scores.

The aim of this article is to give an in-depth analysis of the SHG imaging modalities used to work out our SHG scoring method of collagen in liver fibrosis, and to present new results that adress the influence of resolution and spatial sampling to the robustness of the score. First, we introduce, for the first time, a simple model to rationalize the SHG intensity as a function of the different experimental parameters and collagen content in human liver. Second, this model is validated from experimental measurements using careful characterization of the experimental setup. And third we show that the intensity generated by fibrillar collagen in human liver can be quantified rather precisely, and compared from one objective to another for low NA values.

2. Experimental methods

2.1. Sample preparation

Large surgical biopsies were collected from 12 patients regardless of fibrosis origin. Patients presented chronic liver disease with various degrees of severity related to excessive alcohol intake, hepatitis B or C viruses. Criterion for adequacy of the biopsy specimens was only low steatosis level ($< 5\%$). For the Metavir assessment, tissue samples were fixed in formaldehyde 4%, embedded in paraffin, and sections stained with Sirius red. Liver histological status was assessed by a pathologist using the Fibrosis-Metavir scoring system. Fibrosis was staged on a scale from F0 to F4 : F0 = no Fibrosis, F1 = fibrosis without septa, F2 = few septa, F3 = numerous septa without cirrhosis, and F4 = cirrhosis. The necroinflammatory activity was not considered in this study. For SHG scoring, paraffin embedded acute sections from selected patients were performed and deposited on a glass slide. Due to the scarcity of the biological material obtained from surgical and needle biopsies of human patients, it was not possible to obtain samples of larger thickness than $50\ \mu\text{m}$. Thus, in order to study the effect of sample thickness and imaging depth, we mainly used $50\ \mu\text{m}$ thick samples, and some $5\ \mu\text{m}$ or $15\ \mu\text{m}$ thick samples. Multiphoton imaging by SHG and TPEF was done without dewaxing, staining and coverslide use.

2.2. Experimental setup and imaging conditions

Our imaging setup was based on a modified confocal microscope composed of an Olympus IX71 inverted stand and a FluoView 300 scanning head (Olympus, Hamburg, Germany). A femtosecond Ti:sapphire laser (Mira900-Verdi5 combination, Coherent, Saclay, France) was coupled to the microscope and was tuned at a wavelength of 810 nm for all experiments. The quarter waveplate is settled after the last folding mirror of the microscope, just before the entrance pupil of the objective. This configuration avoids significant ellipticity produced by mirror dichroism and birefringence, the polarisation at the exit of this mirror being almost perfectly linear (less than 1% of ellipticity) [26–28], and not modified during experiments. Thus, circularly polarized 200-fs pulses at a repetition rate of 76 MHz were sent to two different microscope objectives (Table 1).

Table 1. ⁽¹⁾ : Estimation of the lateral and axial $1/e$ radii of the diffraction-limited PSF^2 of the two objective lenses used in the experiments, assuming a Gaussian shape, from ref. [20].
⁽²⁾ Measured value. Excitation wavelength=810nm; FOV=Field of View of our microscope.

Obj	$w_{xy}^{(1)}$ (μm)	$w_z^{(1)}$ (μm)	$w_z^{(2)}$ (μm)	FOV (μm)
10X Plan C-NA 0.25	0.73	9.7	$13.5 \pm .3$	1410
4x Plan CN-NA 0.1	1.83	60.8	65 ± 5	3525

The SHG light was collected in the forward direction by a 0.3 NA condenser (IX-ULWCD, Olympus). The illumination wavelength was blocked by a 2-mm thick BG39 filter (Lambda Research Optics, CA). The SHG was detected through a 405-nm (10-nm full width half-maximum) bandpass filter (Edmund Optics, York, UK) by a PMT module Hamamatsu H7844, with thermoelectric cooler. The PMT assembly was connected to a transimpedance amplifier (C7319, Hamamatsu) so as to match the SHG detection to the internal PMTs of the microscope and to use the full range of Fluoview hardware and software. SHG images were acquired with 12-bit intensity resolution and recorded as Tiff files using the microscope software (Fluoview; Olympus). The slowest scanning speed was selected, corresponding to a fixed pixel dwell-time of about 10 μs or a total acquisition time of 2.68 sec per frame for a full-field image in a 512×512 frame. To cover the same field of view as the 4x-lens with the 10x-lens, nine connexe (512×512) images were automatically acquired with an XY-translation stage (Märzhäuser, Wetzlar, Germany). A montage of the images was then performed, and truncated to the same region of interest to allow comparison.

Table 2. Data of the experimental set up. ⁽¹⁾ : Average laser power at the *focal plane* measured with a calibrated power-meter OPHIR PD300; ⁽²⁾ PMT supplied voltage used for experiments performed with 4% SHG mean intensity from the basal collagen of the ECM.

Obj	Avg. Laser Power ⁽¹⁾ (mW)	VPM (V) 4% ⁽²⁾
10X Plan C-NA 0.25	51.2	798
4x Plan CN-NA 0.1	106.6	726

We carefully measured the PMT gain as a function of the high voltage supply, and obtained $g_{PMT} \sim (V_{PMT})^\alpha$ with an exponent $\alpha=8$, in agreement with the exponent derived from the constructor's data sheet of the PMT. The reproducibility of the gain settings was excellent. Thus, we choose to modify the intensity of the images by adjusting the PMT voltage supply at *constant laser excitation power*, allowing direct comparisons of SHG intensity values.

The z-profile of the objective two-photon Point Spread Function (PSF^2) was measured using two-photon excited fluorescence (TPEF) of a thin layer of an aqueous solution of 10^{-4} mol/l sulforhodamine B dye (kitonred620, *Exciton*, Daton, OH, USA). The axial $1/e$ extend w_z of the diffraction-limited PSF^2 was estimated by least-square fitting assuming a Gaussian shape [20]. Important data of the experimental setup are summarized in Tables 1 and 2. The thickness of the samples was verified from z-stack images acquired with the 10x-NA0.25 objective. The intensity of the z-profile was least-square fitted with the following formula $p(z) = \frac{1}{2}A [erf((z - z_i)/w_z) + erf((z_f - z)/w_z)]$, which is valid for a uniform layer of thickness e lying between z_i and $z_f = z_i + e$, and a gaussian PSF^2 profile. A is an intensity and erf the *Error Function*. w_z was fixed to its experimental value. The size of the objective PSF^2 and sample typical profiles are compared on Fig. 2(a),(b). For the 4x lens, the PSF^2 was carefully settled at the center of the sample.

Although the axial z-sectioning of the 10x-objective is small enough to perform depth studies

of the samples (see for instance the z-profiles in Fig. 2(b)), we will only compare the 4x images to the 10x ones acquired when the objective geometrical focus was at the center of the sample. In fact, z-stack studies of different samples, as for instance the different images shown in Fig. 2(c), indicate that at the scale of the thickness sample ($\sim 50\mu\text{m}$), the collagen structures are almost invariant as a function of depth. This is not astonishing because most of the structures have a radius of gyration much larger than the sample thickness.

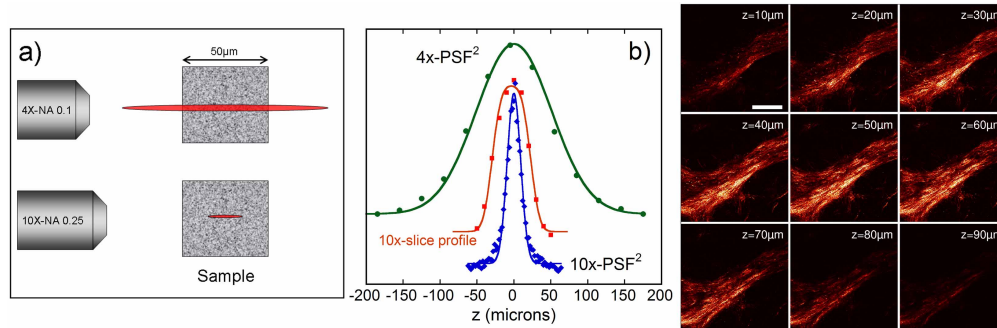


Fig. 2. (Color inline) (a) Schematic representation of the PSF^2 of the objectives ($1/e$ contour lines) as compared to the sample thickness ; (b) Experimental z-profiles of the PSF^2 of the 4x(●) and 10x(◆) objectives. Intensity z-profile (■) of a $50\mu\text{m}$ thick sample reconstructed from slices acquired with the 10x objective. Continuous lines are fits as explained in the text. (c) Typical images of a z-stack recorded with the 10x objective (scale bar= $400\mu\text{m}$).

Image processing and analysis (thresholding and SHG scoring) were performed with home-made routines programmed in Matlab (The MathWorks, Natick, MA).

3. Experimental results and discussion

In a previous study [35], we showed that our scoring method was able to assess fibrotic-cirrhotic stage of human liver using a 4x-NA 0.1 objective, in good correlation with Metavir grade. However, different questions concerning the robustness of our method to different experimental parameters may arise. In particular the NA of the objective has dramatic effects on the sampling volume, as schematized by Fig. 2(a), where the axial extent of the PSF^2 may be much larger or much smaller than the sample thickness. In this section, we show that our SHG scoring method can be rationalized by a simple model of SHG intensity. After proving that this simple model is supported by experiments carried with the two different objectives (table.1), we show that the scores can be indeed compared, despite the differences in the spatial lateral sampling rate or volume.

3.1. SHG intensity and collagen concentration

Estimation of the influence of the objective parameters (NA, \dots) and harmonophore concentration on the real SHG intensity is complicated by the fact that SHG is a coherent process (the word “harmonophore” is generally used to distinguish molecules or supramolecules that provide harmonic generation of light from “fluorophores” that emit fluorescence). Nevertheless, all the liver images show a strong positional and orientational disorder of the collagen clusters. Moreover, the low-NA objectives have PSF^2 axial size much larger than typical wavelengths ($w_z \gg \lambda$, see Table.1). Therefore, we expect the intensity to be the incoherent summation of coherent signals produced by very small clusters of collagen. For m_N clusters of N harmonophores

that scatter coherently, the intensity is proportional to $m_N N^2$. The number of clusters m_N is proportional to the PSF^2 volume $m_N \propto \Omega$. For a distribution of clusters scattering incoherently, the resulting intensity is a weighted sum of such contributions. With such an hypothesis, the overall intensity is still proportional to Ω , and is expected to behave like the $TPEF$ intensity as a function of NA . We will verify *a posteriori* this hypothesis from experimental results. The SHG intensity can be written as

$$I_{SHG} \propto g_{PMT} T^2 I_o^2 \int d^3 r P(r) \bar{C} \quad (2)$$

where g_{PMT} is the PMT gain, T is the overall transmission of the optical setup, \bar{C} an effective cluster concentration, $P(r) = \exp(-(x^2 + y^2)/w_{xy}^2) \exp(-z^2/w_z^2)$ is the PSF^2 assuming a gaussian approximation which was shown to be valid for two-photon excitation [20], and I_o the intensity at the focal plane.

In the paraxial approximation, $w_{xy} \sim NA^{-1}$ and $w_z \sim NA^{-2}$. Assuming in a first step that the sample thickness is much larger than the PSF^2 axial extend, the PSF^2 volume is $\Omega \sim w_{xy}^2 w_z \sim NA^{-4}$, and the intensity $I_o \sim NA^2$. Therefore, $I_{SHG} \sim NA^{-4} (NA^2)^2$ is independent of the NA .

This is valid for the 10x objective, but for the 4x objective, the sample thickness e is smaller than the PSF^2 axial extend ($e < w_z$), and only part of the PSF^2 is used. The effective volume $f\Omega$ probed by the PSF^2 is then

$$f\Omega \sim NA^{-2} \int_{-e/2}^{e/2} \exp(-z^2/w_z^2) dz \sim NA^{-2} w_z \operatorname{erf}\left(\frac{e}{2w_z}\right) \sim NA^{-4} \operatorname{erf}\left(\frac{e}{2w_z}\right) \quad (3)$$

with $f = \operatorname{erf}(\frac{e}{2w_z})$ a geometrical factor. Note that $\operatorname{erf}(x) \rightarrow 1$ when $x \rightarrow \infty$, that is when $e \gg w_z$.

Therefore, for each objective, the intensity can be written as

$$I_{SHG} \propto g_{PMT} T^2 f \bar{C} \quad (4)$$

Under such approximations, the SHG intensity depends on NA only through the geometrical factor f , and is proportional to an effective cluster concentration. By using a reference sample (that is a non fibrotic sample with Metavir grade F0), it is then possible to set the PMT gain by adjusting the supply voltage to obtain an average intensity $\langle I_{SHG} \rangle$ over a given ROI that is a fixed percentage r of the full intensity range corresponding to the saturation intensity I_{sat} (which is 4095 for our 12-bit system). We will call r the reference intensity. Thus, $\langle I_{SHG} \rangle = r I_{sat} = g_{ref} T^2 f \bar{C}_{ref}$, and the intensity for other collagen concentrations, at the same PMT gain g_{ref} , is expected to follow

$$I_{SHG} = r I_{sat} \frac{\bar{C}}{\bar{C}_{ref}} \quad (5)$$

The normalized intensity is proportional to both the effective collagen concentration \bar{C} and the reference value r .

It is important to note that this model assumes that the PSF^2 size is large enough to average the signal over many collagen clusters. It is thus expected to be valid only for low NA lenses. The question of how low the NA should be is difficult to answer, but we remark that the volume of the PSF^2 of a $NA=0.25$ lens is about $\approx 60 \mu m^3$. Since this volume $\sim NA^{-4}$, it is divided by ≈ 2 when $NA=0.3$ and by ≈ 6.6 when $NA=0.4$. In order to keep a large sampling volume, we do not expect our model to be valid for $NA \geq 0.3 - 0.35$. In the following, we verify experimentally the hypothesis underlying the simple model presented in this section, and compare the collagen scores obtained with the two objectives.

3.2. Using nonfibrotic ECM regions to set the reference intensity

We remarked that the SHG signal arising from the basal contribution of the ECM fibrillar collagen present in liver is almost always the same [35], for regions outside portal areas, whatever the Metavir grade. In this section, we show that these nonfibrotic regions can be used as internal reference to calibrate the experimental setup. Typical nonfibrotic ECM ROIs of a F4 sample are shown in Fig. 3(a).

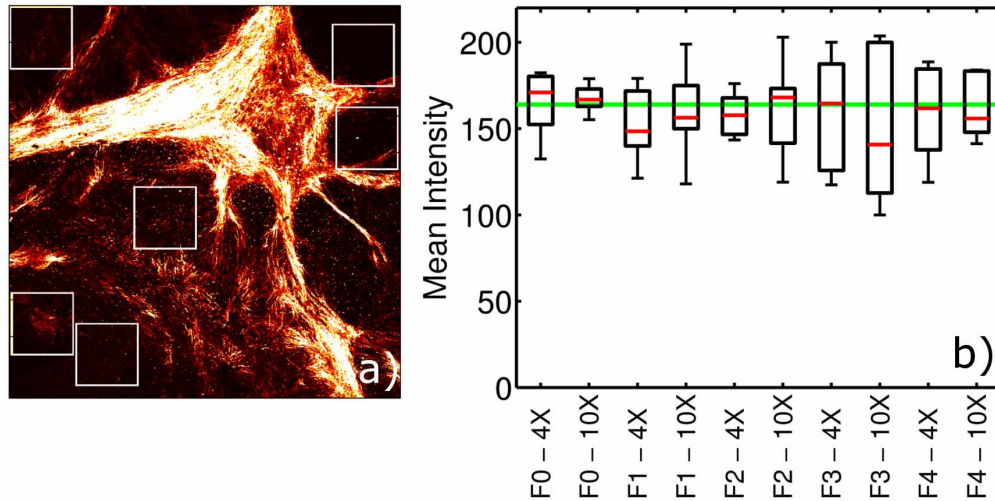


Fig. 3. a) Example of non fibrotic ECM on a F4 sample, delineated by white boxes ; b) Boxplot of mean intensities of different non-fibrotic ECM ROIs selected from different Metavir-graded samples. The PMT gain is the same for all samples, and was set to obtain $r=4\%$ ($\langle I_{SHG} \rangle = 164$, horizontal line) on a F0 sample with the 4x objective.

The PMT reference gain was set from a non-fibrotic sample (F0-Metavir grade) to obtain an average SHG signal of $r=4\%$ of the full intensity range (that is 164). Then, using the same reference PMT gain, images of samples from patients with different Metavir grades F1-F4 were acquired. Six ROIs of 80×80 pixels were selected on each image, and the mean intensity was calculated. The dispersion of these mean intensities are summarized in a boxplot on Fig. 3(b). The mean or median of the intensities, for the two objectives, are always very close to the expected value r . It is a strong indication that nonfibrotic ECM ROIs have an almost uniform collagen content. Although it clearly shows that such ECM regions can be used to set a reference PMT gain, we would however recommend in practice to set the reference intensity by using a F0 sample. In the following, our SHG microscope setup will be always calibrated on the same F0 sample.

3.3. Comparing 4x and 10x intensities using ECM references

It is now possible to check the validity of Eqs. (4) and (5). Using regions of interests containing only normal ECM fibrillar collagen, the PMT supply voltage was adjusted to obtain an average intensity of 2,3,4, and 5% of the full intensity range. Larger values $\geq 6\%$ introduce too much saturation, and were considered as unreliable. The gain is calculated from $g_{PMT} = (V_{PMT}/1000)^8$. Typical images are shown on Fig. 4(a) and (b).

Assuming that Eq. (5) is valid, and that the effective concentration of the ECM collagen \bar{C}_{ref} does not depend on the experiment as indicated by Fig. 3(b), we expect that the PMT gain should be a linear function of r . This is verified, as indicated by Fig. 4(c), where both gains g_{PMT}^{4x}

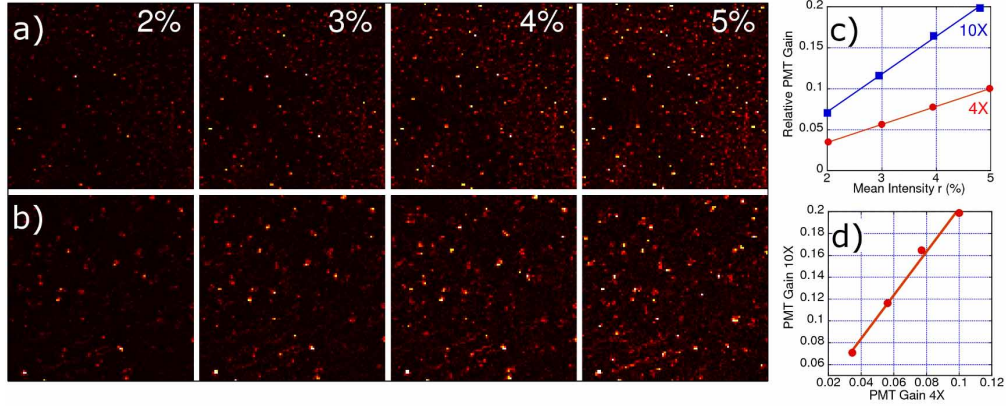


Fig. 4. Images of non-fibrotic ECM ROIs for different PMT gains corresponding to $r = 2, 3, 4, 5\%$: (a) for the 4x objective lens ; (b) 10x objective lens. c) Corresponding PMT gains ; (d) Correlation between the PMT gains of the 10x and 4x objective lens.

and g_{PMT}^{10x} vary linearly with r (linear correlation with $R^2 > 0.998$). More generally, for each r value, the intensity histograms from the 4x and 10x objectives were almost comparable. This is an indication that on such non fibrotic ECM regions, the collagen is distributed uniformly at the two different scales probed by the objectives.

Moreover, using Eq. (4), the PMT gains measured with the two objectives should follow the relationship

$$g_{PMT}^{10x} = \left(\frac{T_{4x}}{T_{10x}} \right)^2 f_{4x} g_{PMT}^{4x} \quad (6)$$

From Tables. 1 and 2, the geometrical factor of the 4x objective is $f_{4x} \approx 0.41$, and we found that $\left(\frac{T_{4x}}{T_{10x}} \right)^2 \approx 4.33$, thus our model predicts $g_{PMT}^{10x} \approx 1.8 g_{PMT}^{4x}$. As indicated by Fig. 4(d), the correlation between the two gains is linear ($R^2 \approx 0.992$), with a slope ≈ 1.95 , which value is slightly higher than the predicted value. Many reasons may explain the discrepancy : (i) the concentration of collagen may be slightly different from one region to another because the 4x and 10x do not probe the same volume; (ii) the collection of the SHG signal may differ from one objective to another due to the difference in NA values; (iii) the geometric correction assuming a gaussian profile is only an approximation; (iv) the real PMT gain may have an error due to the strong voltage dependence; (v) coherence effects, that were neglected by our model, may introduce a dependence on NA. Despite the small discrepancy between the expected value derived from the model and the experiment, the data are well linearly correlated, supporting the assumptions used to derive our model.

With the same methodology, we checked the influence of the geometric factor f appearing in Eq. (3), using three histological serial sections of thicknesses 5, 15 and $50 \mu\text{m}$ respectively, obtained from surgical biopsies of patients with two metavir grades F3 and F4. An example of SHG images of such surgical biopsies is shown in Fig. 5(a-c) for the F4 grade patient. Again, for each sample, the gain of the PMT was adjusted on normal ECM to obtain an average intensity of $r=4\%$. As expected from Eq. (4), g_{PMT} should be proportional to $1/f$, $g_{PMT} \propto 1/f$. As indicated by Fig. 5(d), this is verified for both metavir grades F3 and F4, since the PMT gains show a good linear correlation with the geometric factor $1/f$ ($R^2 = 0.997$ and 0.9944 respectively). These results demonstrate the pertinence of Eq. (4) to account for the SHG intensity from human liver.

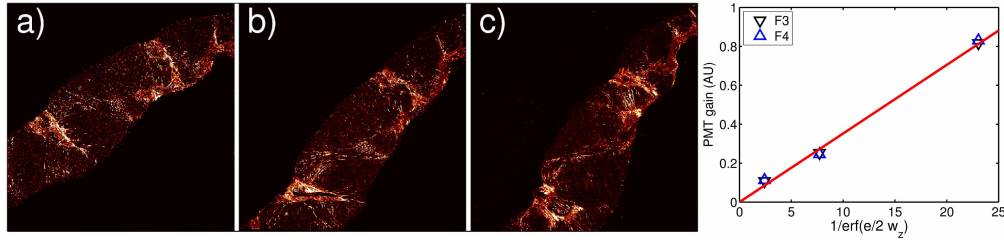


Fig. 5. SHG images of histological serial sections of a surgical biopsy of a F4 patient, with thicknesses : (a) $5\mu\text{m}$, (b) $15\mu\text{m}$, (c) $50\mu\text{m}$. d) Correlation between the PMT gain and the inverse of the geometrical factor f (Eq. (3)).

3.4. Comparison of 4x and 10x scoring, and discussion

In the previous sections, we showed that our model of SHG intensity, summarized by Eq. (4) and 5, well describes our experimental findings in human liver biopsies. The last step is thus to compare the values of the scores, as defined by Eq. (1), of the same ROIs, with different reference intensities and objective lenses.

In evaluating the score $\bar{S}(t_r)$, we select intensities such that $I > t_r$ where t_r is a given threshold, which may be optimized for a given reference intensity r . Then, from Eq. (5), we expect that if the reference intensity is changed to a different value r' , then the new score $\bar{S}(t_{r'})$ obtained with the new threshold $t_{r'} = t_r r' / r$ should be identical to $\bar{S}(t_r)$. Clearly, a direct consequence of Eq. (5) is that the score as defined by Eq. (1) is only a function of t_r / r .

To verify this relationship, SHG images of the large surgical biopsies at our disposal were acquired with the two objectives, with intensity references set to $r = 2, 3, 4$ and 5% . For the 10x lens, we choose the images of the stack that correspond to the center of the sample, and reconstruct an image that covers the same sample area as the one obtained from the 4x lens. Typical images for two different Metavir grades F1 and F4 are presented in Fig. 6(a),(b). Although the spatial sampling is different (1300^2 and 512^2 pixels for the 10x and 4x images respectively), the images appear quite similar. This indicates that collagen aggregates are rather homogeneous at the scale of the sample thickness ($\sim 50\mu\text{m}$), which should be expected because the lateral size of such collagen structures is typically of a few hundreds of microns. Therefore, it is not astonishing that the projected area of collagen aggregates obtained from scoring with 4x or 10x objectives are so similar.

After reconstruction, the scores were calculated with threshold $t_r = 4 t_4 / r$, where $t_4 = 1024$ is the threshold used for $r = 4\%$. It was shown in [35] that this threshold value give a good correlation between our scoring and the Metavir assessment.

The correlation between the scores was quantified by both the Pearson's correlation coefficient ρ and the Lin's concordance correlation coefficient ρ_L [38,39]. Roughly, the Pearson's R measures how two quantities are linearly correlated ($Y = aX + b + \epsilon$ where ϵ is a noise), that is the more linearly correlated are the data, the closest to one is $|R|$, whatever the values of the slope a and intercept b . On the other hand, the Lin's concordance correlation coefficient measures the vertical departure from the bisector, therefore ρ_L tests whether $Y = X + \epsilon$, which is a linear correlation with slope 1 and zero intercept. Therefore, ρ_L should be more appropriate to quantify whether two scores are equal. It can be shown that $\rho_L \leq R$, which can be understood by the fact that the general linear correlation leaves more flexibility to correlate the data than the more stringent condition that the data should align along the bisector.

On Figs. 7(a) and (b) are presented the correlations, for a given objective lens, between the score obtained when $r = 4\%$ and those obtained at different r values. The experimental

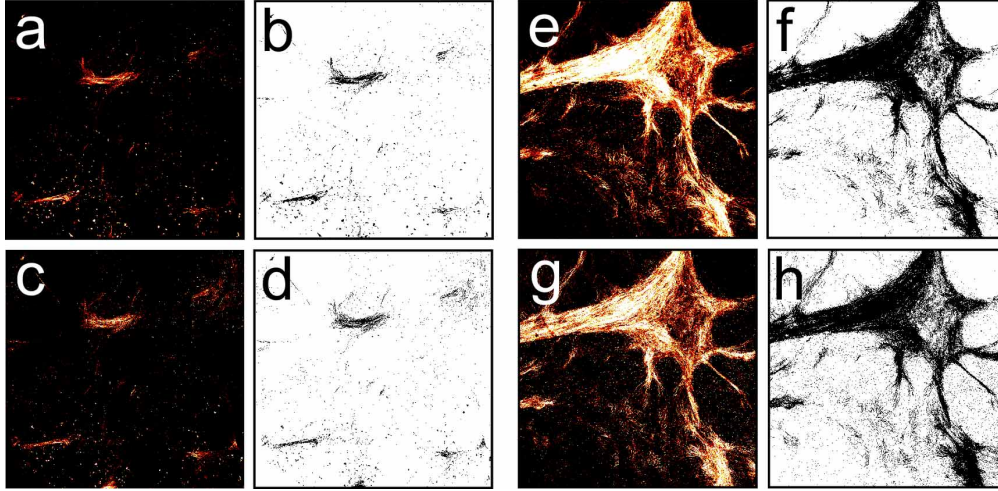


Fig. 6. Example of 4x (first line) and 10x (second line) images when $r=4\%$, for the same field of view $3525 \times 3525 \mu\text{m}^2$ (4x : 512×512 pixels; 10x : 1300×1300 pixels). (a,b,c,d) F1 Metavir grade. (e,f,g,h) F4 Metavir grade. (a,b,e,f) 4x-obj. (c,d,g,h) 10x-obj

values almost align along the bisector, and both Pearson's R or Lin's concordance correlation coefficient ρ_L are, as indicated by table.3, very close to one, indicating a very good concordance. These data clearly support the model presented in the beginning of the article.

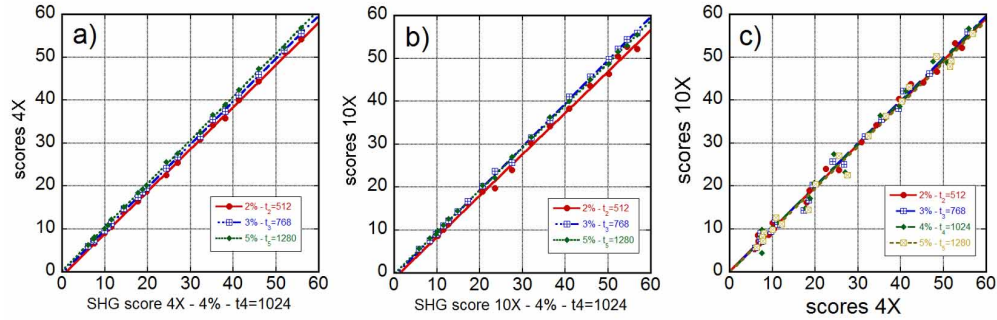


Fig. 7. Comparison of the scores obtained at $r=2,3,5\%$ with the score at $r=4\%$ for : (a) the 4x lens ; (b) the 10x lens . (c) Comparison of 4x and 10x scores.

We also compared the scores obtained for a given r with the two objectives, and using an identical area as indicated before, in Fig. 7(c). Again, the scores obtained from the 4x and 10x objectives align on the bisector for all reference values r , with good concordance correlation coefficients (see Table.3 for the values of R and ρ_L). This figure shows that the scores are almost identical for the two objectives, despite the difference in NA , and thus in axial and lateral resolutions, as well as in lateral spatial sampling rate (512 or 1300 pixels for the same scale).

4. Conclusion

In this article, we presented an in-depth study of a scoring method that we developed to quantify collagen in human liver fibrosis, based on SHG microscopy. This scoring method is based on

Table 3. Slope from a least-square linear fit $Y = a \times X$ and Pearson's linear correlation coefficient R between different scores, and the sample concordance correlation coefficient ρ_L proposed by Lin (1989) [38].

r t_r	2%	3%	4%	5%
	512	768	1024	1280
4x-4%/4x-r% : a	0.963	.994		1.02
: R	0.9996	0.9999		0.9998
4x-4%/4x-r% : ρ_L	0.9980	0.9998		0.9989
10x-4%/10x-r% : a	.936	.989		0.98
: R	0.9986	0.9997		0.9998
10x-4%/10x-r% : ρ_L	0.9904	0.9993		0.9990
4x-r%/10x-r% : a	0.990	0.996	0.992	0.976
: R	0.9977	0.9964	0.9932	0.9939
4x-r%/10x-r% : ρ_L	0.9975	0.9963	0.9932	0.9930

the acquisition of SHG images with a given setup calibrated from a reference sample, and by counting the mean number of pixels above a given threshold. Once set on a given sample, the imaging settings are kept for all further acquisitions. Although developed for SHG imaging, the scoring method described in this article may also be applied, to some extent, to Two-Photon-Excitation-Fluorescence (TPEF) imaging. In fact, the SHG model presented below is incoherent, and applies directly to TPEF imaging. On the one hand, since our scoring method uses saturated images, it rules out the quantification of fluorophores concentrations. But on the other hand, it may give a method to measure the surface of the image occupied by fluorophores which intensity is above a given threshold, on the condition that a reference intensity can be used to set the gains of the experimental setup.

One of the originality of our scoring method comes from the fact that the choice of the reference intensity leads to the acquisition of SHG images with large regions of saturated values, while keeping the nonfibrotic ECM regions within the intensity range of the imaging system. Combined with a scoring based on thresholding the images, this method gives high reproducibility since the threshold is fixed as soon as an imaging setup is chosen. This is in contrast to conventional automatized scoring programs or SHG scoring [34], that need to adapt the segmentation threshold to each image, depending on staining process, illumination, camera properties, leading to uncertainties in the definition of fibrotic and nonfibrotic regions.

We showed that an internal reference can be obtained by using the nonfibrotic ECM areas of any sample. By carefully characterizing the experimental setup, and using two different objective lenses, we validated experimentally a simple model of the measured SHG intensity. Although simple, this model accounts well for the different behaviors of the observed SHG intensity in human liver, at least in the limit of low NA. Thus, the effect of different sampling volume (NA), sample thickness and reference intensity on the SHG intensity was explained and quantified. We were thus able to show that our scoring method is robust to differences in NA and spatial sampling, with a very high concordance of the score values between our 4x-NA 0.1 and 10x-NA 0.25 objectives.

SHG scoring methods presents multiple advantages in comparison to the histologic numerical systems. In particular, it does not require dewaxing or specific staining steps, and SHG imaging displays a high SHG signal-to-background ratio making a specific detection of fibrillar collagen very easy. To sum up, the main advantages of our scoring method which relies on using low-NA and large field of view lenses, as compared to the methods presented in [31, 32, 34]

come from (i) the simplicity and speed of acquisition and treatment (typically less than 10 images permit to score a complete histological slice. No image processing, except binarisation of the raw image data, is needed.) ; (ii) Robustness of the collagen scoring to changes in the objective features, spatial sampling, sample thickness, detection settings. (iii) Reliability and Portability : our scoring method uses a reference (collagen from non fibrotic liver extracellular matrix) to calibrate the setup. This study shows that our SHG collagen scoring method can be highly standardized, and provides a unique score independently of operators, or experimental conditions. Moreover, we proposed a simple model which could be used to compare results obtained from different experimental setups. Therefore, we expect that SHG scoring will give the opportunity to quantify fibrosis regression in response to an antifibrotic treatment for validating experimental protocols especially for cirrhotic patients. Although only validated for human liver fibrosis, we sense that our scoring method can also be applied to other fibrosis in lung, kidney... The simplicity of our scoring method also makes it a good candidate to be implemented with nonlinear endomicroscopes [40].

Acknowledgments

This work was supported by Région Bretagne and Rennes Métropole, CNRS project " Interface Physique-Chimie-Biologie : soutien à la prise de risque ", and CPER PONANT. T. Guilbert has a grant from the "Conseil Régional de Bretagne" for his PhD thesis.