Measurement of muscle disease by quantitative second-harmonic generation imaging

Sergey V. Plotnikov*
University of Connecticut Health Center
Department of Genetics and Developmental Biology and Center for Cell Analysis and Modeling
Farmington, Connecticut 06030

Anne M. Kenny
University of Connecticut Health Center
Center on Aging
Farmington, Connecticut 06030

Stephen J. Walsh
University of Connecticut Health Center
Center for Community Medicine and Health Care
Farmington, Connecticut 06030

Beata Zubrowski
Cherian Joseph
University of Connecticut Health Center
Center on Aging
Farmington, Connecticut 06030

Victoria L. Scranton
University of Connecticut Health Center
Department of Genetics and Developmental Biology
Farmington, Connecticut 06030

George A. Kuchel
University of Connecticut Health Center
Center on Aging
Farmington, Connecticut 06030

Deborah Dauser
University of Connecticut Health Center
Center for Community Medicine and Health Care
Farmington, Connecticut 06030

Manshan Xu
Carol C. Pilbeam
University of Connecticut Health Center
Department of Medicine
Farmington, Connecticut 06030

Douglas J. Adams
University of Connecticut Health Center
Department of Orthopedic Surgery
Farmington, Connecticut 06030

Robert P. Dougherty
OptiNav, Inc.
10900 NE 8th Street, Suite 900
Bellevue, Washington 98004

*Current address: Laboratory of Cell and Tissue Morphodynamics, NHLBI, 50 South Drive, Building 50, MSC-8019, Bethesda, Maryland 20892-8019. Address all correspondence to: William A. Mohler, Univ. of Connecticut Health Center, 263 Farmington Ave., Farmington, CN 06030; E-mail: wmohler@neuron.uchc.edu.

Abstract. Determining the health of muscle cells by in vivo imaging could impact the diagnosis and monitoring of a large number of congenital and acquired muscular or cardiac disorders. However, currently used technologies are hampered by insufficient resolution, lack of specificity, or invasiveness. We have combined intrinsic optical second-harmonic generation from sarcomeric myosin with a novel mathematical treatment of striation pattern analysis, to obtain measures of muscle contractile integrity that correlate strongly with the neuromuscular health of mice suffering from genetic, acquired, and age-related decline in skeletal muscle function. Analysis of biopsies from a pilot group of human volunteers suggests a similar power in quantifying sarcopenic changes in muscle integrity. These results provide the first strong evidence that quantitative image analysis of sarcomere pattern can be correlated with physiological function, and they invite the application of SHG imaging in clinical practice, either in biopsy samples or via microendoscopy. © 2008 Society of Photo-Optical Instrumentation Engineers. [DOI: 10.1117/1.2967536]

Keywords: second-harmonic generation; pattern recognition; scanning microscopy; lasers in medicine; laser application.

Paper 07438R received Oct. 25, 2007; revised manuscript received Feb. 7, 2008; accepted for publication Feb. 27, 2008; published online Aug. 20, 2008.

1 Introduction

Muscular disorders (MDs) are a large heterogeneous group of diseases characterized by progressive weakening and degeneration of skeletal muscle tissue. Many forms of congenital MDs are now known to arise from identified single-gene defects. Molecular diagnosis can clarify medical decisions as to the course of treatment, long-term prognosis, and genetic counseling. Still other MDs are associated only secondarily with the onset of disease elsewhere in the patient, e.g., cachexia in cancer, atrophy due to injury or neuropathy, or sarcopenia of aging. For both genetic and nongenetic MDs, monitoring the course of disease and judging the success of treatment are most often approached through methods that are either low-resolution or invasive, without assessing the integrity of muscle fibers directly. Much attention is focused on the prospect of monitoring disease through microarray gene-
expression profiling. However, a recent effort to differentiate expression patterns of early and advanced human Duchenne Muscular Dystrophy showed a remarkable constancy in expression profiles, even as disease worsened over many years. Several different imaging modalities, such as quantitative computed tomography, dual energy X-ray absorptiometry, and magnetic resonance imaging, provide information about the geometry, mass, and tissue composition of bone and muscle. These noninvasive techniques are suitable for rough quantitative measurements of tissue degradation/repair. However, low spatial resolution prevents these methods from probing the microscopic features and molecular structure of skeletal muscle.

Sarcomeres of striated muscle produce the force that is ultimately responsible for contractile function. Expression and catalysis of sarcomeric proteins are well documented to respond to neuromuscular activity or disease, making the sarcomere pattern a prime candidate for a visible indicator of myofiber integrity. In fact, some published evidence correlates muscular dystrophy with changes in the sarcomere pattern of biopsy samples. However, quantitative microscopic assessment of muscle sarcomere damage has never been adopted as a routine diagnostic tool. The absence of appropriately specific, objective, and automated imaging techniques has been the major limitation for the development of such approaches. Traditional histochemical staining to reveal striations (e.g., eosin or Alizarin blue) lacks molecular specificity, as it also labels additional nonsarcomeric intracellular structures and organelles. Laser diffraction has been used to measure a globally averaged sarcomere length in muscle tissue, and has even been applied to live measurement of muscle extension during orthopedic surgery, but it does not allow for a micron-by-micron analysis of the contractile lattice within tissue. Observation of sarcomere pattern by either electron, immunofluorescence, or polarization microscopy has very limited applicability in the clinic: none of these imaging techniques can be used to image directly within thick-muscle biopsy samples, and obtaining statistically reliable data requires analysis of hundreds of histological sections. Moreover, no robust means for quantifying the differences between sarcomere patterns in healthy and diseased muscle—to yield objective scores of muscle fiber structure—has been described for any of these techniques.

Fig. 1 SHG imaging of normal and diseased muscle. Single optical sections of gastrocnemius muscle dissected from (a) 5-week-old control, (b) mdx−/−;utr−/+; and (c) mdx−/−;utr−/− animals, respectively. Yellow arrowheads mark gaps between myofibers; yellow and blue arrows label splits and ruptures, respectively, within the contractile lattice of myofibers; green arrows show myofiber areas with bright unstriated SHG. Scale = 20 μm. (Color online only.)

We present a novel approach to the problem of quantitative diagnostic muscle imaging, by combining a deep-sectioning, high-resolution optical imaging technique with mathematical and statistical analysis of the resulting intrinsic contrast images from muscle tissue. The myosin thick filaments of sarcomeres produce second-harmonic generation (SHG) when excited by a focused ultrafast-pulsed infrared laser, and the synchronized alignment of repeating sarcomeric A-bands in adjacent myofibrils yields a pattern of high-contrast striation within normal muscle fibers that is striking in its regularity of period and orientation. Furthermore, the method yields three-dimensional stacks comprising relatively large volumes of intact tissue, since the deep-sectioning power of SHG imaging allows image acquisition at a submicron resolution up to 600 μm into a specimen.

Here, we have assessed the potential of monitoring the progression of muscular disease by quantitative analysis of sarcomere striation patterns in SHG images from muscle samples. By applying the Helmholtz equation for wavenumber to collections of SHG images from groups of individuals, we correlated the distribution of lengths of sarcomeres to the severity of disease, comparing mouse or human muscle suffering atrophic, dystrophic, or sarcopenic decline to matched control specimens. Our data show that quantitative SHG imaging of sarcomeric myosin provides robust discrimination between muscle of healthy animals and those with even mild forms of MDs. Because SHG optics can be adapted by miniaturization for real-time in vivo imaging, this method should become valuable for minimally invasive monitoring of muscle structure during medical treatment.

2 Results

2.1 Quantification and Classification of Striation Patterns in SHG Images of Muscle

We used the SHG microscope to capture digital image volumes from skeletal muscle tissue from a range of normal and diseased sources. In healthy myofibers, bands of sarcomeric SHG are straight and evenly spaced, with each band lying virtually orthogonal to the long axis (axis of contraction) of the cell. Damaged cells display a range of visible deviations from this norm, sometimes localized to subdomains within a giant myofiber: hypercontraction or hyperextension of stria-
tion spacing; misorientation, skewing, or contortion of bands; diminution or complete loss of SHG intensity (Fig. 1). An individual region of contractile apparatus can be found to display any one or a combination of these defects. To measure these changes and to assign criteria for scoring normal and abnormal structure, we applied the Helmholtz equation for wavenumber to calculate the local striation spacing and angle of orientation, thus creating new image maps for each of these metrics, in addition to the primary map of SHG intensity [Figs. 2(a) and 2(b); see Section 4 for details of the algorithm].

The deep optical sectioning power of the SHG microscope produced multislice image stacks that typically comprised more than $5 \times 10^5 \mu m^2$ of tissue. Thus, the fine spatial resolution of this sarcomere pattern quantitation (SPQ) algorithm allowed automatic assessment, by three distinct numerical criteria, of each of the tens of thousands of sarcomeres within the group of muscle fibers that were sampled by a single image volume. Unfortunately, the nonlinear dependence of measured SHG intensity upon sample thickness deterred us from assigning reliable abnormal/normal cut-off levels to make diagnostic distinctions based on image brightness. However, by reference to the literature on the physiology and ultrastructure of mammalian sarcomeres, and by empirical fitting to control images, we defined thresholds for scoring of sarcomere pattern relative to controls was detected in biopsies of healthy muscles, were pattern. Even more strikingly, both mdx and mdx/utr (severe) dystrophic animals were absolutely discriminated from wild-type controls by a logistic regression.

2.2 Correlation of Sarcomere Pattern Scores with Damage Due to MDs in Mice

We tested the applicability of the SPQ scoring algorithm to detecting and quantifying disease in muscle. We imaged specimens of gastrocnemius muscle by SHG, comparing SHG image sets from control mice against SHG image data acquired from mice with each of three established models of muscular disorders: disuse-induced atrophy, hereditary muscular dystrophy, and aging (see descriptions of disease models in Section 4). The selected models each incur a characteristic range of damage within the full spectrum of phenotypic severity, allowing us to rate the sensitivity of SHG image analysis in detecting injury. Efforts to combine binary thresholds for both SPQ criteria (length and angle) to measure the abundance of normal versus abnormal sarcomeres provided rather reliable differentiation between healthy and diseased muscle samples (data not shown). However, we found that a detailed analysis focusing specifically on the distribution of sarcomere lengths within specimens produced even better distinctions. A consistent negative correlation between the severity of a disease and the mean sarcomere length was revealed for all disorder models we tested [Fig. 2(d)]. While sarcomere lengths spanned from 1.4 to 2.4 $\mu m$ in biopsies of healthy muscles, mild and severe forms of MDs showed a noticeable reduction in the range of sarcomere lengths and an increased fraction of hypercontracted sarcomeres with lengths below 1.6 $\mu m$ [Figs. 2(c) and 2(d)].

To employ these differences in sarcomere length as a diagnostic marker, we extracted three distinct characteristics of the distribution of sarcomere lengths for each sample volume: the overall mean length (ML), the fraction of all sarcomeres that were not hypercontracted below 1.6 $\mu m$ in length (FNH), and the mean length of the fraction of sarcomeres that were not hypercontracted (MNHL). Clear discrimination between healthy and disordered muscle could be achieved using either ML or FNH alone when control mice were compared to mdx/utr double mutants, which lack functional dystrophin and utrophin proteins and develop a severe form of Duchenne-like dystrophy [Fig. 2(d)]. Similarly, ML showed a significant difference between controls and animals subjected to atrophy by hindlimb suspension (HLS) [Fig. 2(d)], a protocol that decreased femur bone mineral density by an average of 12% and muscle fiber size by 13% (see Fig. 3). Consistent, but not statistically significant, alteration of sarcomere pattern relative to controls was detected in biopsies affected by other milder muscular disorders (mice with mdx single mutations lacking only dystrophin, wild-type mice recovering from atrophy, or wild-type mice of advanced age). These milder forms of disease also displayed a broader range of FNH scores among the acquired image volumes, suggesting a mixture of normal and abnormal sarcomere pattern throughout the tissue of these animals.

To more carefully test the diagnostic efficacy of SPQ scores, we performed receiver operating characteristic (ROC) analysis—both nonparametric and using multivariate logistic regression—to assess the specificity and sensitivity of SPQ measures in distinguishing groups by a clinically relevant statistical test (Table 1). As expected, evaluation of muscle damage based on a single sarcomere pattern parameter (FNH or MNHL) was sufficient to reliably detect severe MDs, such as dystrophy in mdx/utr mutants and muscle atrophy in hindlimb-suspended mice [ROC areas under the curve (AUC) $\geq 0.89$, with 95% confidence $>0.790$ for both FNH and MNHL in both comparisons; Table 1], at late stages when other signs of disease were clearly observed. Yet the diagnostic power of either of these parameters alone was not sufficient for clear discrimination of any moderate or mild form of MD from control samples (ROC AUC $\leq 0.755$, 95% confidence $\leq 0.620$ for all cases; Table 1).

Remarkably, however, bivariate logistic regression combining both FNH and MNHL substantially improved the impact of SPQ scoring and allowed us in several cases to efficiently discriminate intermediate forms of MDs from either control samples or severely affected individuals. For example, in a test for the subtle changes brought on by aging, we achieved quite effective differentiation of 24-month-old mouse muscle from 10-month-old specimens [ROC AUC 0.866, 95% confidence 0.786; Table 1]. This contrast was detected despite our finding of no significant difference in either agility or the percentage of fat-free body mass between breed-matched animals of these approximate age groups (Fig. 3) and the absence of any visibly recognizable alterations of sarcomere pattern. Even more strikingly, both mdx (mild) and mdx/utr (severe) dystrophic animals were absolutely discriminated from wild-type controls by a logistic regression...
Fig. 2 Sarcomere pattern quantification and correlation of sarcomere-length distribution with severity of muscular disorders. Single SHG optical sections (left panels) and corresponding sarcomere length maps (center panels) and angle maps (right panels) from (a) 5-week-old control and (b) *mdx*−/− *utr*−/− mice. Pseudocolor scale shows sarcomere lengths in μm (center) and angle of striations, relative to the long axis of the muscle fiber, in degrees (right). (c) Sarcomere length distributions affected by both mild and severe forms of muscular dystrophy. Histograms of sarcomere length were assembled from the full collection of SHG image stacks for each condition (see Section 4 for numbers of animals, specimens, and images). Fractions of hypercontracted sarcomeres (<1.6 μm) are shown; distributions of non-hypercontracted sarcomeres are filled with color matching the pseudocolor scale shown in (a). (d) Distribution of sarcomere length correlates with severity of various muscle disorders. Two characteristics of the length histograms (FNH and MNHL) were averaged for each entire image stack. Mean and standard deviation were calculated from the full collection of image stacks for each condition (see Section 4 for numbers of animals, specimens, and images). Asterisks (*) indicate data significantly different from control (p<0.05), as measured by a t-test for independent samples (Statistica 6.0, Statsoft, Inc.), with each image stack treated as an independent sample.
model combining all three measures of sarcomere length distribution (ROC AUC 1.000, 95% confidence 1.000; Table 1). The definitiveness of this result stands out, especially with regard to mdx, because the pathology of mdx mice is known to be subtle in the early weeks of life.

To examine the applicability of SPQ scoring in monitoring muscle healing after damage, we analyzed the alteration of sarcomere pattern associated with reloading of HLS-atrophied muscles. Two days of recovery did not change the cross-sectional diameter of muscle fibers in the gastrocnemius (Fig. 3), and has been shown in the literature to yield no improvement in either muscle mass or bone mineral density in adult C57BL6 mice (Ref. 36 and our own unpublished data). Nonetheless, we found a clear improvement of muscle SPQ values within hindlimb-reloaded muscle tissue imaged by SHG. In fact, recuperation of pattern within just two days of reloading was sufficient to reliably segregate recovering animals from the fully atrophic (ROC AUC 0.866, 95% confidence 0.799 in the trivariate regression model, and recovering limbs were actually more accurately discerned from HLS-atrophy specimens than from controls (Table 1). This result is remarkable, for it reveals both the speed of contractile adaptation to newly regained mobility and the sensitivity of SHG/SPQ to the early onset of musculoskeletal remodeling during physiological recovery from disease.

### 2.3 SPQ of Changes in Human Aging

To consider whether these methods were also applicable in human muscle, and to ask if SPQ would correlate with functional performance, we recruited groups of unrelated young adult and elderly volunteers who underwent tests of physical performance and provided a needle biopsy of the quadriceps muscle, vastus lateralis. We acquired SHG images and calculated SPQ values from each of the biopsies. As was seen for mice, SPQ appeared to efficiently resolve specimens from the young and old groups, even by univariate analysis of individual scores (ROC AUC values FNH=0.866, MNHL-

### Table 1 Discrimination of control and diseased mouse muscle by SPQ measurements. Receiver operating characteristic area-under-the-curve (ROC AUC) values are shown in bold face for comparisons among control, mild, and severe disorders. Confidence intervals for each of these calculated values were derived by 1000-fold bootstrap resampling with replacement. The 95% confidence interval for the ROC AUC of each comparison is shown in parentheses under the measured ROC AUC value.

<table>
<thead>
<tr>
<th>Muscular Dystrophy</th>
<th>FNH</th>
<th>MNHL</th>
<th>BVLR: FNH &amp; MNHL</th>
<th>TVLR: FNH &amp; MNHL &amp; ML</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>mdx</em> vs control</td>
<td>0.745</td>
<td>0.677</td>
<td>0.882</td>
<td>1.000</td>
</tr>
<tr>
<td>(0.619 to 0.862)</td>
<td>(0.543 to 0.805)</td>
<td>(0.794 to 0.963)</td>
<td>(1.000 to 1.000)</td>
<td></td>
</tr>
<tr>
<td><em>mdx</em>:uttr vs control</td>
<td>0.962</td>
<td>0.898</td>
<td>0.997</td>
<td>1.000</td>
</tr>
<tr>
<td>(0.908 to 0.999)</td>
<td>(0.791 to 0.986)</td>
<td>(0.987 to 1.000)</td>
<td>(1.000 to 1.000)</td>
<td></td>
</tr>
<tr>
<td><em>mdx</em>:uttr vs <em>mdx</em></td>
<td>0.796</td>
<td>0.794</td>
<td>0.797</td>
<td>0.795</td>
</tr>
<tr>
<td>(0.661 to 0.910)</td>
<td>(0.658 to 0.907)</td>
<td>(0.672 to 0.914)</td>
<td>(0.684 to 0.924)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Atrophy and Recovery</th>
<th>FNH</th>
<th>MNHL</th>
<th>BVLR: FNH &amp; MNHL</th>
<th>TVLR: FNH &amp; MNHL &amp; ML</th>
</tr>
</thead>
<tbody>
<tr>
<td>HLS vs control</td>
<td>0.890</td>
<td>0.896</td>
<td>0.908</td>
<td>0.910</td>
</tr>
<tr>
<td>(0.798 to 0.965)</td>
<td>(0.808 to 0.967)</td>
<td>(0.825 to 0.974)</td>
<td>(0.831 to 0.977)</td>
<td></td>
</tr>
<tr>
<td>Reload vs control</td>
<td>0.711</td>
<td>0.693</td>
<td>0.729</td>
<td>0.723</td>
</tr>
<tr>
<td>(0.567 to 0.841)</td>
<td>(0.564 to 0.816)</td>
<td>(0.577 to 0.884)</td>
<td>(0.580 to 0.893)</td>
<td></td>
</tr>
<tr>
<td>HLS vs reload</td>
<td>0.724</td>
<td>0.746</td>
<td>0.860</td>
<td>0.866</td>
</tr>
<tr>
<td>(0.539 to 0.878)</td>
<td>(0.573 to 0.889)</td>
<td>(0.742 to 0.960)</td>
<td>(0.779 to 0.960)</td>
<td></td>
</tr>
<tr>
<td>Aging</td>
<td>FNH</td>
<td>MNHL</td>
<td>BVLR: FNH &amp; MNHL</td>
<td>TVLR: FNH &amp; MNHL &amp; ML</td>
</tr>
<tr>
<td>----------------------</td>
<td>-----</td>
<td>------</td>
<td>-----------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>24 months vs 10 months</td>
<td>0.679</td>
<td>0.755</td>
<td>0.866</td>
<td>0.876</td>
</tr>
<tr>
<td>(0.528 to 0.802)</td>
<td>(0.618 to 0.864)</td>
<td>(0.786 to 0.936)</td>
<td>(0.803 to 0.941)</td>
<td></td>
</tr>
</tbody>
</table>

ML=mean sarcomere length; FNH=fraction of all sarcomeres that were not hypercontracted; MNHL=mean length of no-hypercontracted sarcomeres; BVLR=bivariate logistic regression; TVLR=trivariate logistic regression. Values highlighted in grey have lower limits for 95% confidence which are above 0.74, indicating a good diagnostic test. Numbers of animals, muscle samples, and image stacks analyzed for each group are given in the subsection of Section 4 specific to each mouse model.
small sample sizes for this study led to decreased lower bounds for confidence in ROC analysis of any of these individual scores. However, we found for human specimens that bivariate and trivariate logistic regression (combining FNH, MNHL, and/or ML) substantially improved the impact of SPQ scoring, to quite closely parallel the results seen for aging in mice [Fig. 5(b)].

3 Discussion

Three major conclusions arise from these studies that should impact the future of high-resolution imaging in the assessment of striated muscle disease. First, striation pattern is generally indicative of a broad range of muscle disease. Second, these changes in striation pattern can be quantified automatically to yield statistically powerful measures of muscle health. Third, endogenous SHG offers a means to extract these quantifiable patterns as high-contrast images from muscle tissue in its native state or in vivo.

The periodic structure of contractile machinery lies at the core of muscle function, and this assembly depends upon continual signaling, gene regulation, and biomechanical feedback. We found that three different disorders leading to loss of muscle performance were detectable by SPQ analysis, and quantifiable in their progression, even though none of their etiologies involves a primary molecular defect within the contractile apparatus. Thus, other deficits in muscle homeostasis or function will likely be amenable to SPQ, including neuropathies and cachexia secondary to cancer. The SPQ algorithm automatically extracts multiple measured variables from the striation pattern and allows one to transition from raw image to statistical analysis between subjects in a way that can be completely independent of operator judgment or bias. In this study, we analyzed the correlation of a number of SPQ variables (and combinations thereof) with muscle disease. From our current experience, sarcomere length gives the most robust quantification of progression from healthy to mild to severe disease. Why all three disorders studied manifest hypercontraction of sarcomeres is not certain, although dystrophin/utrophin deficiency is known to disrupt myofibril interaction with the cell cortex and to accelerate leakage of...
calcium into myofibers, and it is possible that similar effects are wrought by atrophy and sarcopenia. Nevertheless, this relatively subtle change in banding pattern was the most informative distinction between healthy and compromised muscle, apparently more sensitive than measurements of myofiber or tissue morphology or overt disruptions of the contractile lattice. It is known that muscles change their fiber-type composition in response to some of the disease conditions that we have observed here and such changes might possibly affect our measurements. However, we have examined the sarcomere pattern in muscles with varying proportions of different fiber types, as well as in mice where fiber types are marked by contrasting fluorescent reporter genes, and have not noted any distinctions between the striation patterns of different muscle fiber types (data not shown). We believe, based on these observations, that changes in fiber type probably do not alter the sarcomere length distributions that we have measured. Whatever may be the cause of the sarcomere-shortening effect that we have detected, a relationship between sarcomere length and contraction strength has been well documented, suggesting that the loss of performance in muscular disorders may be linked to such structural changes within the cells.

Many mathematical approaches can be applied to texture and pattern analysis in images. The Helmholtz equation works especially well in the context of sarcomeric banding, because the pattern so closely approximates a sine wave in one direction and is uniform in the orthogonal direction. Our analysis method specifically seeks such patterns and identifies the direction and period of the sine wave. In this instance, Helmholtz analysis is more effective than other methods, such as Fourier and wavelet analysis, that decompose the pattern into a sum of waves of predetermined spatial frequencies and orientations, none of which may exactly match the pattern of the image. Already, we have found that Helmholtz approximation yields greater than a 10-fold improvement in pattern-detection sensitivity over an application of 2D wavelet analysis that we have used previously (data not shown). Interestingly, efforts at combining measures of sarcomere length with both the angle and SHG intensity of individual striations in a multivariate treatment of the image pattern [an approach designed to capitalize on all of the variables portrayed in Fig. 2(a)] failed under ROC analysis to equal the power of the approach to SPQ that we have employed here: multivariate logistic regression focused on specific regions of the distribution of sarcomere length alone.

Further improvements upon the performance achieved in this study are still likely, through continued refinement of SPQ algorithms. For example, we have attempted to differentiate muscle biopsies from adult human volunteers diagnosed as normal versus intermediate-frail by Fried’s criteria for frailty. Unfortunately, SPQ analyses have not yielded a notable discrimination between two groups of these classes (five and six patients in each group; ROC AUCs: FNH=0.68, MNHL=0.67, ML=0.62). Similarly, we have not achieved high-quality ROC distinctions between pairs of several of the MD states that we have analyzed: e.g., mdx vs. HLS (ROC AUCs: FNH=0.65, MNHL=0.66, ML=0.68), or HLS vs. 24-month-old mice (ROC AUCs: FNH=0.73, MNHL=0.75, ML=0.79). Beyond the elementary level of analysis that we have explored in this paper, more diagnostic potential should result from spatial-distribution analysis of abnormal regions throughout the volume of tissue, possibly allowing distinction among different etiologies of disease.

Our assessment of the Helmholtz approach also defines some criteria for the design of imaging systems used for capturing sarcomere pattern. As discussed in Section 4, the image magnification and signal-to-noise ratio in our current images approach the lower limits at which this analysis works efficiently. Reducing image noise could allow for some decrease in magnification, and therefore measurement of more sarcomeres per image. Alternatively, lower noise at the current magnification could allow more detailed analysis of subpopulations of hypercontracted sarcomeres.

While SPQ can be applied to any imaging regime that highlights sarcomere striations, the advantages of SHG contrast make it an especially strong choice for future diagnostic muscle imaging. Reports from several labs have used precise morphometric, polarization anisotropy, biochemical, and genetic methods to show that SHG arises from the myosin thick filaments of sarcomeres intrinsically SHG from myosin harmonophores can be elicited and imaged in living tissue without dyes, and the deep sectioning of SHG imaging yields the equivalent of hundreds of serial histological sections within each spatially unified image volume of tissue. In our work, and in principle within a pathology laboratory, SHG imaging can be done immediately following biopsy collection, and pattern analysis is completed the same day. With a continuation of recent advances in microendoscopy, employing SHG-compatible GRIN lens technology imaging of sarcomere patterns will likely become possible in the tissue of live patients without removing biopsies. Streamlining of automated SPQ algorithms should then allow real-time quantification of myocyte health in the clinic and provide a rapid and unbiased tool for assessing diagnosis and prognosis in a wide range of muscular and myocardial disorders.

4 Methods

4.1 Animal Models of Atrophy, Dystrophy, and Sarcopenia

Muscle atrophy was induced in 6-month-old C57BL6 mice by a 14-day course of HLS, as described in the following section. HLS in both mice and rats has been shown to result in substantial atrophy of both muscle and bone within the hind legs during this time span. Returning animals to walking on all fours induces a recovery of muscle force, mass, and fiber size, as well as bone mass recovery, at a rate that is comparable to or slower than the initial rate of atrophy.

Mice of the mdx−/− genotype lack functional dystrophin protein and show continual necrosis and regeneration of muscle fibers starting at about 20 days of age and extending throughout the full course of life. Yet mdx−/− animals show normal muscle strength through much of life and can achieve close to 80% of the life spans of controls. In contrast, mdx−/−; utr−/− mice, which lack both dystrophin and the paralogous protein utrophin, display much more rapid incidence of muscle fiber damage and have a maximum lifespan of 20 weeks. Sarcopenia, the progressive loss of muscle mass and strength with aging, has been found to affect animals across many phyla. To evaluate the effectiveness of SHG
imaging in recognizing sarcopenic changes in the mouse, we examined muscle from mice at mid-adulthood (10 months old) and at an advanced age (24 months) close to the average lifespan of the wild type. All animal protocols were approved by the Animal Care and Use Committee of the University of Connecticut Health Center (UCHC), Farmington, CT.

4.1.1 Hindlimb suspension (HLS) protocol
Cages for hindlimb suspension (HLS) of mice were constructed based on the standard HLS cage design for unloading hindlimbs of rats. In addition to miniaturizing components, modifications included suspending the mice from a lightweight (26-g) roller ball carriage, comprising two spherical polymer balls rolling on twin rails, that allowed the suspended mice to walk freely around in the rat-sized wire bottom cages.

Six-month-old C57Bl/6 females were used. Mice were moved into HLS cages (one mouse per cage) two days before suspension and tails were taped one day before suspension to allow mice to become acclimated to tape. Mice were suspended for 14 days. Control mice, and those released from HLS for recovery, resided in normal mouse cages. During this protocol, all mice were housed in a room in which the ambient temperature was maintained at 80°F. Mice had free access to regular food and water. Body weights and behavior were monitored daily during suspension. Mice of three different groups were sacrificed for imaging of muscle from the gastrocnemius of each hind leg: control (n=7 mice, 14 muscle samples, 119 image stacks analyzed); mice suspended for 14 days and sacrificed without return to normal walking (“HLS;” n=6 mice, 12 muscle samples, 89 image stacks analyzed); and mice suspended for 14 days and then released for 2 days of normal walking (“reload;” n=5 mice, 10 muscle samples, 66 image stacks analyzed).

4.1.2 Breeding of dystrophic mouse models, mdx and mdx;utr
Founders for the dystrophy model breeding colony were kindly provided by Dr. Melissa Spencer (UCLA). Initial crosses of donated mdx−/−;utr+ mice were mated to mdx males ordered from the Jackson Laboratory (Bar Harbor, ME) and rederivation of mdx−/−;utr+ embryos was carried out by staff of the Center for Laboratory Animal Care and the Gene Targeting and Transgenic Facility at UConn Health Center. Genotypes of progeny were determined in each litter by PCR and/or sequencing. Five-week-old mice, both male and female, of three different genotypes were sacrificed for imaging of muscle from the gastrocnemius of each hind leg: wild-type C57Bl/6 (n=6 mice, 12 muscle samples, 128 image stacks analyzed); single-mutant mdx+/− (n=8 mice, 16 muscle samples, 101 image stacks analyzed); and double-mutant mdx−/−;utr−/− (n=7 mice, 14 muscle samples, 42 image stacks analyzed).

4.1.3 Adult and aged mice
For aging studies, 10- and 24 month-old C57BL6 males raised at Harlan, Inc. (Indianapolis, IN) were purchased through the National Institute on Aging. The animals were received and housed for acclimatization 5–7 days before each imaging experiment. Animals of two groups were sacrificed for imaging of muscle from the gastrocnemius of each hind leg: 10 month-old C57Bl/6 (n=5 mice, 10 muscle samples, 89 image stacks analyzed); and 24 month-old C57Bl/6 (n=10 mice, 20 muscle samples, 192 image stacks analyzed).

4.1.4 Mouse muscle samples
Mice were sacrificed by CO2 narcosis, and lower hindlimbs were removed and skinned for dissection. Small specimens of muscle (∼2 mm × 1 mm × 0.5 mm) were dissected from the gastrocnemius using the cutting blade of a 4.0-mm OD × 80-mm skeletal muscle biopsy needle (Popper & Sons Inc.) and briefly washed in ice-cold phosphate buffered saline (Invitrogen). Adipose and connective tissue were removed by scalpel under a dissecting microscope, and trimmed samples were stored in optical clearing buffer containing 50% glycerol at −20 °C until imaging. Optical clearing yields more than a two-fold increase in SHG image contrast for deep optical sections in muscle tissue and substantially diminishes SHG from collagen fibrils within muscle. Previous work has shown virtually no effect of glycogen upon myosin assembly, enzymatic activity, or sarcomere pattern.

4.1.5 Human patient biopsy samples
Fifteen patient volunteers were recruited for muscle biopsy. Four young (age 20–35 years) and 11 elderly subjects (age 65 years and older; average age 79 years) were free of physical disabilities, metastatic cancer, and medication known to affect muscle health. The skin and subcutaneous layer surrounding the biopsy area were anesthetized with 1% lidocaine (Elkins-Sinn, Cherry Hill, NJ), a small incision was made, and a deep-muscle biopsy was taken from the lateral portion of the vastus lateralis (=20 cm above the knee) using a Bergström 5-mm cannula biopsy needle (Depuy Orthopedics, Warsaw, IN) with suction. The tissue sample was immediately trimmed of any visible fat or connective tissue and mounted on the microscope slide in cold sterile saline solution. SHG imaging began within 30 minutes of sampling. The numbers of image stacks collected from the young and elderly groups were 22 and 39, respectively. The protocol for this study was approved by the Institutional Review Board of the University of Connecticut Health Center (UCHC), Farmington, CT.

4.1.6 SHG microscope setup and imaging conditions
All SHG experiments used a nonlinear optical imaging system described previously an Olympus BX61WI upright microscope equipped with a Fluoview 300 (Olympus USA) scanning head and Mira 900 Ti-sapphire laser pumped by a 10W, 532-nm Verdi (Coherent). The laser was tuned at 900 nm, and average power at the sample plane was adjusted to 15–25 mW, depending on the sample thickness. A long working-distance 40X, 0.8 N.A. water immersion objective lens and a 0.9 N.A. dry condenser (Olympus USA) were used for excitation and forward-propagating signal collection, respectively. The SHG signal was reflected with a 450-nm hard reflector (TLM2; bandwidth ±45 nm; CVI Laser), isolated from the laser fundamental and any fluorescence by a 450-nm bandpass filter (10-nm FWHM, CVI Laser), and detected by a

Journal of Biomedical Optics 044018-8 July/August 2008 • Vol. 13(4)
photon-counting photomultiplier module (Hamamatsu 7421, Bridgewater, NJ). Three-dimensional stacks containing up to 150 optical sections at 2-μm increments were acquired at slow scan speed (8 s per 512 × 512 pixel image; 30.5 μs per pixel) with a resolution of 0.29 μm per pixel.

### 4.1.7 Sarcomere pattern quantification (SPQ)

Computed analysis of sarcomere pattern was performed on SHG image stacks using a custom-designed plug-in collection within the Java-based program ImageJ (http://rsb.info.nih.gov/ij). This MuscleTone package quantifies the tonal component of sarcomere pattern by using the 2D Helmholtz equation to estimate the square of the wavenumber on a pixel-by-pixel basis according to \( k^2 = -\nabla^2 \phi / \rho \), where \( \rho \) is the pixel intensity. The mean and variance of the estimate for the wavelength (i.e., sarcomere length), \( \lambda = 2\pi / k \), are computed over a small neighborhood of each pixel. If the variance falls below a threshold, then the sarcomere length for that pixel is taken as the mean wavelength. MuscleTone computes 32-bit image maps of the following parameters for the neighborhood of every pixel of the raw image: average pixel intensity, sarcomere length, and striation angle (angle between A-bands and the long axis of the myofiber, assumed to be 90° for the ideal case). The striation angle is estimated by averaging \( \phi \) from \( \tan(\phi) = (\nabla_x \phi) / (\nabla_y \phi) \) over the pixel neighborhoods. Images were processed in three spatial frequency bands spanning a range of expected wavelengths between 4 and 12 pixels. The radius of the averaging neighborhood for each band was taken as 125% of the upper wavelength limit of that band (up to 15 pixels). The results for the various bands were combined by selecting the band giving the smallest standard deviation of the computed wavelength for each pixel. The maximum standard deviations allowed for assigning a nonzero score were 1.5 pixels and 50° for \( \lambda \) and \( \phi \), respectively. The accuracy of this algorithm and its sensitivity to noise were tested with computer-generated chirped sinusoidal patterns convolved with a level of random noise similar to that in our SHG images, and reliable measurements were obtained for sarcomere lengths of 5 to 11 pixels (1.45–3.2 μm). This range encompasses the majority of sarcomeres in normal muscle samples. Sarcomere length distributions were extracted from the generated maps with the ImageJ histogram tool.

In order to further streamline use of the MuscleTone algorithm on images randomly oriented in the xy-plane, the following method was used to automatically determine the angle perpendicular to the long axis of muscle fibers within the sample. A Gaussian filter with a radius equal to the upper wavelength limit was applied to each slice of the stack. The filtered image was thresholded, using ImageJ’s auto-threshold function to set the limits for the conversion to a binary intensity scale. Finally, the Find Edges filter of ImageJ was applied, and a Radon transform calculation was used to find the angle in the xy-plane with the largest variance along a projection. All sarcomere lengths and striation angles were then determined, as described above, relative to this cross-fiber axis, \( \phi_0 \).

### 4.2 ROC Analysis and Bootstrapped Estimation of Error

The receiver operating characteristic curve compares the sensitivity and specificity of a classifier system as the threshold for discrimination is varied, indicating the true-positive and false-positive rates for classification of members of two distinct groups. A test with no power in distinguishing between two groups yields a value of 0.5 for the area under the ROC curve (AUC), while a perfect classifier yields 1.0 for the ROC AUC. We calculated ROC curves and AUC values for primary data using STATA and Microsoft Excel. For estimation of error from our relatively small samples (and taking into account dependencies between image volumes from the same animal/patient, leg, or biopsy), we generated 1,000 model data sets for each class of specimen by bootstrap selection of data points with replacement, using custom-written C++ code. Logistic regression models were generated and ROC curve errors were calculated from the bootstrapped data sets using STATA.

### 4.2.1 Mouse mobility performance testing

These studies were conducted using a Rotamex 4/8 RotaRod instrument from Columbus Instruments (Columbus, OH). The testing protocol was designed in order to reliably diagnose the presence of declines in mobility and endurance in mice. Following a brief warm-up session (a linear acceleration from 0 to 12 RPM performed over 5 minutes), each animal underwent five consecutive increasingly challenging trials, each lasting 10 minutes. Trial 1 began at 0 RPM and accelerated to 24 RPM, trials 2–3 began at 0 RPM and accelerated to 36 RPM, while trials 4–5 began at 12 RPM and accelerated to 48 RPM. Testing sessions were standardized in terms of time of day, room, and study investigator. Results were analyzed using a repeated-measures ANOVA, allowing for the examination of both between-group effects and within-subject effects. A full model allowed for a simultaneous examination of between-group and within-subject effects.

### 4.2.2 Body composition and bone mineral density of mice

Body composition was determined using peripheral dual-energy X-ray absorptiometry (pDXA; PIXImus II; GE-Lunar Corp., Madison, WI). Prior to each series of scans, a tissue calibration scan was performed using the manufacturer’s provided phantom. Mice were anesthetized using 2.5% Isoflurane (IsoFlo; Abbott Laboratories, North Chicago, IL) mixed with oxygen (1.5 L/min) for a period of 8–10 min, including induction and scanning. The mice were then placed in the prone position on a specimen tray and scanned. The head was excluded from total body scans. Information was provided on fat, lean body mass, and bone mineral density involving total body, as well as femoral diaphysis. HLS mice were measured initially after taping the tails, but before suspension, and again at the end of the period of hindlimb suspension.

**Acknowledgments**

We thank Melissa Spencer for providing founder mice and George Keech and the staff of the Center for Laboratory Animal Care for assistance in establishing the mdx/utr mouse.
References


2. V. Preedy and T. Peters, Skeletal Muscle: Pathology, Diagnosis and Management of Disease (2002).


36. J. Machackova, J. Barta, and N. S. Dhalla, “Myofibrillar remodelling

Journal of Biomedical Optics 040418-10 July/August 2008 • Vol. 13(4)


