Rat-tail tendon fiber SAXS high-order diffraction peaks recovered by a superbright laboratory source and a novel restoration algorithm

Liberato De Caro,^a Davide Altamura,^a Teresa Sibillano,^a Dritan Siliqi,^a Giovanni Filograsso,^a Oliver Bunk^b and Cinzia Giannini^{a*}

^aIstituto di Cristallografia (IC-CNR), via Amendola 122/O, Bari, 70126, Italy, and ^bPaul Scherrer Institut, Swiss Light Source, Villigen PSI, 5232, Switzerland. Correspondence e-mail: cinzia.giannini@ic.cnr.it

Received 24 December 2012
Accepted 20 March 2013

The nanoscale structural order of air-dried rat-tail tendon is investigated using small-angle X-ray scattering (SAXS). SAXS fiber diffraction patterns were collected with a superbright laboratory microsource at XMI-LAB [Altamura, Lassandro, Vittoria, De Caro, Siliqi, Ladisa & Giannini (2012). *J. Appl. Cryst.* **45**, 869–873] for increasing integration times (up to 10 h) and a novel algorithm was used to estimate and subtract background, and to deconvolve the beam-divergence effects. Once the algorithm is applied, the peak visibility improves considerably and reciprocal space information up to the 22nd diffraction order is retrieved ($q = 0.21 \text{ \AA}^{-1}$, $d = 29 \text{ \AA}$) for an 8–10 h integration time. The gain in the visibility is already significant for patterns collected for 0.5 h, at least on the more intense peaks. This demonstrates the viability of detecting structural changes on a molecular/nanoscale level in tissues with state-of-the-art laboratory sources and also the technical feasibility to adopt SAXS fiber diffraction as a future potential clinical indicator for disease.

© 2013 International Union of Crystallography
Printed in Singapore – all rights reserved

1. Introduction

Collagen is among the most abundant proteins in vertebrates and a major constituent in many hierarchically structured biological tissues. More than 20 types of collagen (Myllyharju & Kivirikko, 2001; Shoulders & Raines, 2009) are known, which are distributed across different tissues: bone, tendon, ligament, dermis, cartilage, vitreous, skin, dentin, cell basement membrane, hair, cornea, placenta *etc.* In recent years, many collagen-related diseases have been identified as being either of genetic origin (Kuivaniemi *et al.*, 1991) or based on collagen mutations (Myllyharju & Kivirikko, 2004; Mahajan *et al.*, 2010) occurring during collagen production.

Collagen mutations can lead to various diseases at the tissue level: *Osteogenesis imperfecta*, which causes bone brittleness (Gajko-Galicka, 2002), chondrodysplasias, a skeletal disorder (Horton *et al.*, 1989), Ehler–Danlos syndrome, which causes deformities in connective tissue (Hamel *et al.*, 1998), Alport syndrome, which causes progressive hearing loss and kidney problems (Kashtan, 1993), osteoporosis, characterized by bone loss and increased risk of fracture (Shuster, 2005; Feng & McDonald, 2011), and Knobloch syndrome, where retina degeneration causes extreme nearsightedness (high myopia) (Mahajan *et al.*, 2010), to cite only a few relevant examples.

In tissues, collagen often forms fibers and networks that add structural strength to many body parts (Shoulders & Raines, 2009). In 1982, for the first time, the nanoscale structure of a

rat-tail collagen fiber was determined by small-angle X-ray scattering (SAXS), pointing out relevant differences between diffraction maxima of native chordae tendineae, native rat-tail tendon, when air dried or wet (Bigi *et al.*, 1982).

Indeed SAXS is a very promising characterization technique to study fibers as (i) it provides access to nano-scale structures; (ii) the experimental setup is rather simple; (iii) the sample does not need any specific preparation; and (iv) the diffraction pattern contains maxima in scattered intensity whose sharpness, positions and widths reflect the regularity, order and periodicity of the electron density along the fiber and the fiber packing (Suhonen *et al.*, 2005).

SAXS has been widely used together with wide-angle X-ray diffraction to inspect the structural properties of mineralized type I collagen fibers in healthy and pathologic bone (Fratzl *et al.*, 1993, 1998, 2004; Roschger *et al.*, 2008; Orgel *et al.*, 2006; Wess *et al.*, 1998; Giannini *et al.*, 2012), which is the most abundant among the major fibrous collagen types (I, II and III). Other tissue components, such as muscle (Reconditi, 2006), keratin (Feughelman *et al.*, 2005) or corneal tissue (Boote *et al.*, 2011), have been characterized by means of their typical fiber small-angle diffraction patterns.

Recently, in breast cancer patients, it has been hypothesized that changes of the molecular structure in hair (James *et al.*, 1999), skin derma and nails (James, 2009) could be related to the cancer malignancy. The hair fiber structure associated with colon cancer and Alzheimer's disease (James, 2003; James *et*

al., 2005) was indicated as a possible signature-at-a-distance biomarker. An additional diffractive ring, superimposed on the normal hard alpha keratin fiber pattern, was found in pathologic cases (James, 2006, 2009). Specifically to prostate cancer, a disease-related extra diffuse ring superimposed on the keratin pattern and falling between the 13th and 14th orders of the fiber pattern was measured (James, 2009). There is still a controversial and open debate as to whether keratin structural modifications could quantitatively and strictly be related to breast or other types of disease (Laaziri *et al.*, 2002; Corino & French, 2008), although structure–disease correlations have been definitively proved as detectable in SAXS experiments. However, whenever the relevant information is hidden under the high-order diffraction rings, which typically are mostly affected by noise, SAXS measurements have to be collected with synchrotron radiation sources.

Synchrotron radiation provides higher-quality diffraction patterns with respect to laboratory systems, owing to the application of a small and collimated beam of high flux, but the use of a synchrotron facility for examining fiber diffraction from tissues is not a practical means to provide daily clinical information. The alternative, novel, high-brilliance laboratory microsources available nowadays could offer a solution for future clinical applications, if the possibility of adopting SAXS fiber diffraction as a clinical biomarker is confirmed.

In this context, a recent superbright X-ray laboratory microsource (XMI-LAB) has been successfully used for SAXS experiments on bio- and nanomaterials (Altamura *et al.*, 2012). The already significant source brilliance was further enhanced by a novel restoration algorithm which improved the laboratory performance, in terms of signal-to-noise ratio and fringe visibility, towards third-generation synchrotron radiation beamlines, as shown by tests on a self-assembled nanoscale inorganic superstructure (De Caro *et al.*, 2012).

In the present work, the XMI-LAB instrumentation was used to inspect the nanoscale order of an air-dried rat-tail tendon, *i.e.* type I un-stretched collagen. SAXS data were collected for increasing integration times: 0.5, 1, 2, 3, 8, 10 h. An improved version of the algorithm described by De Caro *et al.* (2012), here applied for the first time to SAXS data, allowed reciprocal space information to be retrieved up to the 22nd diffraction order of the collagen fiber, at least for those data collected for 8 h or more. This result was obtained through estimation and subtraction of the background and an iterative deconvolution of beam-divergence effects from data by means of a Richardson–Lucy approach (De Caro *et al.*, 2012). The gain in the diffraction peak visibility was already considerable for patterns collected for 0.5 h, at least on the more intense peaks.

The paper is organized as follows: the algorithm developed to estimate and subtract the background and deconvolve the beam divergence from data and its application on simulated data are described in the *Methods* section; the experimental details on the SAXS data acquired on rat-tail tendon are provided in the *Experimental* section; the raw, background-subtracted and beam-divergence-deconvoluted SAXS profiles are described in the *Results* section.

2. Methods

2.1. Algorithm description

After integration of the measured SAXS patterns over the azimuthal angle, one-dimensional profiles $I(q)$ are obtained as a function of the scattering vector modulus q . Following the same notation as used by De Caro *et al.* (2012), the measured integrated intensity $I(q)$ is related to $I_{\text{ideal}}(q)$ by

$$I(q) = I_{\text{ideal}}(q) \otimes I_S(q) + I_{\text{bk}}(q) + n(q). \quad (1)$$

Here $\otimes$ denotes the one-dimensional convolution product. The difference between the ‘ideal’ profile $I_{\text{ideal}}(q)$ and the measured one $I(q)$ is caused by the finite beam divergence and/or detector point-spread function, the latter being here negligible with respect to the former. Therefore the instrumental broadening considered in the following is just due to the incident beam divergence. Such broadening can be described by means of a one-dimensional far-field source-intensity angular profile $I_S(q)$, which is obtained directly as the azimuthal average of the measured two-dimensional far-field source intensity angular distribution. Moreover, background intensity $I_{\text{bk}}(q)$ and noise $n(q)$ are superimposed on the ideal data. The ideal experimental task is to extract $I_{\text{ideal}}(q)$ from $I(q)$, assuming that either experimental measurements or estimations of the functions $I_S(q)$ and $I_{\text{bk}}(q)$ are available. This task can be achieved by a maximum likelihood estimation of $I_{\text{ideal}}(q)$, with respect to a Poisson noise modeling of $n(q)$, once the functions $I(q)$, $I_S(q)$ and $I_{\text{bk}}(q)$ are known, following the Richardson–Lucy (RL) approach, as described by De Caro *et al.* (2012). The obtained equation, which gives the maximum likelihood estimation of $I_{\text{ideal}}(q)$, can be solved iteratively, giving at the $(n + 1)$ th cycle

$$I_{\text{ideal},n+1}(q) = I_{\text{ideal},n}(q) \left\{ \bar{I}_S(q) \otimes \left[\frac{I(q)}{I_{\text{ideal},n}(q) \otimes I_S(q) + I_{\text{bk}}(q)} \right] \right\}. \quad (2)$$

When the background intensity $I_{\text{bk}}(q)$ is not experimentally available it can be estimated at any peak position by calculation. The simplest approaches to perform this task are to estimate $I_{\text{bk}}(q)$ from the slope of the background, or to interpolate $I_{\text{bk}}(q)$ from the intensities at higher- and lower-angle sides of a peak. The former method is effective in the presence of peak overlapping. More sophisticated and/or iterative approaches devoted to a successful removal of background intensity in crystallography are discussed by Steenstrup (1981), Brückner (2000) and David & Sivia (2001).

Here we have used a new iterative numerical method to estimate the background intensities usually involved in one-dimensional integrated SAXS measurements, generated by substrates – used as sample holders – such as Kapton foils, microscope glass slides and Lindemann glass capillary tubes. Typical backgrounds in SAXS one-dimensional integrated measurements are thus characterized by almost everywhere monotonically decreasing profiles, excepting a large peak at low q values caused by the presence of the beam stopper in the two-dimensional map data collection. Some examples of

measured SAXS one-dimensional integrated backgrounds at different acquisition times are shown in Fig. 1.

In order to estimate the background intensities directly from the one-dimensional SAXS as-measured profiles we need only *a priori* information about the dynamical range (DR) of the background profile, evaluated as the ratio of the maximum background intensity divided by the average background intensity at large q . DR can be obtained without necessarily measuring the actual background profile for every sample. Indeed DR is a function of the type of detector, the acquisition time, the source properties and the substrate used to hold the sample.

Thus, after having evaluated DR, to estimate the background profile we need to average each intensity value $I(q_i)$, in the one-dimensional SAXS profile obtained for the investigated sample, with the two neighboring values at q_{i-1} and q_{i+1} . Denoting by $\langle I(q_i) \rangle$ the local averaged values, we replace the intensity value $I(q_i)$ with the minimum value between $I(q_i)$ and $\langle I(q_i) \rangle$, i.e. $I(q_i) \rightarrow I'(q_i) = \min[I(q_i), \langle I(q_i) \rangle]$. This averaging step mainly affects intensity peak values and it has been repeated several times until the number of iterations is large enough to damp out all the peaks in the $I(q_i)$ curve, giving as final result a non-constant background without peaks. The iterations can be stopped when $\sum_i |I(q_i) - I'(q_i)|$ falls below a suitable threshold, except in the neighborhood of the main low- q peak, where the iterations have to be repeated until the dynamical range of the estimated background profile is lower than DR.

Let us denote with $\Delta I(q)$ the standard deviation on the measured data. The one-dimensional intensity at a particular q is given by the azimuthal integration over a specific Δq range in the two-dimensional map, in the annulus from q to $q + \Delta q$. Since the pixel intensity fluctuates in this annulus as a result of noise, the integrated one-dimensional profile will be the mean of a statistical distribution of all measurements that contribute to that annulus. From this statistical distribution it will be possible to estimate the intensity standard deviation. A good data set should exhibit small standard deviations (relative to the intensity).

Concerning the evaluation of the error $\Delta I_{\text{bk}}(q)$ on the background intensity $I_{\text{bk}}(q)$, assuming data affected by Poisson noise, one should put $\Delta I_{\text{bk}}(q) \simeq I_{\text{bk}}(q)^{1/2}$. However this is a coarse approximation. Indeed, just as with the one-dimensional integrated-intensity profile, the unknown background intensity can also be considered as the average of the statistically distributed background intensities across a specific Δq range of the two-dimensional map. Therefore, the corresponding error in the determination of $I_{\text{bk}}(q)$ would be the standard deviation of this statistical distribution, which is much smaller than $I_{\text{bk}}(q)^{1/2}$. In the unfavorable case where we do not have a direct measurement of the two-dimensional background map, we cannot rely on an immediate determination of $\Delta I_{\text{bk}}(q)$ by statistical calculations. Nevertheless, even in the absence of this measurement, if $I_{\text{bk}}(q)$ has to be mathematically evaluated by an $I(q)$ that is obtained as an azimuthally integrated value across a specific Δq range of the two-dimensional map, the obtained $I_{\text{bk}}(q)$ value could still be

assumed to be an average value obtained by several measurements. Thus the error $\Delta I_{\text{bk}}(q)$ on the background intensity $I_{\text{bk}}(q)$ can be set proportional to $I_{\text{bk}}(q)$, i.e.

$$\Delta I_{\text{bk}}(q) \simeq c(q)I_{\text{bk}}(q), \quad (3)$$

where $c(q)$ indicates a proportionality function which takes into account different numbers of measurements at different q values. Since the maximum error on an average I value obtained by N measurements is $I/N^{1/2}$, $c(q)$ can be set equal to $N(q)^{-1/2}$, where $N(q)$ is the number of two-dimensional intensity measurements, for a given q value, that have been azimuthally averaged.

After estimating the background intensity $I_{\text{bk}}(q)$, as previously described, the finite source size effects can be deconvolved from the measured $I(q)$ by means of the iterative equation (2), to obtain $I_{\text{ideal}}(q)$. In order to estimate the error made in the deconvolution of $I(q)$, due to the not exact knowledge of $I_{\text{bk}}(q)$ and $I(q)$, we have to consider how errors on the determination of the background $\Delta I_{\text{bk}}(q)$ and on the measured intensities $\Delta I(q)$, already discussed, influence the deconvolution process. They can be estimated as follows. After a suitable number of iterations, the deconvolution algorithm converges when

$$\bar{I}_s(q) \otimes \left[\frac{I(q)}{I_{\text{ideal},\infty}(q) \otimes I_s(q) + I_{\text{bk}}(q)} \right] \rightarrow 1, \quad (4)$$

where $I_{\text{ideal},\infty}(q)$ is the deconvolved intensity at convergence. In the presence of a background error estimation $\Delta I_{\text{bk}}(q)$ and an error $\Delta I(q)$ on the measured intensity, at convergence one would have

$$\bar{I}_s(q) \otimes \left[\frac{I(q) + \Delta I(q)}{I_{\Delta,\infty}(q) \otimes I_s(q) + I_{\text{bk}}(q) + \Delta I_{\text{bk}}(q)} \right] \rightarrow 1, \quad (5)$$

where $I_{\Delta,\infty}(q)$ is the deconvolved intensity at convergence in the presence of both $\Delta I_{\text{bk}}(q) \neq 0$ and $\Delta I(q) \neq 0$. After equating equations (4) and (5), and applying the Fourier convolution theorem, straightforward calculations lead to

Figure 1
Examples of one-dimensional integrated SAXS background profiles measured at XMI-LAB.

$$\Delta I_{\text{ideal}}(q) = \text{FFT}^{-1} \left\{ \frac{1}{\text{FFT}[I_s(q)]} \right\} \otimes [\Delta I_{\text{bk}}(q) + \Delta I(q)], \quad (6)$$

where FFT denotes the fast Fourier transform, FFT^{-1} the inverse FFT and $\Delta I_{\text{ideal}}(q)$ the error on the deconvolved intensity due to $\Delta I_{\text{bk}}(q)$ and $\Delta I(q)$ at convergence.

2.2. Simulations

In order to verify the reliability of the results obtainable by the previously described algorithm we have realized some tests on simulated data. Fig. 2(a) compares a one-dimensional SAXS experimental profile (red curve), obtained from a piece of air-dried rat-tail tendon (see *Experimental* for more details), with simulated data (black curve) characterized by peaks, background intensity and noise level chosen to simulate the actual experimental situations under investigation.

The simulated data $I_{\text{sim}}(q_i)$ have been obtained by convolving an ideal profile $I_{\text{ideal}}(q_i)$ – which should be retrieved by the algorithm – with a source-profile function, by adding a

background $I_{\text{bk}}(q_i)$ and by generating a noise level comparable to those of the considered experimental situations. Fig. 2(b) compares in a log–log plot, to better show small differences, the ‘true’ background input profile (blue curve) and the ‘recovered’ one (green curve) obtained by the proposed algorithm. The black curve is the simulated starting input $I(q_i)$ profile, whose enlarged view is reported in Fig. 2(a) (black curve) for a smaller q range, as schematically shown by the red dashed lines. The tests performed on several simulated data sets show that the proposed algorithm is usually either slightly underestimating or slightly overestimating the background as a function of different q regions. The residual factor (R factor) $R_{\text{bk}} = \sum_i |I_{\text{bk}}(q_i) - I_{\text{bk,rec}}(q_i)| / \sum_i I_{\text{bk}}(q_i)$ between the recovered and the ‘true’ background profiles can be used to evaluate the quality of the background estimates obtained by the proposed algorithm. Our tests give R values ranging from 0.03 to 0.08. The example shown in Fig. 2(b) has $R_{\text{bk}} = 0.075$. As SAXS one-dimensional profiles usually scale as a power of q_i ,

Figure 2

(a) Example of a one-dimensional SAXS experimental profile (red curve), plotted in a log scale, obtained from a piece of air-dried rat-tail tendon, and of simulated data (black curve), showing similar features of noise level, background and peak intensities; (b) Comparison in a log–log scale between the ‘true’ background profile (blue curve) and the ‘recovered’ profile (green curve) obtained by the proposed algorithm. The black curve is the starting input $I(q_i)$ profile, the same as Fig. 2(a) just plotted in a log–log scale and in a different pixel range, as schematically shown by the dashed red lines.

Figure 3

(a) In a log–log plot, the green curve shows the R factor between the simulated ideal $I_{\text{ideal}}(q_i)$ profile and the $I_{\text{dec}}(q_i)$ profile obtained by deconvolving the starting input $I(q_i)$ profile. The black curve shows the figure of merit (FOM) used to stop the deconvolution iterations. (b) Log plot of the $I_{\text{dec}}(q_i)$ profile (blue curve) obtained by deconvolving $I(q_i)$ (black curve), corresponding to the minimum of FOM. The comparison with the true solution $I_{\text{ideal}}(q_i)$ (red curve) shows a quantitative agreement on the maximum intensity values and positions of the peaks. The red-boxed inset shows the effects of noise amplification in $I_{\text{dec}}(q_i)$, caused by too many deconvolution iterations.

i.e. $I(q_i) \propto q_i^{-n}$, it is useful to calculate the background on $I(q_i)q_i^n$ instead of $I(q_i)$, with the exponent n ranging between two and three. In this way, all peaks are scaled to show fewer differences in the intensity variations and the background estimation is improved. After $I_{\text{bk}}(q_i)$ has been determined it can be rescaled on the correct values: $I_{\text{bk}}(q_i) \rightarrow I_{\text{bk}}(q_i)q_i^{-n}$.

After the background profile estimation, the deconvolution can be performed. The results obtained for the simulated example of Fig. 2(b) are shown in Fig. 3 where, in Fig. 3(a), we have plotted, as a function of the number of deconvolution iterations, the residual R factor $R_{\text{sol}} = \sum_i |I_{\text{ideal}}(q_i) - I_{\text{dec}}(q_i)| / \sum_i I_{\text{ideal}}(q_i)$ between the simulated ideal $I_{\text{ideal}}(q_i)$ and the $I_{\text{dec}}(q_i)$ profiles obtained by deconvolving the starting input $I(q_i)$ profile. In the deconvolution procedure we have used the previously estimated $I_{\text{bk,rec}}(q_i)$ background profile. Our simulated tests show that at the beginning the residual R_{sol} decreases as a function of the number of deconvolution cycles but after some iterations the residual factor increases, as shown in Fig. 3(a). In fact, during deconvolution by using an RL approach, as described by De Caro *et al.* (2012), one should handle the problem of noise amplification. The increase of the residual R_{sol} in Fig. 3(a) is just due to this amplification. Regularization of the solution is a widely used scheme to avoid noise amplification, but in our tests we followed a different approach. Usually the starting input profile $I_{\text{start}}(q_i)$ at the first cycle of an RL algorithm is the function $I(q_i)$ that has to be deconvolved. In this way, without any regularization, noise is readily amplified from the first deconvolution iteration. However, if one uses $I_{\text{start}}(q_i) = \langle I(q_i) \rangle$, i.e. equal to the average value of the function that has to be deconvolved, at the first iterations of the deconvolution procedure only low-spatial-frequency features of $I_{\text{dec}}(q_i)$ are actually restored. These components are not affected by noise. After several deconvolution iterations also higher-spatial-frequency components of $I_{\text{dec}}(q_i)$ begin to be restored, and if they are affected by noise, several artefacts are generated as a result of noise amplification. However, our tests show that it is possible to use a suitable figure of merit (FOM), plotted in Fig. 3(a) (black curve), which reaches the minimum just before the noise amplification starts.

This FOM is defined as a combination of two other figures of merit usually used to test the convergence of RL algorithms (De Caro *et al.*, 2012):

$$F_n(q_i) = \left(\bar{I}_s(q_i) \otimes \left\{ \frac{I(q_i)}{[I_{\text{dec},n}(q_i) \otimes I_s(q_i)] + I_{\text{bk}}(q_i)} \right\} \right) \rightarrow 1, \quad (7)$$

and

$$\Delta I_n = \sum_i |I_{\text{dec},n+1}(q_i) - I_{\text{dec},n}(q_i)| / \sum_i |I_{\text{dec},n}(q_i)| \rightarrow 0, \quad (8)$$

$$\text{FOM} = 1 - F_n - \Delta I_n \rightarrow 0. \quad (9)$$

Thus we have adopted the following strategy to avoid noise amplification. Instead of regularizing the solution, we have used the original RL recursive algorithm given by equation (2) and considered the best deconvolved solution as that obtained

after a number of deconvolution iterations for which the defined FOM given by equation (9) reaches its minimum value. The result obtained is shown in Fig. 3(b). The blue curve is the $I_{\text{dec}}(q_i)$ profile obtained after 225 deconvolution iterations, corresponding to the minimum of the FOM. The comparison with the true solution $I_{\text{ideal}}(q_i)$ (red curve) shows that the small non-correct estimate of the background intensity, realized by the first step of the algorithm, influences essentially the minima of the profiles but neither the maxima intensity values nor their positions. A minimum residual factor R_{sol} of about 0.075 is obtained. The inset of Fig. 3(b) shows the effects of noise amplification in the $I_{\text{dec}}(q_i)$ profile after 2000 deconvolution iterations. Our tests on several sets of simulated data have always confirmed this kind of behavior, proving that the deconvolved profiles are quantitatively related to the correct solution at the maximum peak values. Only the minima of the reconstructed profiles are affected by the background error estimate. Moreover our simulated tests indicate that a denoising of the input data leads to the possibility to perform many more iterations – usually a few thousand – because noise amplification is suppressed. Nevertheless, the quality of the reconstructed profiles $I_{\text{dec}}(q_i)$ is not improved with respect to that obtained by the above-described approach because the error in the estimate of the background prevents smaller R_{sol} values being reached. Further studies are needed to verify the reliability of deconvolved data obtained by the proposed algorithm in more difficult experimental situations, in which the convolved peaks of $I(q_i)$ are much smaller than those considered in this work.

3. Experimental

3.1. Laboratory setup

The rotating anode X-ray microsource Rigaku FR-E+ SuperBright at XMI-LAB (Altamura *et al.*, 2012) is coupled through focusing high-flux multilayer optics [Confocal Max-Flux (CMF 15-105) for Cu radiation] to a three-pinhole SAXS camera (SMAX-3000). A Triton 20 gas-filled proportional counter SAXS detector, 20 cm in diameter and 195 μm effective pixel size, was positioned for the present experiment at ~ 1920 mm from the sample to collect up to 22 diffraction orders of the SAXS pattern from a piece of air-dried rat-tail tendon. The illuminated area was about 100 μm in diameter. The collected two-dimensional images, once the beam center had been determined, were calibrated with a silver behenate standard (Blanton *et al.*, 1995) and azimuthally integrated into one-dimensional radial profiles.

3.2. Synchrotron setup

Experiments were also performed at the cSAXS beamline of the Swiss Light Source (SLS), Paul Scherrer Institut, Villigen, Switzerland (Bunk *et al.*, 2009). A monochromatic X-ray beam ($E = 12.398$ keV) was focused down to about $(100 \mu\text{m})^2$ by a bent monochromator crystal and a bent mirror. A Pilatus 2M detector was placed at a sample–detector distance of 2159 mm.

4. Results

Figs. 4(a) and 4(b) show the raw SAXS one-dimensional profiles (black curves), collected for 8 h and 100 s acquisition times, at XMI-LAB and at the cSAXS beamline, respectively. The background intensity profiles (green curves), calculated as described in the *Methods* section, are also shown. They have been obtained by stopping the background determination procedure when $\sum_i |I(q_i) - I'(q_i)| \simeq 0.000001$. The blue curve in Fig. 4(a) is the deconvolved intensity obtained for the XMI-LAB raw data by equation (2) after about 1000 iterations, corresponding to the minimum of the FOM. Error bars are

Figure 4

(a) SAXS one-dimensional profiles $I(q)$ collected at XMI-LAB from air-dried rat-tail tendon for a 10 h long acquisition (black curve). The green curve shows the background intensity $I_{bk}(q_i)$ obtained by applying the algorithm described in the *Methods* section. The blue curve is the calculated intensity $I_{ideal}(q_i)$ obtained by applying the RL deconvolution algorithm [see equation (2)]. Error bars are shown in black. The numbers in the upper side of the figure denote the diffraction orders. (b) SAXS one-dimensional profiles $I(q)$ collected at the SLS synchrotron beamline from air-dried rat-tail tendon for an acquisition time of 100 s (black curve). The green curve shows the background intensity $I_{bk}(q_i)$ obtained by applying the algorithm described in the *Methods* section. The red curve is the calculated intensity $I_{ideal}(q_i)$ obtained by subtracting the background intensity. Error bars are shown in black. The numbers on the upper side of the figure denote the diffraction orders.

Figure 5

(a) Comparison of one-dimensional SAXS profiles, obtained from Figs. 1(a) and 1(b), after background subtraction and deconvolution (only for XMI-LAB data). (b) Comparison of one-dimensional SAXS profiles obtained after background subtraction and deconvolution of XMI-LAB data collected for different acquisition times, namely 2 and 8 h. (c) Comparison of one-dimensional SAXS profiles obtained after background subtraction and deconvolution of XMI-LAB data collected for different acquisition times, namely 0.5 and 8 h.

shown in black. The background-subtracted profile (red curve) shown in Fig. 4(b) has been obtained from the raw synchrotron data without any deconvolution. The numbers on the upper side of Figs. 4(a) and 4(b) denote the diffraction orders.

Fig. 5(a) compares the SAXS profiles collected with 8 h acquisition time at XMI-LAB after the deconvolution procedure (blue curve) with the synchrotron data (red curve) obtained after background subtraction and rescaled by a constant factor. From the comparison between the two curves it can be evinced that the deconvolved data collected for 8 h at XMI-LAB contain information about the rat collagen fibers up to the 22nd diffraction order. The two profiles show many common features, several of them being visible only after this data treatment. For example, the fourth order, which was not directly visible in the raw XMI-LAB data, became visible after the deconvolution procedure. The main differences are confined in the peak-minima regions and are due to the background intensity determination. Figs. 5(b) and 5(c) show the SAXS profiles collected for 2 and 0.5 h (red curves) and 8 h (blue curve) acquisition times, respectively, at XMI-LAB after deconvolution. It should be noted that the reduction of acquisition time mainly affects the intensity determination of low-intensity peaks, the more intense peaks being well determined at 2 h and even at 0.5 h. Conversely, smaller peaks need a longer acquisition time to be correctly evaluated.

5. Conclusions

SAXS data collected for dried rat-tail collagen samples by using a laboratory-based superbright microfocus X-ray source and a synchrotron radiation SAXS beamline have been compared. It has been demonstrated that a novel algorithm for deconvolving the laboratory data from beam divergence can yield data quality comparable to the synchrotron radiation source. This may represent an important milestone towards future applications where SAXS laboratory equipment is needed to carry out diagnostic processes, as in nanomedicine research.

Work is in progress to verify if this approach could be used for other experimental data, such as wide-angle X-ray scattering and X-ray powder diffraction.

The authors acknowledge financial support from the SEED project 'X-ray synchrotron class rotating anode microsource for the structural micro imaging of nanomaterials and engineered biotissues (XMI-LAB)' IIT Protocol No. 21537 of 23/12/2009, and from the FIRB 2009/2010 project 'Rete integrata per la Nano Medicina (RINAME)' – RBAP114AMK_006. Rocco Lassandro is acknowledged for his technical support at XMI-LAB. Andreas Menzel is gratefully acknowledged for setting up the experiment at the cSAXS beamline and for providing beamtime for the measurements.

References

Altamura, D., Lassandro, R., Vittoria, F. A., De Caro, L., Siliqi, D., Ladisa, M. & Giannini, C. (2012). *J. Appl. Cryst.* **45**, 869–873.

- Bigi, A., Ripamonti, A., Roveri, N., Compostella, L., Roncon, L. & Schivazappa, L. (1982). *Int. J. Biol. Macromol.* **4**, 387–392.
- Blanton, T. N., Huang, T. C., Toraya, H., Hubbard, C. R., Robie, S. B., Louër, D., Göbel, H. E., Will, G., Gilles, R. & Raftery, T. (1995). *Powder Diffr.* **10**, 91–95.
- Boote, C., Kamma-Lorger, C. S., Hayes, S., Harris, J., Burghammer, M., Hiller, J., Terrill, N. J. & Meek, K. M. (2011). *Biophys. J.* **101**, 33–42.
- Brückner, S. (2000). *J. Appl. Cryst.* **33**, 977–979.
- Bunk, O., Bech, M., Jensen, T. H., Feidenhans'l, R., Binderup, T., Menzel, A. & Pfeiffer, F. (2009). *New J. Phys.* **11**, 123016.
- Corino, G. L. & French, P. W. (2008). *Int. J. Cancer*, **122**, 847–856.
- David, W. I. F. & Sivia, D. S. (2001). *J. Appl. Cryst.* **34**, 318–324.
- De Caro, L., Altamura, D., Vittoria, F. A., Carbone, G., Qiao, F., Manna, L. & Giannini, C. (2012). *J. Appl. Cryst.* **45**, 1228–1235.
- Feng, X. & McDonald, J. M. (2011). *Annu. Rev. Pathol. Mech. Dis.* **6**, 121–145.
- Feughelman, M., Willis, B. & James, V. (2005). *Proceeding of the 11th International Wool Research Conference*, Leeds, UK, September 2005, 65FI, electronic publication.
- Fratzl, P., Fratzl-Zelman, N. & Klaushofer, K. (1993). *Biophys. J.* **64**, 260–266.
- Fratzl, P., Gupta, H. S., Paschalis, E. P. & Roschger, P. (2004). *J. Mater. Chem.* **14**, 2115.
- Fratzl, P., Misof, K., Zizak, I., Rapp, G., Amenitsch, H. & Bernstorff, S. (1998). *J. Struct. Biol.* **122**, 119–122.
- Gajko-Galicka, A. (2002). *Acta Biochim. Pol.* **49**, 433–441.
- Giannini, C., Siliqi, D., Bunk, O., Beraudi, A., Ladisa, M., Altamura, D., Stea, S. & Baruffaldi, F. (2012). *Sci. Rep.* **2**, 435.
- Hamel, B. C., Pals, G., Engels, C. H., van den Akker, E., Boers, G. H., van Dongen, P. W. & Steijlen, P. M. (1998). *Clin. Genet.* **53**, 440–446.
- Horton, W. A., Campbell, D., Machado, M. A. & Chou, J. (1989). *Am. J. Med. Genet.* **34**, 579–583.
- James, V. J. (2003). *Med. Sci. Monit.* **9**, 79–84.
- James, V. J. (2006). *Cancer Detect. Prev.* **30**, 233–238.
- James, V. J. (2009). *Int. J. Cancer*, **125**, 133–138.
- James, V., Kearsley, J., Irving, T., Amemiya, Y. & Cookson, D. (1999). *Nature*, **398**, 33–34.
- James, V. J., Richardson, J. C., Robertson, T. A., Papadimitriou, J. M., Dutton, N. S., Maley, M. A. L., Berstein, L. M., Lantseva, O. E. & Martins, R. N. (2005). *Med. Sci. Monit.* **11**, 53–57.
- Kashtan, C. E. (1993). *Alport Syndrome and Thin Basement Membrane Nephropathy*. In *GeneReviews*, edited by R. A. Pagon, T. D. Bird, C. R. Dolan, K. Stephens & M. P. Adam. Seattle: University of Washington.
- Kuivaniemi, H., Tromp, G. & Prockop, D. J. (1991). *FASEB J.* **5**, 2052–2060.
- Laaziri, K., Sutton, M., Ghadirian, P., Scott, A. S., Paradis, A. J., Tonin, P. N. & Foulkes, W. D. (2002). *Phys. Med. Biol.* **47**, 1623–1632.
- Mahajan, V. B., Olney, A. H., Garrett, P., Chary, A., Dragan, E., Lerner, G., Murray, J. & Bassuk, A. G. (2010). *Am. J. Med. Genet.* **152**, 2875–2879.
- Myllyharju, J. & Kivirikko, K. I. (2001). *Ann. Med.* **33**, 7–21.
- Myllyharju, J. & Kivirikko, K. I. (2004). *Trends Genet.* **20**, 33–43.
- Orgel, J. P. R., Irving, T. C., Miller, A. & Wess, T. J. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 9001–9005.
- Reconditi, M. (2006). *Rep. Prog. Phys.* **69**, 2709–2759.
- Roschger, P., Paschalis, E. P., Fratzl, P. & Klaushofer, K. (2008). *Bone*, **42**, 456–466.
- Shoulders, M. D. & Raines, R. T. (2009). *Annu. Rev. Biochem.* **78**, 929–958.
- Shuster, S. (2005). *Med. Hypoth.* **65**, 426–432.
- Steenstrup, S. (1981). *J. Appl. Cryst.* **14**, 226–229.
- Suhonen, H., Fernández, M., Serimaa, R. & Suortti, P. (2005). *Phys. Med. Biol.* **50**, 5401–5416.
- Wess, T. J., Hammersley, A. P., Wess, L. & Miller, A. (1998). *J. Struct. Biol.* **122**, 92–100.