the frequency is unchanged, the secondary radiation from spatial positions is not necessarily in phase due to the internal structure of the collagen fibrils. This leads to interference effects between the scattered waves. Below 10keV, X-ray photons are primarily absorbed or phase shifted as they pass through the sample. This means that X-ray microscopes offer unique opportunities for quantitative imaging without inelastic or multiple elastic scattering effects. (Jacobson & Kirz, 1998)

A kinematical approximation will be adopted throughout which assumes that a photon can only experience a single, weak scattering event in its passage through the specimen. Validity of the single-scattering approximation implies that the amplitude of the singly scattered radiation is very small compared with the incident beam amplitude. Then the amplitude of the doubly and multiply scattered radiation is very small and negligible compared with the singly scattered.(Cowley, 1995)

Whilst the model is generally of little quantitative use, it does provide a simple interpretation of the geometry of diffraction patterns. In addition kinematical theory is useful for interpreting weak beam images. (Cowley, 1992) Consider a collimated beam of X-rays, wavelength λ, and incident on a single collagen fibril. The amplitude and phase of the scattered X-rays are represented by

$$A = \sum_{j=1}^{N} F_j e^{-\Delta\phi} \tag{1}$$

Where F_j is the structure factor at a particular distance $\boldsymbol{r}_j$. The phase difference between different waves scattered by the same structure displaced by a distance $\boldsymbol{r}_j$ from each other is,

$$\Delta\phi = (2\pi/\lambda)\cdot(\mathbf{r}_j\cos\theta' - \mathbf{r}_j\cos\phi) \tag{2}$$

$$\Delta\phi = (\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{r}_j \tag{3}$$

Where the corresponding angles are shown in Figure 5. Writing $\boldsymbol{Q}=(\boldsymbol{k}-\boldsymbol{k_0})$ the phase difference becomes,

$$\Delta\phi = \mathbf{Q}\cdot\mathbf{r}_j \tag{4}$$

$\boldsymbol{Q}$ is the scattering vector and contains all the necessary information about $\boldsymbol{k}$, $\boldsymbol{k_0}$ and λ.

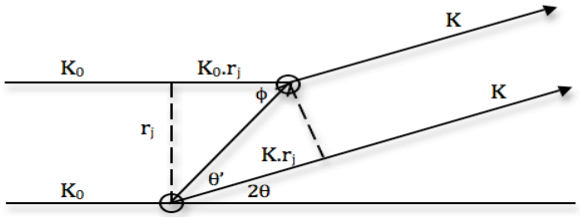

Figure5: Schematic representation of elastic scattering from two identical points, showing all interactions.

The modulus of $\boldsymbol{Q}$ is related to the scattering angle 2θ by,

$$Q = (4\pi/\lambda)\sin\theta \tag{5}$$

Where λ is the X-ray wavelength. If z is the specimen to detector distance and Z is of a diffraction maximum form the centre of the detector, 2θ can be measured from the relationship

$$2\theta = \tan^{-1}(Z/z) \tag{6}$$

Hence the d-spacing can be determined using Bragg's law,

$$2d\sin\theta = n\lambda \tag{7}$$

Where θ is the Bragg angle and n in this instance is 1 as only the first order reflection in the meridian

directional is under investigation.
The amplitude $|F(\mathbf{Q})|$ of a particular structure factor indicates the extent to which the electron density is concentrated on planes parallel to the Bragg planes, while its phase indicates the position of high electron density relative to the Bragg planes.

For an assemblage of N fibrils, the intensity of the resultant scattered wave is equal to the square modulus of the structure factor,

$$I(\mathbf{Q}) = |A(\mathbf{Q})|^2 = \sum_{j=1}^{N} F_j(\mathbf{Q})F_k^*(\mathbf{Q}) + \sum_{j\neq k}^{N-1}\sum F_j(\mathbf{Q})F_k^*(\mathbf{Q})e^{-i(\mathbf{r}_j-\mathbf{r}_k)\cdot\mathbf{Q}/N} \tag{8}$$

Where $\boldsymbol{r_j}$ defines the position of the j^{th} fibril and $F^*_j(\boldsymbol{Q})$ is the complex conjugate of $F_j(\boldsymbol{Q})$.(Aspden & Hukins, 1979)(Meek & Quantock, 2001) When $j=k$, there are N terms whose sum is NF^2 and when $j\neq k$ there are $(N\text{-}1)$ terms of which two correspond to the vectors joining a given pair. One is the complex conjugate of the other. When $\boldsymbol{r}_k$ is not equal to $\boldsymbol{r}_j$, the components of $(\boldsymbol{r}_j\text{-}\boldsymbol{r}_k)$ will be non-zero integers. The summation will disappear as $\boldsymbol{r}_k$ and $\boldsymbol{r}_j$ are orthogonal to each other.

Re-expressing the far-field (Fraunhofer) diffraction pattern of the structure factor equation in terms of electron density and the reciprocal lattice, $F(\boldsymbol{Q})$ can be represented by,

$$F_j(\mathbf{Q}) = \sum_{j=1}^{N} f_j(\mathbf{Q})e^{i(\mathbf{Q}\cdot\mathbf{r}_j/N)} \tag{9}$$

Where $f_j(\boldsymbol{Q})$ is the axially projected atomic form factor and $\boldsymbol{r}_j$ is the axial position along the D-period. Thus, the diffraction amplitude is essentially determined by the Fourier transform of the atomic form factor which is related to the electron density $\rho_j(\boldsymbol{r})$. Since the intensity $I(\boldsymbol{Q})=F(\boldsymbol{Q})F^*(\boldsymbol{Q})$ is measured, only amplitude information on the electron density distribution is available directly,

$$I(\mathbf{Q}) = \sum_{j=1}^{N} f_j(\mathbf{r}_j)e^{i(\mathbf{Q}\cdot\mathbf{r}_j/N)} f_j^*(\mathbf{r}_j)e^{-i(\mathbf{Q}\cdot\mathbf{r}_j/N)} = \sum_{j=1}^{N} |f_j(\mathbf{r}_j)|^2 \tag{10}$$

Thus retrieval of the relative phase in the Fraunhofer diffraction plane is equivalent to obtaining a full description of the wave function, which has originated from the specimen and the electron distribution within the D period can be reconstructed.

Despite the early pioneering work on X-ray interferometry in the 1960s (Bonse & Hart, 1965), the majority dark-field imaging (Morrison & Browne, 1992)(Chapman, Jacobsen, & Williams, 1996) methods were introduced in the late 1990s. The development of such advanced imaging methods is particularly difficult for hard X-rays, because of the lack of efficient X-ray optics. Existing hard-X-ray dark-field-imaging methods, for example, rely on the use of crystal optics that can accept only a very narrow energy bandwidth (~0.01%) and angular divergence(~1 arcsec)(Pfeiffer, 2008). This is why dark-field or scattering-based imaging is currently restricted in practice to applications at highly brilliant synchrotron X-ray sources and is not available for widespread applications that require a method applicable to standard X-ray tubes.

The development of techniques using coherent X-rays is of considerable current interest. This is largely because of new undulator based synchrotron X-ray sources are capable of providing coherent hard X-ray beams several orders of magnitude more intense than previously available.

The prerequisite for ptychography is a well known substantially localised and mainly coherent illumination. If an object is illuminated with coherent electromagnetic radiation such as highly brilliant X-rays, a diffraction pattern is formed in the Fraunhofer far-field that is related via a Fourier transform. The aim of coherent diffractive imaging or so called lens-less imaging is to directly reconstruct the original transmission function of the specimen from its measured diffraction pattern. The great experimental advantage of diffraction is that the interference condition is determined only by scattering within the specimen itself. This technique requires no optics and does not impose stringent resolution requirements on the detector.(Miao, Charalambous, Kirz, & Sayre, 1999) Reinterference of the beams that have traveled large distances through optical apparatus is not required. This means that the

effective spatial resolution can be in the sub-angstrom level. (Faulkner & Rodenburg, 2004)

The physical basis of ptychography relies on being able to move a probe across the specimen (or change the illumination conditions) in such a way that a multiple set of intensity measurements can be made in the Fraunhofer diffraction plane. (Rodenburg, 1992) Each pattern originates from a different but overlapping region of the specimen, which is moved laterally across the illuminating beam. Using multiple overlapping apertures allows examination of a wide field if view.

Once such an oversampled diffraction pattern is obtained, the iterative phasing method starts with a predefined set of phases for diffraction amplitudes and Fourier transforms back and forth between amplitudes in reciprocal space and density in real space. In each iteration the real space density is confined to within the finite specimen size and the square of the diffraction amplitudes in reciprocal space is made equal to the experimentally measured intensities. This iterative process continues until the RMS error is sufficiently low.

The condition of the overlap requires the iterative algorithm to find a solution which is consistent with all parts of the object function illuminated at a specific probe position. The requirement for consistency between different measurements eliminates possible uniqueness problems, by breaking the symmetry of the situation. (Faulkner & Rodenburg, 2004) Both the RMS error and the conditions for overlap will be discussed in the next section.

When highly brilliant coherent electromagnetic radiation is scattered from an area in a disordered system, it gives rise to random diffraction and the resulting intensity distribution, observed far from this region, will have a granular nature commonly referred to as speckle.(Libbert, 1997) Such a speckle pattern is uniquely related to the exact spatial arrangement of the disorder and is produced whenever randomly distributed regions of a material produce different phase shifts into the scattering of the coherent radiation. (Sutton, 1991) Although the speckle pattern may be difficult to interpret as a spatial structure, a unique advantage in scattering with a coherent beam lies in the ability to observe the dynamics of the disorder. When the spatial arrangement of the disorder evolves with time the speckle pattern also changes.

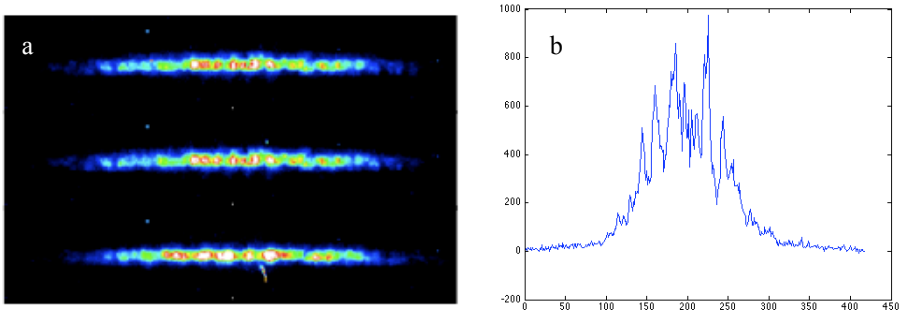

Figure 6: Sample diffraction pattern showing a) an array of reflections from real data and b) a profile through the reflection showing the speckle peaks.

The Phase Retrieval Algorithm

Ptychography was initially proposed by Hergel and Hoppe in 1970 (Hergel & Hoppe, 1970) for use in transmission electron diffraction microscopy. Hoppe (Hoppe, 1969) described how finite, coherent illumination could be used to get intensity from interference between Bragg peaks thereby restoring the phase information. The first widely accepted phase algorithm, mainly for non-periodic objects was put forward in 1972 by Gerchberg and Saxton. (Gerchberg & Saxton, 1972) The idea being that, if partial information about the magnitude of the object density as well as about the magnitude of the object's Fourier transform can be supplied, the phase information may be recovered

A variation of this idea was modelled by Fienup (Fienup, 1982; Fienup, 1987) which only required the support in the real plane to be known, rather than the entire intensity. However the use of larger non-periodic biological specimens, where the sample size is larger than the beam size required alterations of these earlier algorithms.

Let $O(\mathbf{r})$ and $P(\mathbf{r})$ represent two-dimensional complex functions. $O(\mathbf{r})$ is the transmission function of the specimen. $P(\mathbf{r})$ is the illumination function, which in this case is given by the 25μm x 25μm slit

used to define the aperture. Here *O(**r**)* can be moved with respect to *P(**r**)* by various distances ***R***, although *P(**r**)* can equally well be moved. I will use the notation of aperture to define the pupil at $z = 0$.

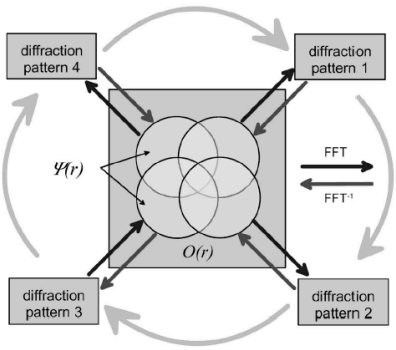

Figure 7: Diagram of the phase retrieval algorithm, where the outer arrows indicate the position stepping within one iteration, while the arrows within indicate (inverse) Fourier transforms and desired input-output information.

The product of the transmission function with the illumination function is formed to produce the exit wave function of *ψ(**r**)*,

$$\psi(\mathbf{r},\mathbf{R}) = O(\mathbf{r}-\mathbf{R})P(\mathbf{r}) \tag{11}$$

This is generally true for a thin sample, however the exit wave for an assemblage of collagen fibrils in a tissue is simply the sum of each individual fibril exit wave. (Aspden & Hukins, 1979) The algorithm works to find the phase and modulus of the complex specimen transmission function *O(**r**)* which is affected by the electron density along the length of the fibril. The pupil function *P(**r**)* needs to be accurately characterized if the algorithm is to converge to give a realistic reconstruction of the sample.

The algorithm proceeds as follows. A random array of amplitudes and phases is used as an initial guess of the object function $O_{g,n}(\mathbf{r})$ where *g* represents the guessed function at the *nth* iteration. This is a real space function. The guessed exit wave function for position ***R*** is calculated as above,

$$\psi_{g,n}(\mathbf{r},\mathbf{R}) = O_{g,n}(\mathbf{r}-\mathbf{R})P(\mathbf{r}) \tag{12}$$

In the Fraunhofer regime, the diffraction pattern in diffraction space (***k***-space) is related to the real space function via a Fourier transform, thus the corresponding wave function in the diffraction space plane for position ***R***,

$$\Psi_{g,n}(\mathbf{k},\mathbf{R}) = \Im[\psi_{g,n}(\mathbf{r},\mathbf{R})] \tag{13}$$

Where ***k*** is the reciprocal space co-ordinate. Here $\Psi_{g,n}(\mathbf{k},\mathbf{R})$ is a guessed version of the actual wavefunction in diffraction space since it has been produced by a guessed object function $O_{g,n}(\mathbf{r}-\mathbf{R})$. Here $\Psi_{g,n}(\mathbf{k},\mathbf{R})$ can also be written as,

$$\Psi_{g,n}(\mathbf{k},\mathbf{R}) = |\Psi_{g,n}(\mathbf{k},\mathbf{R})| e^{i\theta_{g,n}(\mathbf{k},\mathbf{R})} \tag{14}$$

Where $|\Psi_{g,n}(\mathbf{k},\mathbf{R})|$ is the guessed wavefunction amplitude and $\theta_{g,n}(\mathbf{k},\mathbf{R})$ is the guessed phase, both in diffraction space at iteration *n* for position ***R***. The guessed amplitude is then replaced with the actual amplitude taken from experimental results,

$$\Psi_{c,n}(\mathbf{k},\mathbf{R}) = |\Psi(\mathbf{k},\mathbf{R})| e^{i\theta_{g,n}(\mathbf{k},\mathbf{R})} \tag{15}$$

Where $|\Psi(\mathbf{k},\mathbf{R})|$ is the known diffraction space modulus. Taking an inverse transform back into real space obtains a new and improved guess of the exit wave function previously consisting of random amplitudes and phases,

$$\psi_{c,n}(\mathbf{r},\mathbf{R}) = \Im^{-1}[\Psi_{c,n}(\mathbf{k},\mathbf{R})] \tag{16}$$

The guessed object wave function in the area covered by the aperture is updated, using an update

function,

$$O_{g,n+1}(\mathbf{r}-\mathbf{R}) = O_{g,n}(\mathbf{r}-\mathbf{R}) + \frac{|P(\mathbf{r})|}{|P_{\max}(\mathbf{r})|}\frac{P^*(\mathbf{r})}{(|P(\mathbf{r})|^2+\alpha)} \times \beta(\psi_{c,n}(\mathbf{r},\mathbf{R}) - \psi_{g,n}(\mathbf{r},\mathbf{R})) \tag{17}$$

Where β controls the amount of feedback in the algorithm and may be varied between 0.5 and 1. Lower values of β increases the importance of the newest estimate of the object function, whereas higher values increase the importance of the previous estimate. α is used to prevent a divide by zero occurring if $|P(r)| = 0$ and acts as a Wiener filter taking the value $\alpha = 0.001$. The update function is crucial to the success of the algorithm since it makes the effective deconvolution that occurs possible. The expression

$$\frac{|P(\mathbf{r})|}{|P_{\max}(\mathbf{r})|} \tag{18}$$

Attenuates the high errors which arise where the illumination is weak as the function favors the influence of those areas of the specimen which have been strongly illuminated and

$$\frac{P^*(\mathbf{r})}{|P(\mathbf{r})|^2} \tag{19}$$

Removes the initial multiplication with the pupil function in generating the exit wave in equation (11). Given that more than one position ***R*** is used, the algorithm has distinct advantages over the Fienup algorithm in that it doesn't suffer problems of uniqueness and that a wider field of view may be imaged.

The importance of the fine sampling of the intensity of the measured diffraction pattern was recognized at an early stage and has lead to the method being referred to as oversampling. Using the Shannon theory in communication(Shannon, 1949):

If a function d(x) is known to vanish outside the points x=±a/2, then its Fourier transform F(X) is completely specified by the values which it assumes at the points X = 0, ±1/a, ±2/a, . . .

Sayre(Sayre, 1952) suggested that the electron-density function *d(x)* describing a single unit cell vanishes outside the points $x = \pm a/2$ and that direct structure determination for centrosymmetric structures could be accomplished by knowledge of $|F|^2$ at half integral spacing. This implies that successful convergence of the algorithm, giving accurate reconstruction of the phase information, can only be achieved with at least a fifty percent overlap of successive probe positions.

The term overlap implies a linear distance offset and not an area overlap. For two circles of radius r and the centre-to-centre distance a, the absolute overlap o_{abs}, is defined as $o_{abs}=2r-a$. Normalized by the diameter of the circles the relative overlap o, is given by,

$$o = 1 - \frac{a}{2r} \tag{20}$$

The advantage of specifying an overlap in distance rather than area is that this is typically the parameter that is specified for a scan over several object positions. The specifications of relative overlap in the following neglect the change in the incident illumination by free-space propagation from the aperture to the object.

The convergence of the algorithm can be monitored by computing the sum-squared error (SSE). In the Fourier domain the squared error is the sum of the squares of the amounts by which $\Psi_{g,n}(\mathbf{R})$, the guessed Fourier transform , violates the Fourier domain constraints $\Psi(\mathbf{R})$,

$$SSE = \frac{\sum_{R} [\,|\Psi_{g,n}(\mathbf{k},\mathbf{R})| - |\Psi(\mathbf{k},\mathbf{R})|\,]^2}{N^2} \tag{21}$$

Where N^2 is the number of pixels in the reconstruction and comes from normalizing the discrete Fourier transform. In practice the error reduction algorithm usually decreases the error rapidly for the first few iterations but much more slowly for later iterations. (Fienup, 1982) Herein the error will be normalized and considered a root mean square (RMS),

$$RMS = \frac{\sqrt{\sum_{R} [\,|\Psi_{g,n}(\mathbf{k},\mathbf{R})| - |\Psi(\mathbf{k},\mathbf{R})|\,]^2}}{N^2 \sum_{R} |\Psi(\mathbf{k},\mathbf{R})|^2} \tag{22}$$

The advantage of this system is that its output is always an image having a Fourier transform that satisfies the Fourier domain constraints. This can be thought of a hybrid input-output algorithm where the input $O_{g,n}(\mathbf{r}-\mathbf{R})$ can be considered as the driving function for the next output $O_{g,n+1}(\mathbf{r}-\mathbf{R})$. This viewpoint allows for a great deal of flexibility in selecting the next input and allows more rapid convergence to a solution.

Test images used to model the algorithms ability to reconstruct accurately are based on real AFM images generated in the London Center for Nanotechnology (LCN). These have the required periodicity to achieve realistic diffraction patterns from which the model can be refined. Herein all test images will consist of those shown in Figure 8.

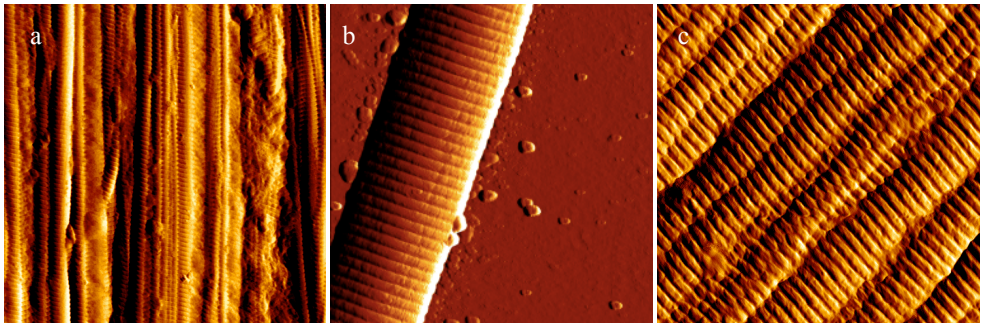

Figure 8: Test images used in algorithm refinement a) AFM of collagen fibre, showing fibrils overlapping and not perfectly aligned b) individual collagen fibril showing clear D-banding c) well aligned sample of collagen fibre, by provided by Cat McCallion (UCL).

A range of relative overlaps were tested using the RMS error to find a compromise between the number of iterations until sufficient convergence and the time lapsed until that convergence. An increase in relative overlap leads directly to an increase in the radiation dose experienced by the sample. As biological samples are of interest in this investigation, limiting radiation damage experienced by the specimen while achieving sufficient reconstruction also needs to be considered. Although the algorithm run times seem comparatively small, this will increase dramatically as the array grows in size when 3D specimens are considered later.

A clear view of the RMS error based on relative overlap along with the associated time to complete can be compared in Figure 9. The oversampling requirement of half integer steps can clearly be seen. With a relative overlap of 10% the RMS error stagnates and convergence to an accurate value is unlikely given it is seven orders of magnitude larger than the smallest RMS generated by different relative overlaps.

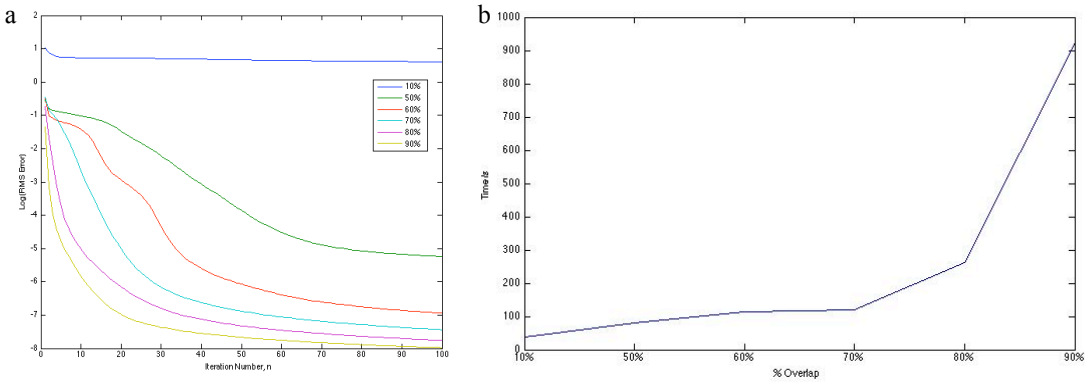

Figure 9: Log(RMS) error shown in a) which clearly shows convergence for the higher relative overlaps and b) shows the corresponding time to complete the reconstruction.

On the consideration of radiation dose, time for reconstruction and convergence of reconstruction algorithm, a relative overlap of 60% will be assumed throughout. This is consistent with other investigations.(Bunk, 2008) After 100 iterations the object was retrieved as shown in Figure 10 and at 60% overlap a RMS of 9.651×10^{-4}.

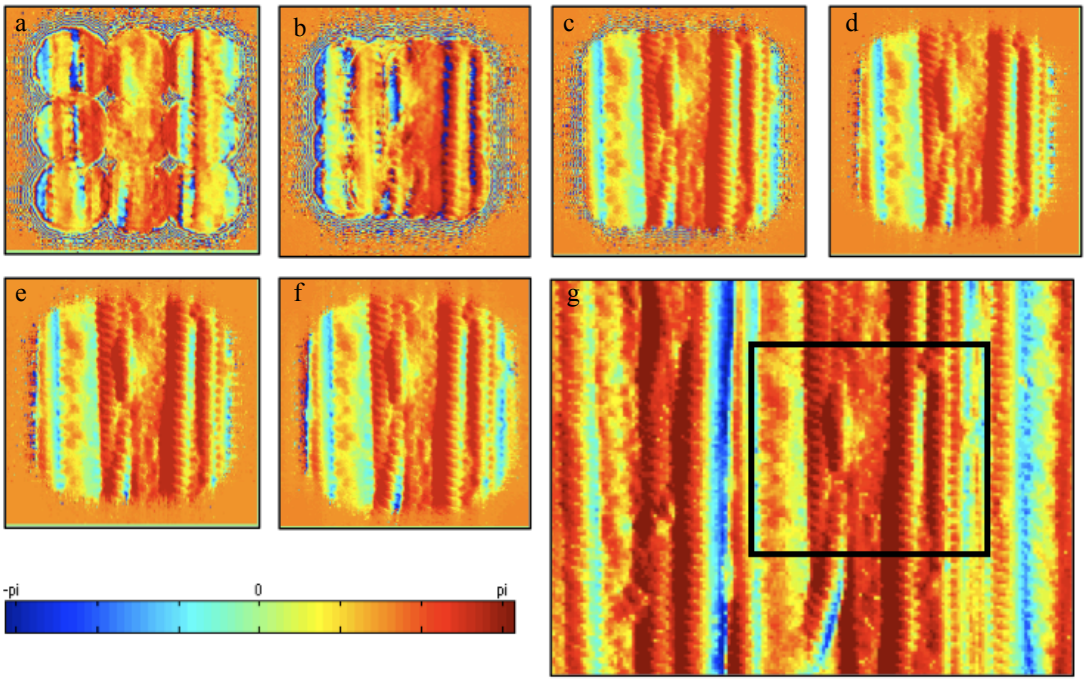

Figure 10: Reconstructions of simulated data for different relative overlaps. Each image has been reconstructed in 100 iterations, covering the same area. The relative overlap was a) 10% b) 50% c) 60% d) 70% e) 80% and f) 90%. The initial start image to be reconstructed is highlighted in g).

Development of the Pupil Function

To further test the algorithms robustness in reconstructing a real space image from a set of Fraunhofer diffraction patterns a realistic pupil function needs to be approximated. The pupil function is convoluted with the object function (specimen) to give the resulting exit wave. Starting with a simple top hat function of a square and circular pupil, propagators were employed to model the beam as it traveled in free space from the aperture to the specimen.

To model the propagation of a plane wave through free space a propagator is used. A diffracting aperture is assumed to lie in the (ξ,η) plane, and is illuminated in the positive z direction. The wavefield across the (x,y) plane which is parallel to the (ξ,η) plane and at a normal distance z from it.

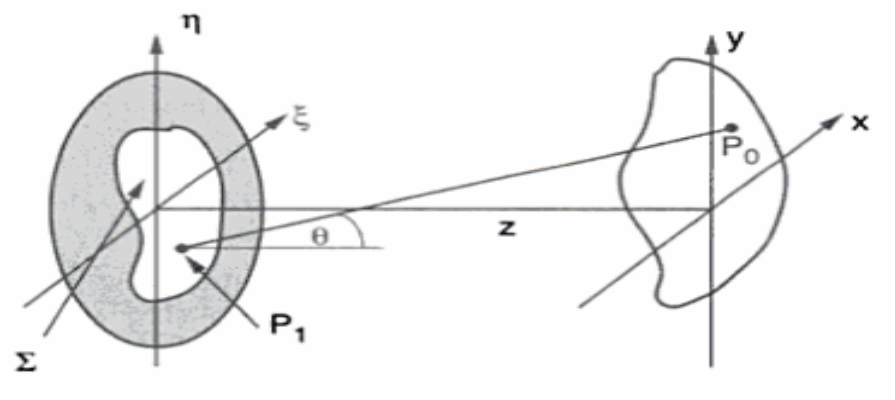

Figure 11: A schematic diagram showing the basic setup employed in propagating a plane wave through free space. Here Σ *is the aperture function and* P_0-P_1 = $\mathbf{r}$.(Goodman, 2005)

The z-axis pierces both axes at their origin. The pupil function at the object $P(\mathbf{r}_o)$ is characterized by,

$$P(r_o) = \frac{e^{ikz} e^{i\frac{kr_o^2}{2z}}}{i\lambda z} \int_{-\infty}^{\infty}\int \left\{ P(r_a) e^{i\frac{kr_a^2}{2z}} \right\} e^{-i\frac{kr_{oa}^2}{z}} d\xi d\eta \tag{23}$$

Where $P(\mathbf{r}_a)$ is the aperture function at $z = 0$, $r_o^2 = (x^2+y^2)$, $r_a^2 = (\xi^2+\eta^2)$ and $r_{oa}^2 = (x\xi+y\eta)$. Which is recognized as the Fourier transform of the product of the complex field just to the right of the aperture and a quadratic phase exponential. However given that the Fresnel Number, F<< 1 in the Fraunhofer plane, an angular spectrum approximation is used to model the free space propagation of the wave,

$$P(r_o) = \frac{e^{ikz} e^{i\frac{kr_o^2}{2z}}}{i\lambda z} \int_{-\infty}^{\infty}\int P(r_a) e^{-i\frac{kr_{oa}^2}{z}} d\xi d\eta \tag{24}$$

Which is simply the Fourier transform of the aperture function evaluated at frequencies $f_x=x/\lambda z$ and $f_y=y/\lambda z$, including multiplicative phase factors. Thus starting out with simple real, square and circular apertures, a complex pupil function at the specimen can be modeled.

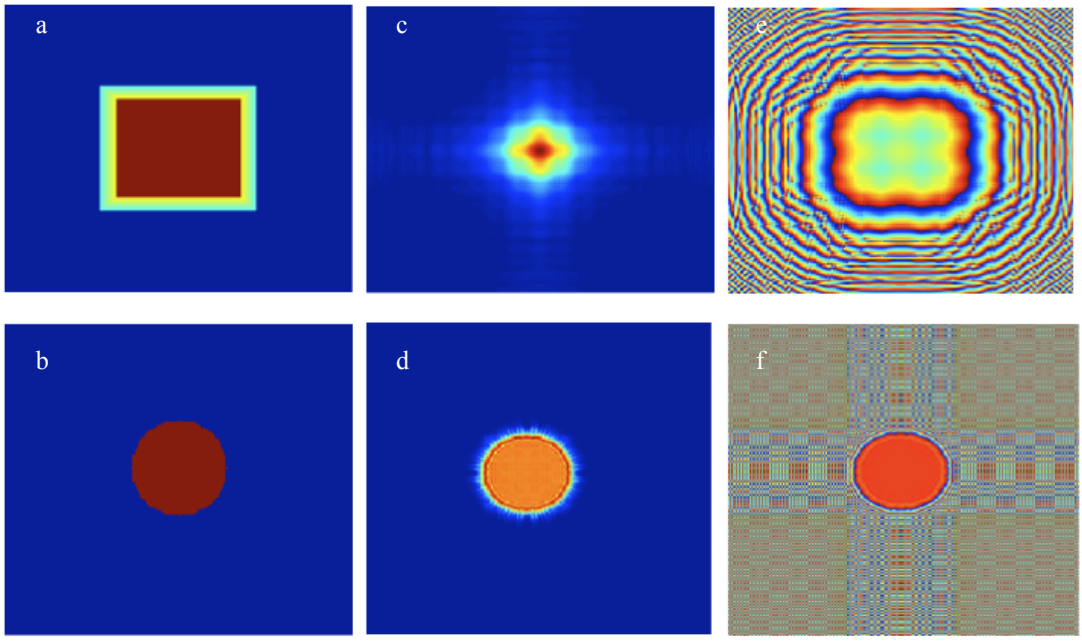

Figure 12: Initial real pupils a) Square pupil with a edge described by a Fermi function b) circular pupil, both with a maximum value of 1 c) & d) show the amplitude of the pupil function just before the object and e) & f) show the phase of the pupil.

The square aperture had roughness modeled following a Fermi function around its edge. This mimics the inability to achieve a completely true edge on a slit. Apertures with a 50-pixel radius are generated corresponding to 25μm in real space. The value of 25μm was chosen, as this is the smallest aperture size available on the beamline when real data was taken. This will be discussed in the next section. The distance from the aperture to the sample, z =1.455m. The X-rays have energy of 8.5keV.

Suitability of Algorithm on Complex Functions

As the function describing the sample is unknown but complex, a range of initial images are tested to observe the convergence of the algorithm on a) real functions b) phase only functions and c) complex functions. All test functions used to generate the diffraction patterns which will be called in the reconstruction algorithm are created by

$$O(\mathbf{r}) = A(\mathbf{r}).*\exp(i\theta(\mathbf{r})) \tag{25}$$

Using the test image in Figure 8c, real, phase only and complex functions were created and reconstructed observing convergence.

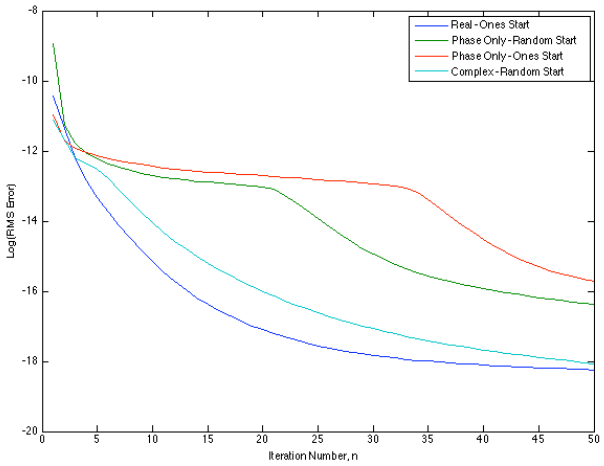

Figure 13: RMS error graph showing how the algorithm converges with different start images with different start guesses. Notice how starting with a real value for a phase image yields a higher error.

It was observed that a real function was reconstructed more accurately when initiated with a real guess function and that phase only functions were reconstructed more accurately when a random phase with constant amplitude was used. The algorithm will converge when initiated with a real function but seems to limit the range of phase reconstruction to $\pm\pi/2$. This may be due to the inability of the algorithm to reduce the real component to zero to achieve a phase only function.

From the RMS error, a real guess on a phase only function produces a higher RMS error that a complex guess for a phase only function. Using the same complex guess on the complex reconstruction shows convergence to that of the RMS error on the reconstruction of the real function. The reconstructions were initiated with a random phase and constant amplitude function, having the value `0.75+5.97*rand*i`. This value was deduced from running the algorithm on a known phase object and adjusting the values accordingly.

Dark Field Imaging

When we discuss the generation of diffraction patterns for use in the reconstruction algorithm, these are of course highly idealized perfect Fourier transforms accessing the full spectrum of diffraction maxima. However employing the use of small angle x-ray diffraction techniques limits the field of view to the first diffraction maxima omitting the zeroth and higher orders. Information from the zeroth order is restricted because of information being obscured behind the beam stop required to prevent camera damage from unscattered radiation.

Using a single diffraction order in the reconstruction algorithm to form an image causes all the regions of the specimen not of the same crystal structure to be represented as dark regions in the final reconstructed image. This forms the basis of dark-field imaging.

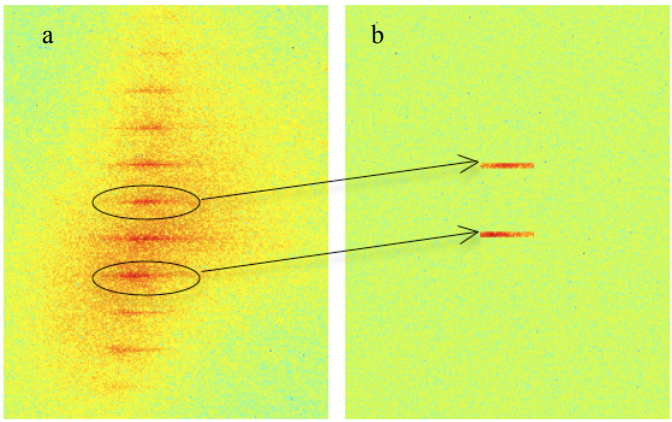

Figure 14: Log view of Fraunhofer diffraction pattern a) Idealised theoretical Fourier transform used in image reconstruction and b) the information we can use in the real world to reconstruct an image.

To isolate the dark field pattern the full range of diffraction patterns is observed and the relative positions of the first reflections recorded. These are then isolated and padded with random values to mimic that observed in the background.

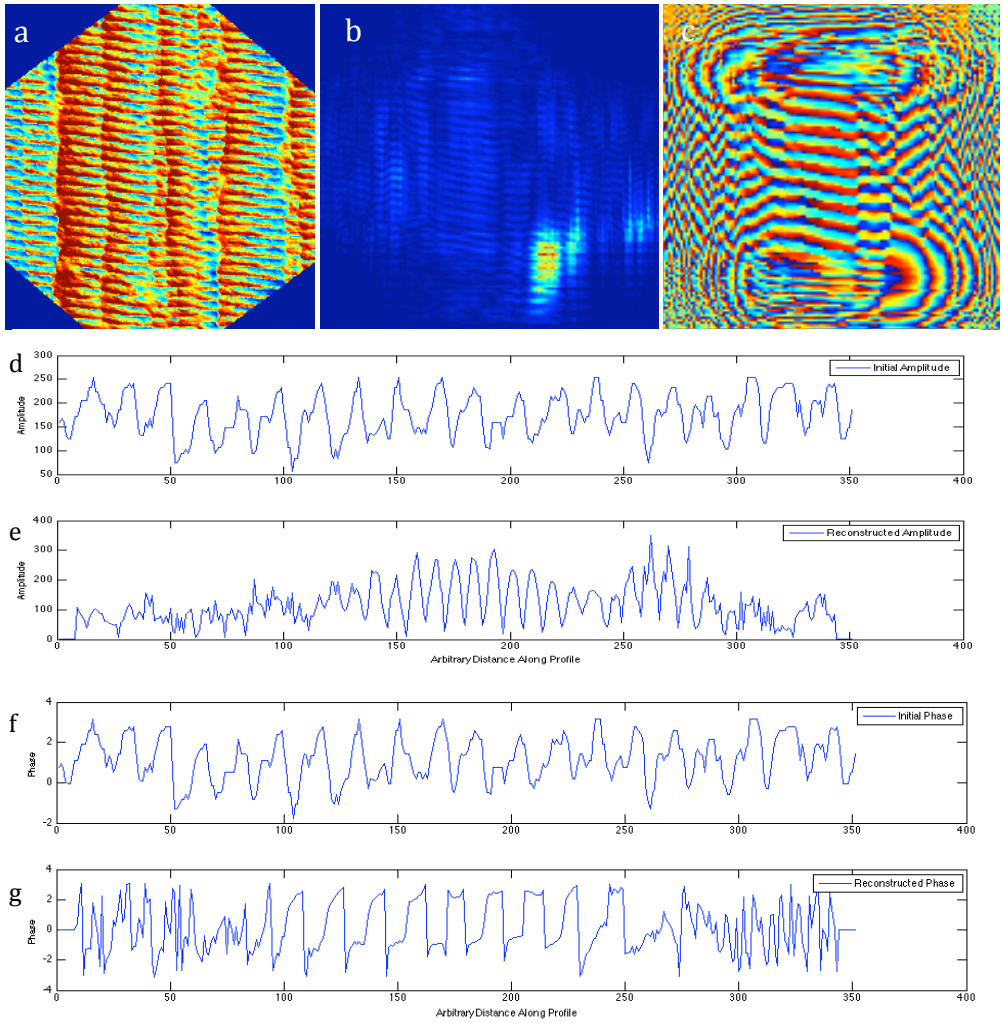

Figure 15:a) Initial start image used to generate diffraction patterns, both the phase and amplitude were constructed with this image with a phase range ±π b) Reconstructed dark field amplitude c) Reconstructed dark field phase d) shows a section through the initial amplitude with e) the corresponding reconstructed amplitude f) shows a section through the initial phase and g) the corresponding reconstructed phase.

The important thing to consider in the amplitude reconstruction is the shape of the reconstruction and the relative positions and heights of the peaks along the profile. Only the periodicity of each collagen

fibril will be reconstructed in the dark field regime. Individual fibrils with D-banding can clearly be distinguished from the dark background. The phase reconstruction shows peaks in same relative position and the magnitude of the phase is consistent with the original start image within error.

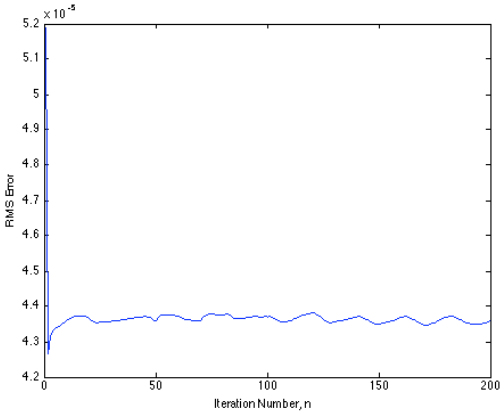

Figure 16: RMS error of dark field reconstruction showing the oscillatory behaviour around the convergence point. Better convergence may be observed at higher iteration numbers.

The convergence oscillates about the point 4.2×10^{-5}, however this may iron out at higher iteration numbers.

Collection of Raw Data and Analysis of Bragg Peak Information

Data was collected at the non-crystalline diffraction beamline I-22 at the Diamond Light Source (DLS), Didcot, United Kingdom. The DLS is a third generation synchrotron providing high brilliance X-rays achievable by the use of an undulator insertion device. Figure 17 shows a schematic diagram of the beamline setup.

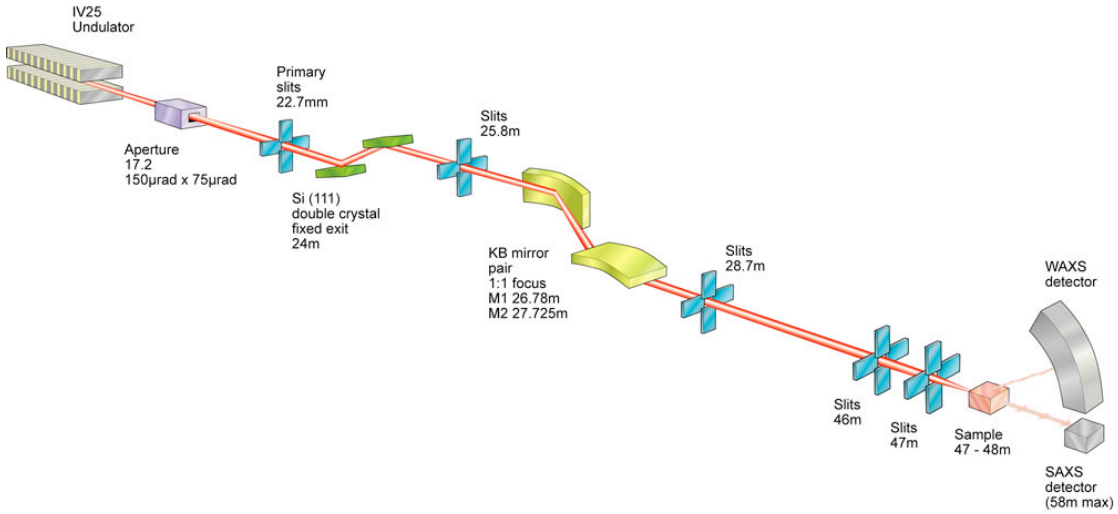

Figure 17: Schematic diagram of beamline I-22, showing the monochromator and the focusing mirrors. The source aperture s3 set at 200µm and the pupil s4 set at 25µm. (Fletcher, 2008)

A wavelength of 0.1462nm was selected by a Si(111) double crystal monochromator and focused using a pair of Kirkpatrick-Baez Mirrors at 26.78m and 27.725m from the undulator source. (Fletcher, 2008) Slits at 28.7m defined the beam size at 200µm x 200µm. The beamline has an energy resolution $\Delta E/E = 1.2 \times 10^{-4}$. A 6mm diameter lead backstop was placed in front of the detector to prevent unscattered radiation damaging the CCD array. Using the backstop as a reference, the real space resolution of the detector was determined to be ~20µm.

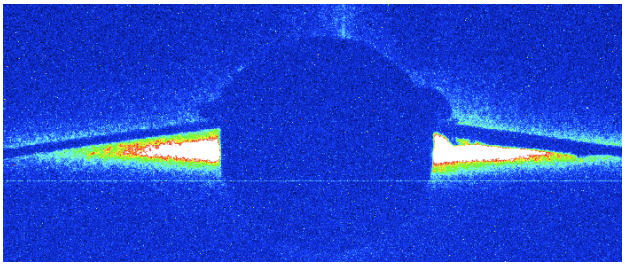

Figure 18: Lead beamstop in front of CCD detector being held in place by guide wires. The zeroth order and unscattered radiation can be seen either side of it.

In order to resolve the speckle the incident beam size must not be much greater than the transverse coherence length of the beam given by $\lambda R_s/d_s$, where R_s is the distance from the source and d_s is the aperture size.(Sutton, 1991) Using Shannon's theorem, a speckle needs to illuminate at least 2px on the CCD camera in order to usable in a reconstruction. Using the real space pixel resolution, a minimum aperture size was found to be ~24μm.

The coherence was defined with the s4 slits, set to 25μm as it was not possible to set the gaps more precisely than 5μm and so this was the smallest practical size. As this is slightly more than the coherence length, we attempted to close the slits at s3 to compensate, however the difference in contrast was inconclusive.

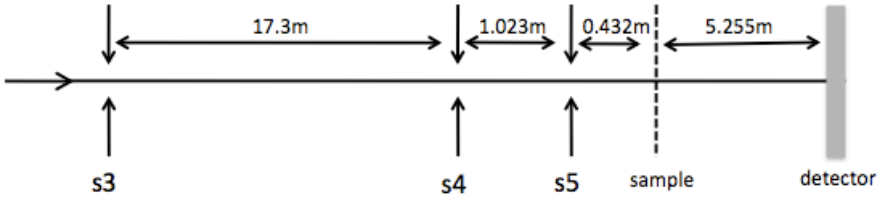

Figure 19: Close up schematic of beamline from s3 slits showing the spacing between slits and the sample.

The s5 slits placed 0.432m before the sample were used as a guard, set to cut the node in the Fresnel diffraction of the s4 slits. Without this, the experiment would have failed due to slit flares far stronger than the signal under investigation.

The beam propagates from the s4 slits to the sample at the standard SAXS position 1.455m away. A Princeton Instruments PI-SCX: 1300 x 1340, 20μm x 20μm air cooled CCD camera positioned 5.255m behind the sample records intensity in a 1x1 binning mode. The CCD camera was cooled to -30° and had a low readout noise. This resulted in a speckle fringe spacing of about 4 pixels, sufficient to assure oversampling of the data. The data was recorded in WinView.

Preparation of Collagen Samples

We used fresh samples of rat-tail collagen, using a trained expert in the dissection to surgically remove the skin, bones and hair from the rat-tail. Dissection was performed using a surgical scalpel and sharp surgical tweezers to isolate bundles of fibrils or lengths of a single fibril. Extracted bundles were then rinsed with deionised water and stored in phosphate buffered saline (PBS)/azide solution. The PBS buffers the solution with the azide kills off any bacteria and prevents the bundles from degrading.

Samples were pulled into alignment using surgical tweezers and placed between two layers of kapton film. The hydrophobic nature of the kapton film necessitated the use of the second layer forming a "sandwich". This method has been successfully implemented in the past to obtain good quality X-ray diffraction patters. (McCallion, 2008)

A range of liquid sample cells were available from the beamline laboratory. PBS filled chambers allowed for a wet environment, which not only provided more native surroundings for the samples but also was thought to minimize radiation damage by improving heat dissipation. This also meant that the natural structure of the fibril bundles could be maintained better.

The cell consisted of a brass body with a window for a mica slide, a space for a washer, a second mica slide and a brass frame. The sample was placed between the first mica slide and the washer and the space filled with PBS solution before placing the second mica slide on top. A brass frame was then used to secure the cell. Both mica cells were 25μm thick. It was advantageous to use the liquid cells as early X-ray diffraction patterns proved that the shape and dimensions of the collagen fibrils changed in hydrated environments. (Bear, 1942)(McCallion, 2008) Samples stored for more than two days in kapton were completely dry and showed meridional reflections corresponding to 64nm D-banding.

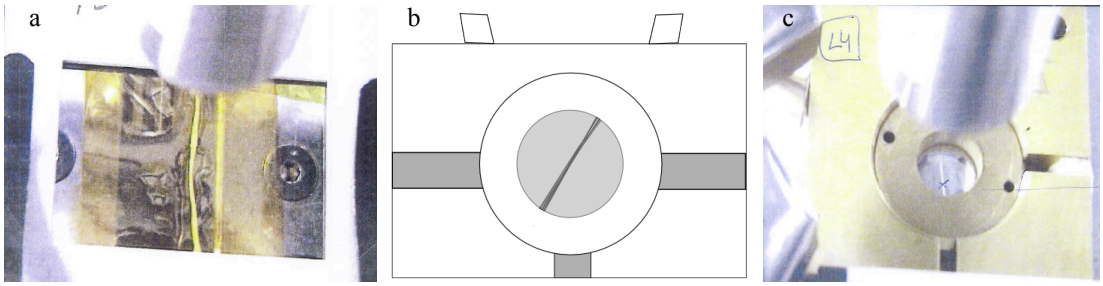

Figure 20: Collagen samples held in beamline I-22 a) collagen bundle in sandwiched between kapton film being held in a picture slide b) a schematic of the liquid cell and c) actual with "X" marked on collagen to position the sample in front of the aperture. (Robinson, 2008)(McCallion, 2008)

Early attempts using the liquid cells failed as the PBS solution attenuated the beam too strongly and no readout was recorded on the CCD camera. However a modified liquid cell with a smaller volume of PBS solution decreased the optical path length.

A sufficient set of data was collected which clearly showed the reflection from the 67nm D-banding. This is easily seen in Figure 21. The first reflection was recorded at 575px above the unscattered beam, gives a real space distance of 11.5mm. Using equations (6) & (7) with z = 5.255m and Z = 11.5mm corresponds to a value of d = 66.8nm

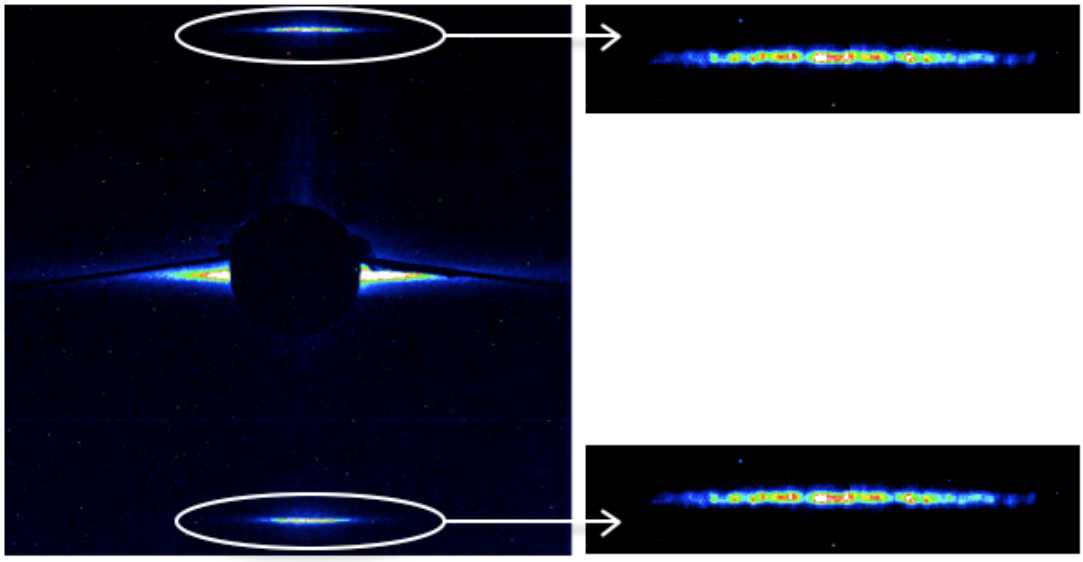

Figure 21: Sample diffraction pattern from rat-tail collagen in liquid cell, showing a magnified image of the first meridional reflections with clearly defined speckles.

The horizontal width of the reflection is related to 1/*width* of the collagen fibril whilst the vertical height of the reflection is related to the 1/*height* of the fibril illuminated by the pupil. The speckles size is related to 1/*beam size*. Taking profiles along the transverse and longitudinal lengths of the speckle patterns allows more detailed information regarding the structure under investigation to be determined.

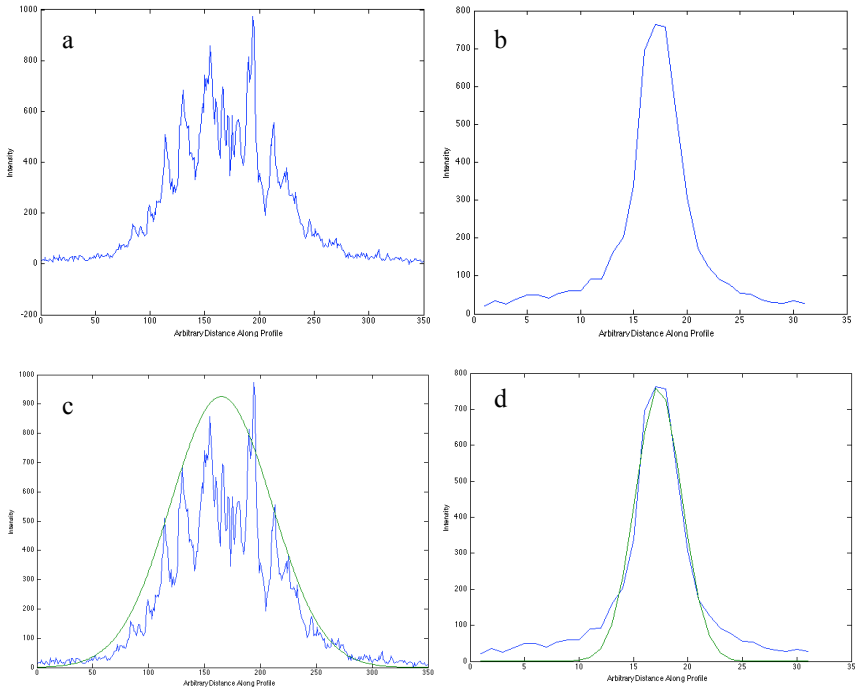

Figure 22: Profile through first meridional reflection showing a) transverse and b) longitudinal. Gaussian fits to the transverse in c) and longitudinal in d).

The fraying that broadened the diffraction in previous attempts was suppressed by using the think end of the tendon without removing the skin. The speckles are confined to a single line effectively making it a 1D pattern. This means that the phase shifted fibrils making up the tendon are perfectly parallel to each other over the area illuminated by the beam. The reflections are also in a straight line suggesting there is no angular distribution of fibres otherwise a curved diffraction plane would be observed.

The Gaussian was fit over the profile using,

$$f(x) = a_1 * \exp(-((x - b_1)/c_1)^2) \tag{26}$$

For Figure 22c: a_1=925, b_1=165, c_1=64 and for Figure 22d: a_1=765.4, b_1=17.3, c_1=3.031. Using this information a FWHM can be found for the transverse and longitudinal dimensions of the reflection and the speckle size related to the beam size.

Transverse FWHM = 106px – 1/(2.12mm)
Longitudinal FWHM = 5px – 1/(100μm)
Average individual speckle = 2.6px -1/(52μm)

The number of speckles is directly related to the number of domains (fibrils) the beam illuminates as it traverses the sample. In this case the number of speckles ~30. The structures within the wide peaks are imperfectly resolved X-ray speckle patterns, formed because two different domains introduce phase shifts into the scattered radiation. (Brauer, 1995) This will be compared with the theoretical model in the next section.

Modelling Collagen Fibrils for Simulations

Initial a very basic collagen model was created using simple values of 0 and 1. Later a more realistic collagen model was constructed using data from AFM images to mimic the results obtained from real data, initially in 2D and then 3D to test the algorithm and the likely output of any reconstruction. Using the test image in Figure 8c, from an AFM, a single "collagen unit" was constructed by observing the profile of an individual collagen fibril. Each pixel in the reconstruction corresponded to 4nm in real space. Thus the collagen unit had dimensions 17x67px.

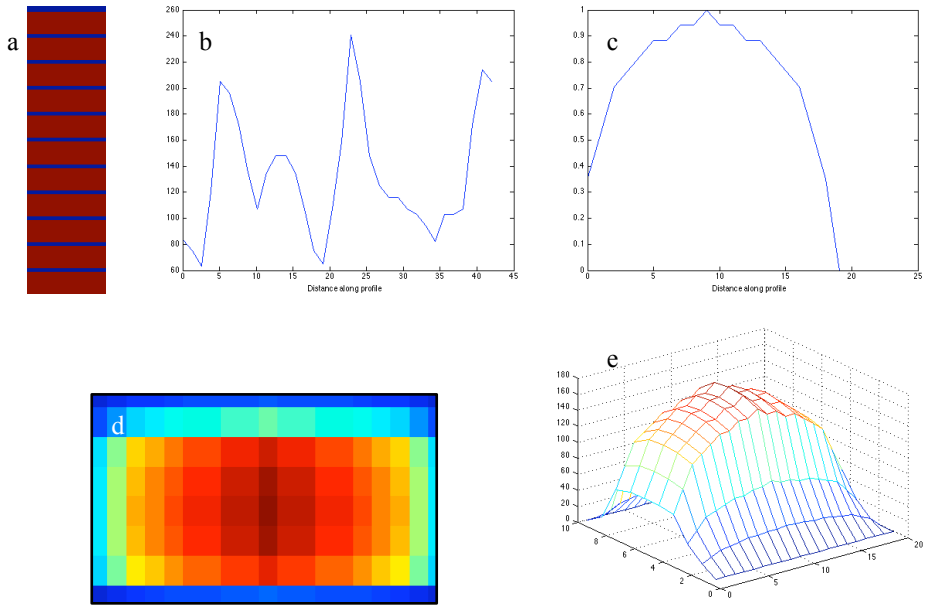

Figure 23: a) Simple model of single fibril and a profile through a single collagen fibril showing b) longitudinal profile and c) transverse profile with d) the reconstructed 2D unit and e) the 3D unit.

Using the "collagen unit" individual fibrils could be assembled into bundles of fibrils not dissimilar to the samples used to take real data. A small routine was made in which collagen fibrils were randomly displaced from the first collagen fibril (reference fibril) in the y-direction. The amount of displacement from the reference fibril directly related to the relative phase attached to each fibril. Since each D-period represented 17px, if a neighbouring fibril was displaced by 5px below the reference fibril, the corresponding phase attached was 5/17 x π radians, in the positive direction. A negative phase was given if the neighbouring fibril was displaced above the reference fibril.

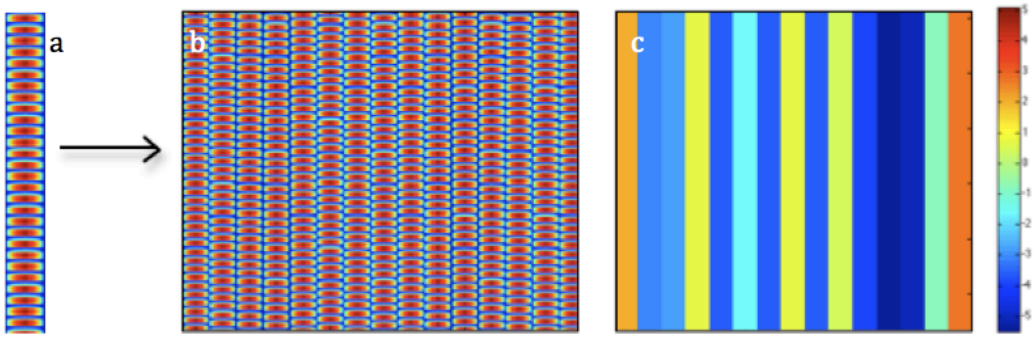

Figure 24: Simulated collagen model a) single collagen fibril b) being built up into a collagen bundle with c) corresponding phase assigned to each fibril with the corresponding range shown.

Generating a diffraction pattern in the Fraunhofer plane using this simple 2D model does not generate realistic results but does mimic the periodicity shown from the real data and the first meridional reflection can easily be seen. The routine needed to be adapted to take very general values depending on the size, thickness and the dimensions of the individual "collagen unit". Thus a 3D model was created. The code for this can be found in Appendix A.

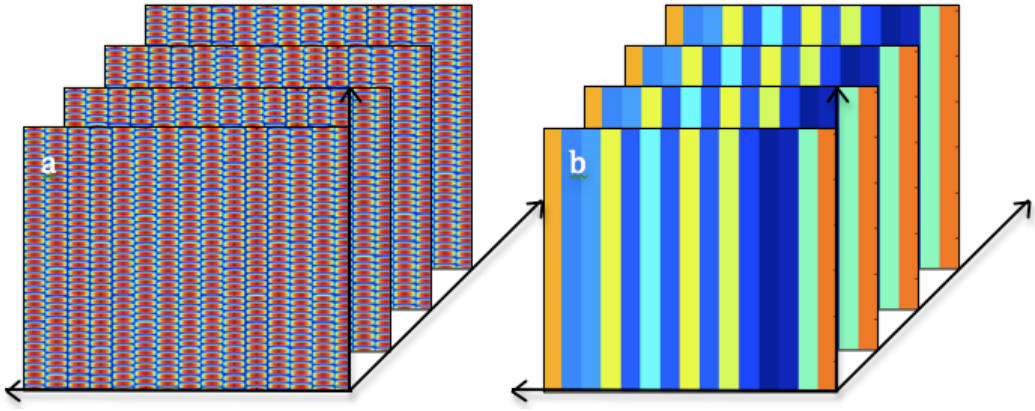

Figure 25: Schematic of 3D collagen model build up showing a) build up of amplitude and b) build up of phase.

Again each phase in each layer was offset from the reference fibril in the left hand position in the first layer. The routine could now handle fibril bundles or any dimension size is only limited by the amount of memory available. Some random noise could be added to the 3D model to characterise small beam instabilities or noise from the readout of the CCD. Comparisons were made each time a refinement was made in the model and compared with the general structure of the real diffraction pattern produced. The diffraction pattern produced by the model was strongly dependant on the size of the pupil used. The advancement of the diffraction pattern produced by each model can be seen in Figure 26.

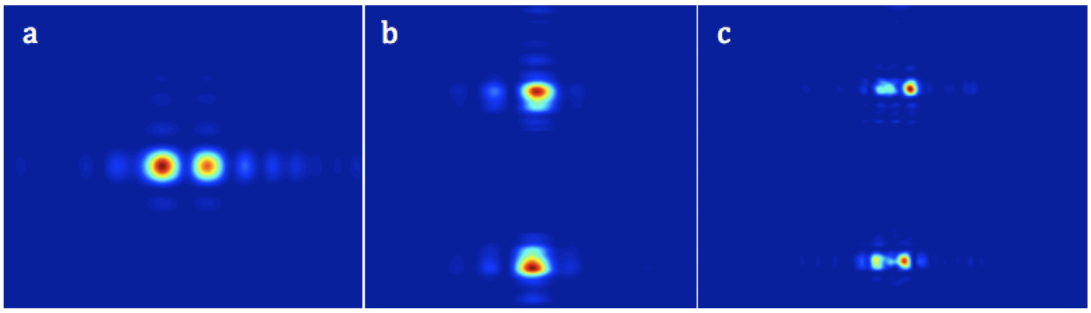

Figure 26: Selection of diffraction patterns from different models a) diffraction pattern made from a 1D layer made up of blocks of Figure : b) diffraction pattern created from Figure and c) diffraction pattern from Figure . Figures b) & c) have had the unscattered region in the centre removed so the speckles have a better contrast.

The most realistic diffraction pattern is created using a 3D model. The 3D model was created with 20 layers although this could be varied to any size, with each of the layers containing 9 collagen fibrils. Using the kinematical approximation described before, each layer was convoluted with the pupil function i.e. layer 1 was multiplied by the pupil, layer 2 was multiply by the pupil etc. and the convolution of the pupil and each layer summed to give a 2D array which was then Fourier transformed. This scattering method provided the most realistic and useful results.

However if the speckle size generated by the routine were to be comparable to those collected at the synchrotron source, an amount of padding needed to be added to the initial start image. The padding is necessary since the simulated sample has dimensions on the order of nm and the pupil is on the order of μm. The amount of padding is calculated by,(Berenguer, 2008)

$$padding = \frac{d_{CCD}\lambda}{\rho_{CCD}\rho_{AFM}} \tag{27}$$

Where d_{CCD} is the CCD distance = *6.71m*, λ=*0.1462nm*, ρ_{CCD} is the resolution of CCD = *20μm* and ρ_{AFM} is the resolution of the AFM = *4nm*, giving an overall padding ~12000 i.e. any start image must be padded with zeros until the array has dimensions 12000x12000 if comparable speckles are to be created. This value is just too large to be handled by a standard computer and so the resolution of the AFM image is adjusted to *16nm* giving an array size of 3200x3200. This means that the speckles generated in the pattern will be 4 times larger than those from the synchrotron. The effect of this padding can be seen in Figure 27.

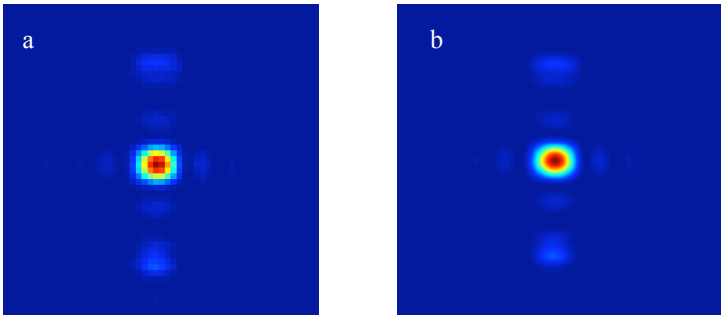

Figure 27: Diffraction pattern generated from the same image with a) no padding and b) padding with zeros creating an array size of 3200x3200.

Creating a simulated collagen fibre bundle which has significantly more fibrils in each layer compared with the amount of layers in the sample enables investigation of where the speckle comes from. A model fibre bundle with 20 layers with 50 fibrils per layers was modelled and the transverse and longitudinal profiles of the first reflection in the diffraction pattern investigated.

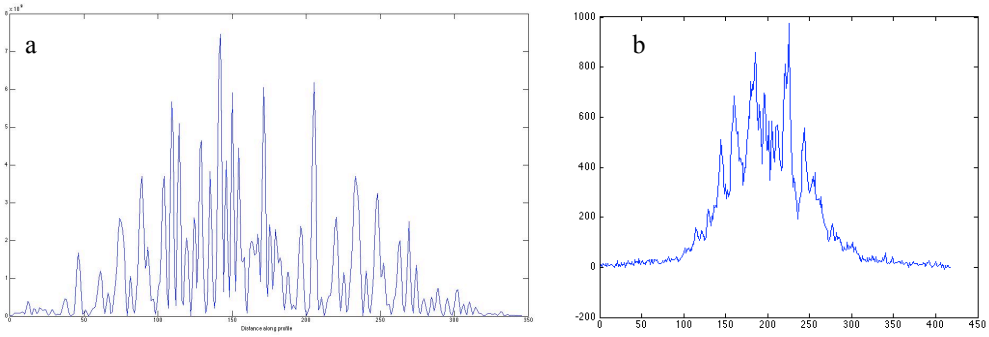

Figure 28: Transverse profiles through a) simulated model showing ~50 individual speckles corresponding to the number of speckles illuminated. Note the intensity scale is arbitrary due the initial values of the amplitude of the initial collagen fibril and b) the actual profile through raw data collected at the synchrotron. This is shown again for ease of comparison.

It can be seen from Figure 28a that there are significantly more than 20 speckles and if they are counted individually there are ~50 suggesting that our assumption made in the last section that the speckle number is in direct relation to the number of domains illuminated. However in the theoretical model, considerably more peaks tend to a minimum of 0 compared with the data from the synchrotron. This is a measure of the visibility and is defined by,

$$V = \frac{(I_{\max} - I_{\min})}{(I_{\max} + I_{\min})} \tag{28}$$

For homogeneous specimens with negligible SAXS contribution the value the visibility remains unchanged. For specimens that show strong internal density fluctuations on the micrometer length scales and produce a strong SAXS signal show a significant decrease of the visibility and yield values of V<1. (Pfeiffer, 2008) More generally the visibility is inversely proportional to the effective integrated local small-angle scattering power of the sample. For the simulated collagen bundle V=1, whereas for the real data the average V~0.48. Given this there are a number of reasons why the model generated deviates from the real sample. The reduced visibility could be due to internal structure which is neglected in the simulation. To investigate this further higher order reflections could be recorded to investigate whether the same visibility reduction is present in the WAXS.

The model assumes a 100 per cent coherent source which may not be the case as demonstrated in the previous section, where the coherent length of the beam ~22μm and the minimum usable aperture size was 25μm. Thus instabilities in the beamline could be portrayed in the reduced visibility.

3D Dark Field Imaging and Application to Real Ptychography Scan

As discussed at the beginning, the dark field removes the unscattered beam and reconstructs an image from higher order reflections. In order to do this effectively for a 3D model a command line was built into the routine to remove the central spot which although contains information cannot be used to reconstruct an image with the real data. Figure 29 shows the four successive ptychography diffraction patterns from a simulated 3D collagen bundle. The speckles are similar as the area is oversampled but also different as the pupil function moves across the image.

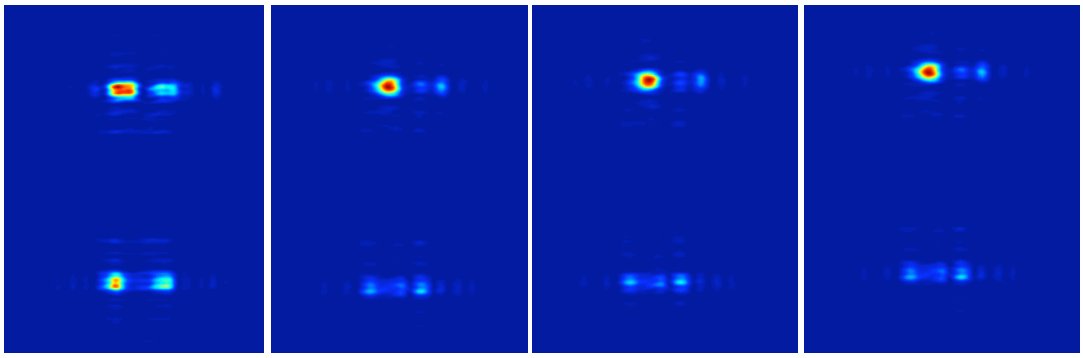

Figure 29: Progressive of ptychography diffraction scan on simulated 3D image using a 50px pupil corresponding to 25μm in real space. Speckles are similar but at the same time different as the pupil is moved across the sample.

Running the reconstruction algorithm with the padding described in the last section is very memory intensive as nine million Fourier transforms are computed per pupil position. This leads to a limited field of view which can be reconstructed using a desktop computer, with each iteration taking ~ 15 minutes. However even with this limited field if view a good reconstruction can be achieved.

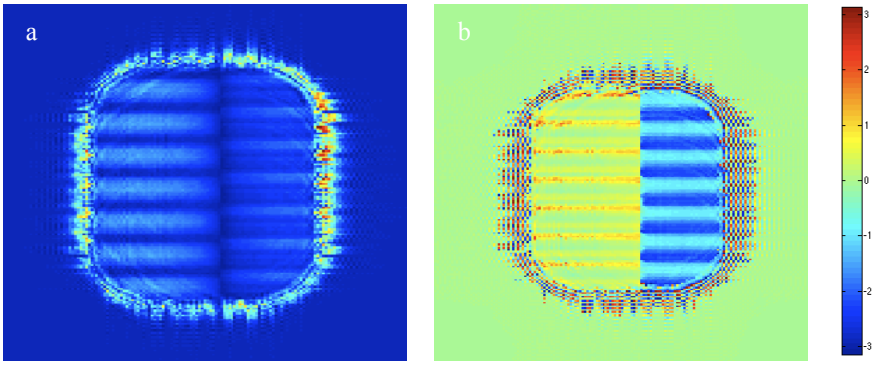

Figure 30: Dark field imaging showing a) amplitude and b) phase reconstruction after 50 iterations. Notice that anything that doesn't contribute to the first reflection is portrayed as a dark area.

The dark field can clearly be seen corresponding to areas which contributed to the unscattered or higher order reflections. The D-banding can be seen in both the amplitude and phase reconstruction. Taking a longitudinal profile of both generates more useable information.

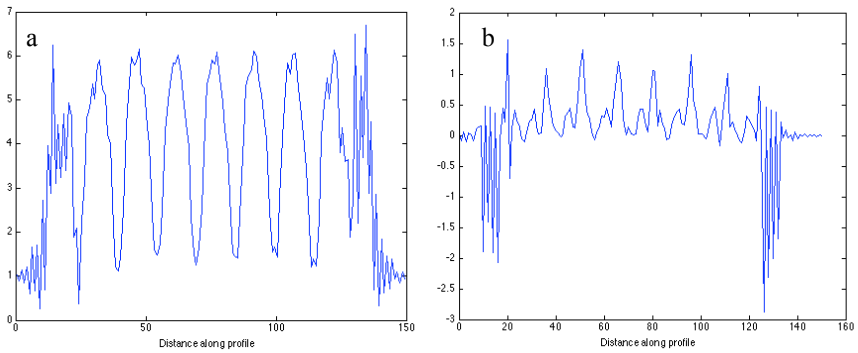

Figure 31: Profile through a) amplitude reconstruction with y-axis intensity and b) phase reconstruction with y-axis in radians.

There are a number of observations to consider from the profile sections through the reconstructed image. Although the initial start image had solid bands of phase, periodicity appears in the reconstruction due to zero values in the amplitude. By simple inspection, the frequency of the periodicity in the phase reconstruction is twice that of the periodicity of the amplitude reconstruction. This could be due to the summing of each layer before taking the Fourier transform which would confirm the small dips in the peaks of the amplitude reconstruction suggesting that the resolution is too low. The fact that there are no negative values for the phase reconstruction suggests that the layer sum has a base at zero. Sampling a different fibril from the bundle shows different base values. The observations seen in this 3D reconstruction will be useful in interpreting real data taken from the synchrotron.

A number of ptychography scans were run. The scan described below was run with an aperture size of 25μm and a step size of 10μm satisfying the oversampling requirement. The motor which moved the sample and the recording software were unable to be synchronised and so the following approach was adopted,

wait time	+	motor movement	=	exposure time	+	readout
30s	+	0.45s	=	28.65s	+	1.8s

A total of 66 frames were recorded but only every third frame was used in the reconstruction. With the central beam stop and the unscattered radiation being removed. The amplitude and phase reconstruction can be seen in Figure 32.

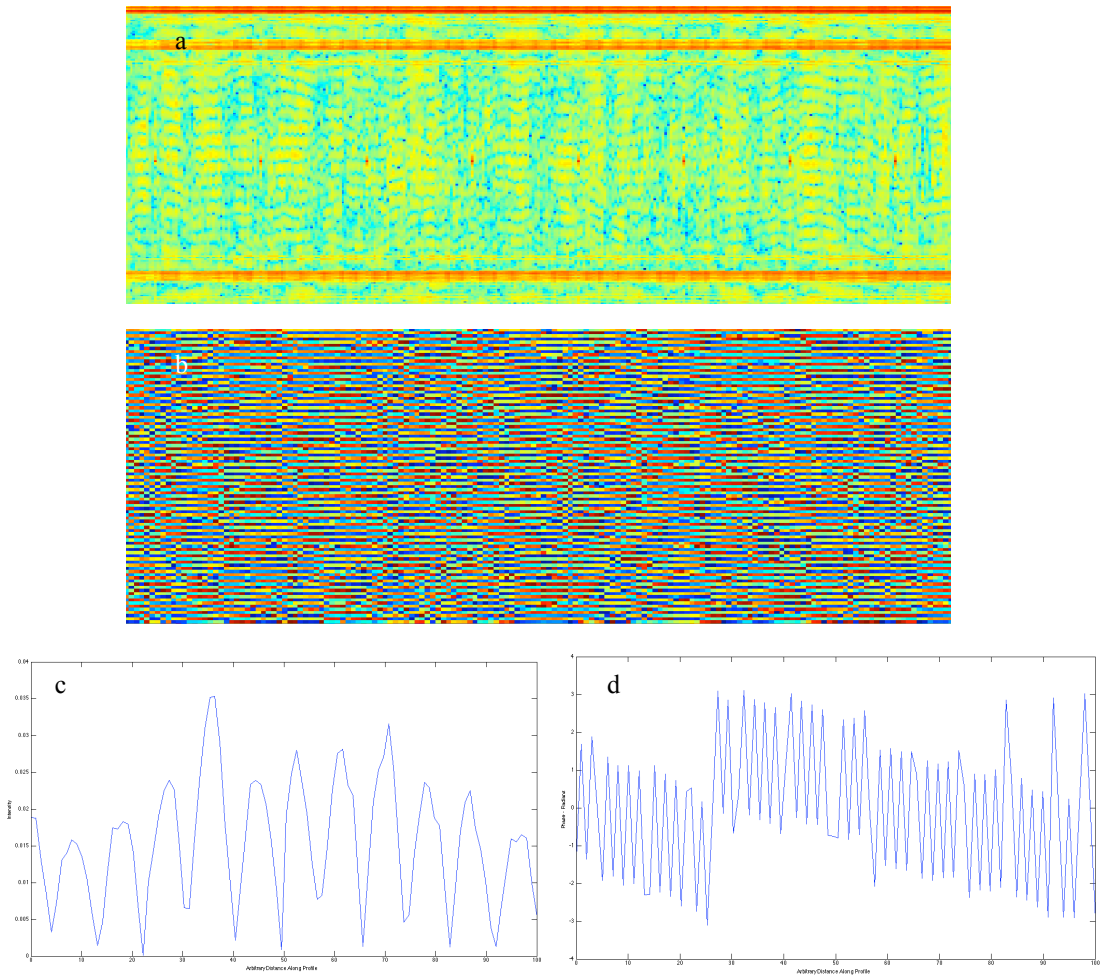

Figure 32: Reconstruction of real ptychography scan showing a) amplitude and b) phase reconstruction. The profile of each in c) & d) respectively.

Regular banding can be seen in the amplitude reconstruction and a profile through the phase shows a regular arrangement. However like the 3D simulated model, the frequency of the phase periodicity is higher.

The resolution of the reconstruction is determined using the following equation,

$$resolution = \frac{d_{CCD}\lambda}{\rho_{CCD}N} \tag{28}$$

Where d_{CCD}, λ, ρ_{CCD} are the same as those described in equation 27 and N is the number of pixels in the x- and y- directions. This gives a resolution in the x-direction of 28.66nm and 29.55nm in the y-direction. Peak to peak distance in the amplitude is on average 9px giving a real space distance of 265nm and in the phase, peak to peak distance is on average 2.2px, giving a real space distance of 65nm. The periodicity in the phase closely matches that of the D-banding, whereas the amplitude periodicity is ~4 higher than expected. This could be due to the kinematical approach adopted in the summing of the individual layers of the fibrils and requires further investigation.

The range of reconstructed amplitude values is quite low with an average height ~ 0.36nm. This could be due to a low intensity pupil function used in the reconstruction or an inaccurate shape assumed for the pupil at the sample. Both these require further investigation.

Given the difficulty in acquiring the raw data the initial reconstructions have be seen to give reasonable results. With further data acquisition and an ability to perfectly synchronise the motor which moves the sample and the recording software, a greater leap could be made in establishing the phase along a length of a single collagen fibril.

Conclusion

The robust theory underlying ptychography suggests that this simple method can be successively employed to recover the phase information along a single collagen fibril. The use of biological rather than crystalline samples introduces additional variables which make successful imaging more difficult. The need to isolate many thin, well aligned wet samples is a difficult process and exposure to high intensity X-rays for a long period of time denatures the underlying structure. The data acquired at DLS suggests that the sectioning techniques used to isolate the fibre bundles have improved dramatically as the speckle pattern observed was confined to a single line. This means the individual fibrils in the bundle were well aligned.

The algorithm used for reconstruction has been shown to cope well with a mixture of real, phase only and complex images. External components such as a pupil function, position function and diffraction patterns can be easily imported and manipulated. Memory requirements rather than the coding for the algorithm was shown to be a limiting factor which could be overcome by running it on a cluster.

The data acquired at DLS was shown to be good enough for some initial reconstructions where a periodicity in the phase ~ 65nm and in the amplitude ~300nm could be recovered. The increased periodicity in the amplitude reconstruction requires further investigation, perhaps on a different synchrotron source.

Initial results point towards time being the only obstacle rather than technological ability. Third generation synchrotron sources provide the pre-requisites of high flux and a high degree of coherence required if ptychography is to be successful. Better modelling of the pupil function at the sample and additional tomography data acquired by moving the sample in 3D rather than 2D could remove any ambiguities in the phase solution.

Appendix A

Reconstruction Algorithm

```
% Ptychoraphy Algorithm (moving object)
% ptycho_script_rjb_modif.m
% see APL, 85(20), pp. 4795-4797 for details and description of algorithm
% numbers in brackets eg. {6} require to step in algorithm from above paper

%% clear workspace, set data location, define parameters

close all              % close all plots / figures
clear                  % clear workspace
%clc                    % clear command screen

data_root = [pwd '/']; % location of data directory % (forward '/' on UNIX
num_expos = [11 1];  % number of exposures [y x]
obj_size = [1300 1340]; % object size [y x]
fft_size = [256 256]; % fourier transform size [y x]

num_iterations = 1000;   % number of iterations
delta = 0.001;         % value of delta (weiner filter - prevents /0 errors in update
function)

tl = 1 + (obj_size - fft_size) / 2;          % position of top left corner of object
region covered by FFT
br = ((obj_size - fft_size) / 2) + fft_size; % position of bottom right corner of
region covered by FFT

obj_guess = ones(obj_size);                  % {1} for first guess at object create
array of ones of same size as object
corrected_obj_guess = ones(obj_size);        % corrected object guess, initially set
to ones
total_sse=zeros(num_expos,1);                % Bin values used throughout algorithm
total_sse1=zeros(num_expos,1);
sse1=zeros(prod(num_expos),num_expos);
sse=zeros(prod(num_expos),num_expos);
fourier_sum=zeros(prod(num_expos),1);
fourier_sum1=zeros(prod(num_expos),1);
load([data_root 'pupil/ASsquarepupil50pxz=1.455.mat'], 'pupil');      % load probe
function from file (check variable name in workspace)
load([data_root 'positions/vertical_pty_scan_50px_pupil.mat'], 'pos');    % load probe
positions from file (check variable name in workspace)

%% results display

fig1 = figure('Position', [0 309 1270 502], ...
    'Name', 'Current Object Guess', 'NumberTitle', 'off');
colormap('default')

axes11 = subplot(2,1,1);                   % subplot 1: amplitude
title('Object Modulus');
set(gca,'Xlim',[0 obj_size(2)]);
set(gca,'Ylim',[0 obj_size(1)]);
set(gca,'YDir','Reverse');
set(gca,'NextPlot','replacechildren')
axis image

axes12 = subplot(2,1,2);                   % subplot 2: phase
title('Object Phase');
set(gca,'Xlim',[0 obj_size(2)]);
set(gca,'Ylim',[0 obj_size(1)]);
set(gca,'YDir','Reverse');
set(gca,'NextPlot','replacechildren')
axis image

%% PIE main loop
for k=1:1300;                               % Initiating random start guess
    for l=1:1340;
        obj_guess(k,l)=0.75+5.97*rand*j;
    end
end
illum_weight = (abs(pupil) ./ max(max(abs(pupil))));                    % weights
update by the relative illumination by probe
update_f = illum_weight .* (conj(pupil) ./ ((abs(pupil) .^ 2) + delta));  % update:
weight x (P*/(P^2 + d))
```

```
fprintf([ num2str(num_iterations) ' iterations']);                      % display
number of iterations

for n = 1:num_iterations                                                % loop over
total iterations
    fprintf(['\niteration ' num2str(n) ' ']);                           % display
current iteration number



    for i = 1:prod(num_expos)                                           % loop over all
exposures (difraction patterns)
        fprintf('.');                                                   % display
progress (a '.' for every exposure)


        current_guess = obj_guess( ...                                  % extract
illluminated part of current object guess
            (tl(1) + pos(i,1)):(br(1) + pos(i,1)), ...                  % Ymin:Ymax for
extract at 'i'th position
            (tl(2) + pos(i,2)):(br(2) + pos(i,2)));                     % Xmin:Xmax for
extract at 'i'th position


        illum_guess1 = current_guess .* pupil;                           % {2} multiply
current object extract with illumination function
        illum_guess=zeros(obj_size);                                     % Inserting
256*256 guess into 1300*1340 for comparison with real data
        illum_guess(523:778,543:798)=illum_guess1;
        diff_guess = fftshift(fft2(fftshift(illum_guess)));             % {3} propgate
to far-field by FFT (diff. pattern)
        diff_guess2=abs(diff_guess).^2;                                 % Used to
calculate the error
        load([data_root 'diffraction_patterns/pty_dp_' num2str(i)], 'pty_dp');% load
diff. pattern for current position
        recorded_dp = pty_dp;                                                % place
actual diff pattern in recorded_dp


        phase_guess = complex(0, 1) .* angle(diff_guess);               % calculate
guessed phase (i*theta(k,R))
        corrected_guess = sqrt(recorded_dp) .* exp(phase_guess);        % {4} replace
guessed amp with modulus from actual..
                                                                       %.. diff
patterns, keeping guessed phases


        corrected_area = ifftshift(ifft2(ifftshift(corrected_guess))); % {5} propogate
to object plane by inv FFT ..
                                                                        % to get
corrected object


        corrected_obj_guess((tl(1) + pos(i,1)):(br(1) + pos(i,1)), ... % {6} correct
whole obj guess at current probe pos.
            (tl(2) + pos(i,2)):(br(2) + pos(i,2))) = ...                % Ymin:Ymax,
Xmin:Xmax of whole object guess
            obj_guess((tl(1) + pos(i,1)):(br(1) + pos(i,1)), ...        % Ymin:Ymax,
Xmin:Xmax of most recent obj guess
            (tl(2) + pos(i,2)):(br(2) + pos(i,2))) + ...
            (update_f .* (corrected_area(523:778,543:798) -
illum_guess(523:778,543:798)));              % update function for current probe
position


        sse1(i,n)=sum(sum(((abs(diff_guess2))-
(abs(recorded_dp))).^2)/((obj_size(1)*obj_size(2)))); % SSE Error
        fourier_sum1(i,1)=sum(sum(recorded_dp).^2);
% Used to normalise SSE to RMS



        obj_guess = corrected_obj_guess;                                % set current
object guess to latest corrected guess
                                                                % {7}
    end


    total_sse1(n,1)=sum(sse1(:,n)); % sum all individual errors ro generate one
overall error
    total_fourier_sum=sum(fourier_sum);   % Sum all diffraction patterns to normalize
    figure(fig2)                % display current object guess
```

```
    subplot(1,3,1)
    imagesc(abs(obj_guess))      % subplot 1: amp
    subplot(1,2,2)
    imagesc(angle(obj_guess))    % subplot 2: phase
    drawnow
end

fprintf('\nPIE reconstruction complete\n');                          % completion
message
```

3D model function

```
%% clear workspace, set data location, define parameters

close all             % close all plots / figures
clear                 % clear workspace
%clc                   % clear command screen

%% Collagen Generator 29-01-2009
% Set up size of array
load('profile2unitcollagen','combinedamp');

vertdistance =690;                                          % Setting y-distance
hordistance = 670;                                           % Setting x-distance
layers = 20;                                                 % Number of layers in
matrix
g=size(combinedamp);
fibrilwidth = g(2);                                          % Width of individual
collagen unit
fibrilheight=g(1);                                           % Height of individual
collagen unit
amp=zeros(vertdistance,hordistance,layers);                  % Set up matrices to be
used later in the function
phase=zeros(vertdistance,hordistance,layers);
p=zeros(hordistance/fibrilwidth,1,layers);
q=zeros(hordistance/fibrilwidth,1,layers);
for l=1:layers;
for j=(1:fibrilwidth:hordistance-(fibrilwidth-1));       % Set up beginning of each
new collagen strip up to 81- each strip is 20
a=zeros(vertdistance,fibrilwidth);                       % Size of collagen fibre
b=(zeros(ceil(30*rand),fibrilwidth));                    % Random start point for
first collagen d-band
c=size(b);
a(1:c(1),1:fibrilwidth)=b;                               % Inserting random start
point into fibre
d=floor(vertdistance-c(1));                              % Calculating how much space
in left in the fibre to fill with d-banding
for k=(1:fibrilheight:d);
f=c(1)+k;
a(f:f+(fibrilheight-1),1:fibrilwidth)=combinedamp;       % Inserting d-banding bloacks
to fill the rest of the fibre
a=a(1:vertdistance,1:fibrilwidth);
end
amp(1:vertdistance,j:j+(fibrilwidth-1),l)=a;             % Selecting fibre so length
is preserved
[m n]=find(amp(1:vertdistance,j:j+(fibrilwidth-1),l));   % Finding non zero elements
within the array and outputting row-column
p(((j+(fibrilwidth-1))/fibrilwidth),1,l)=m(1,1);         % Inserting first non zero
element in each fibre to a table
q((j+(fibrilwidth-1))/fibrilwidth,1,l)=...
    (p((j+(fibrilwidth-1))/fibrilwidth,1,l)-p(1,1,1));   % Calculating the difference
between first non-zero elements in each fibre
phase(1:vertdistance,j:j+(fibrilwidth-1),l)=...
    ones(vertdistance,fibrilwidth)*q((j+(fibrilwidth-1))/...
    fibrilwidth,1,l)*(pi/g(1));
end
amp1(:,:,l)=amp(81:680,1:600,l);
phase1(:,:,l)=phase(81:680,1:600,l);
end
clear a b c d f g h j k l m n p q amp phase combinedamp fibrilheight fibrilwidth
hordistance phase vertdistance

data=amp1.*exp(i*phase1);                                    % Create complex wave
function
clear phase1 amp1
%%
num_expos = [3 5];  % number of exposures [y x]
obj_size = [600 600]; % object size [y x]
fft_size = [256 256]; % fourier transform size [y x]
```

```
tl = 1 + (obj_size - fft_size) / 2;          % position of top left corner of object
region covered by FFT
br = ((obj_size - fft_size) / 2) + fft_size; % position of bottom right corner of
region covered by FFT


load('pupil/ASsquarepupil50pxz=1.455.mat', 'pupil');      % load probe function from
file (check variable name in workspace)
load('positions/squarepupil_60_overlap_pupil_size_50.mat', 'pos');    % load probe
positions from file (check variable name in workspace)


%%
total_sum_current_guess=zeros(3200,3200);
for r = 1:prod(num_expos)
for s=1:layers;


    current_guess1(:,:,s)=
data((tl(1)+pos(r,1)):(br(2)+pos(r,1)),(tl(2)+pos(r,2)):(br(2)+pos(r,2)),s);  % get
the current object times the pupil function for the specific
    total_current_guess(:,:,s)=current_guess1(:,:,s).*pupil; % Applying kinematical
approach
sum_current_guess=cumsum(total_current_guess,3);
total_sum_current_guess(1473:1728,1473:1728)=sum_current_guess(:,:,layers);


dp = abs(fftshift(fft2(fftshift(total_sum_current_guess)))) .^ 2;            % get the
diffraction pattern of the object
end


    save(['diffraction_patterns/dp_' num2str(r)], 'dp');      % save the data
end
```

References

Aspden, R. M., & Hukins, D. (1979). Determination of the Direction of Preferred Orientation and the Orientation Distribution Function of Collagen Fibrils in COnnective Tissues from High-Angle X-ray Diffraction Patterns. *Journal of Applied Crystallography , 12*, 306.

Aspden, R. M., & Hukins, D. (1979). Determination of the Direction of Preferred Orientation and the Orientation Distribution Function of Collagen Fibrils in COnnective Tissues from High-Angle X-ray Diffraction Patterns. *Journal of Applied Crystallography , 12*, 306.

Astbury, W. (1940). *International Society of Leather Trades' Chemists , 24*, 69.

Bear, R. S. (1942). Long X-ray Diffraction Spacings of Collagen. *Journal of the American Chemical Society , 64* (3), 727.

Berenguer, F. (2008). Private Communication.

Bonse, U., & Hart, M. (1965). An X-ray Interferometer with Long Separated Interfering Beam Paths. *Applied Physical Letters , 6*, 155.

Brauer, S. e. (1995). X-ray Intensity Fluctuation Spectroscopy Observations of Critical Dynamics in Fe3Al. *Physical Review Letters , 74* (11).

Brodsky, B., & Eikenberry, E. F. (1980). Characterisation of Fibrous Forms of Collagen. *Methods in Enzymology , 82*, 127.

Bunk, O. e. (2008). Influence of the Overlap Parameter on the Convergence of the Ptychographical Iterative Engine. *Ultramicroscopy , 108* (5), 481.

Burge, R. E., & Randall, J. T. (1955). The Equivalence of Electron-Microscope and X-ray Observations on Collagen Fibres. *Proceedings of The Royal Society A , 233*, 1.

Chapman, H. N., Jacobsen, C., & Williams, S. A. (1996). A Characterisation of Dark Field Imaging of Colloidal Gold Labels in Scanning Transmission X-ray Microscope. *Ultramicroscopy , 62*, 191.

Chapman, J. A. (1974). The Staining Pattern of Collagen Fibrils I. An Analysis of Electron Micrographs. *Connective Tissue Research , 2* (2), 137.

Cowley, J. M. (1995). *Diffraction Physics* (3 ed.). North Holland.

Cowley, J. M. (1992). *Electron Diffraction Techniques: Volume 1* (Vol. 1). Oxford University Press.

Faulkner, H., & Rodenburg, J. (2004). Movable Aperture Lensless Transmission Microscopy: A Novel Phase Retrieval Algorithm. *Physical Review Letters , 93* (2), 023903.

Fienup, J. (1982). Phase retrieval algorithms: a comparison. *Applied Optics , 21* (15), 2758.

Fienup, J. (1987). Reconstruction of a Complex-valued Object from the modulus of its Fourier Transform Using a Support Constraint. *Journal of the Optical Society of America A. , 4* (1), 118.

Fletcher, S. (2008, Aug 25). *Beamline I-22 Design Specifications*. Retrieved Mar 23, 2009, from Diamond Website: http://www.diamond.ac.uk/Beamlines/Beamlineplan/I22/TechSpecs.htm

Fraser, R., MacRae, T. P., Miller, A., & Suzuki, E. (1983). Molecular Conformation and Packing in Collagen Fibrils. *Journal of Microbiology , 167* (2), 497.

Gerchberg, R., & Saxton, W. (1972). A Practical Algorithm for the Determination of Phase from Image and Diffraction Plane Pictures. *Optik , 35*, 237.

Goodman, J. W. (2005). *Introduction to Fourier Optics - 3rd Edition.* Greenwood Village, Colorado: Roberts & Company Publishers.

Hergel, R., & Hoppe, W. (1970). *Ber. Bunsen-Ges , 74*, 1148.

Herzog, R. O., & Jancke, W. (1920). *Chemische Berichte , 53* (10), 2162.

Hodge, A. J., & Petruska, J. A. (1963). Recent Studies with the Electron Microscope on Ordered Aggregates of the Tropocollagen Molecule. *Aspects of Protein Structure* , 289.

Hodge, A. J., & Schmitt, F. O. (1963). The Charge Profile of the Tropocollagen Macromolecule and the Packing Arrangement in Native-Type Collagen Fibrils. *Proceeding of The National Society of Science, Washington , 46* (2), 186.

Hoppe, W. (1969). *Acta Cryst. A. , 25*, 495.

Hulmes, D. (1992). The Collagen Superfamily - Diverse Structures and Assemblies. *Essays in Biochemistry , 27*, 49.

Jacobson, C., & Kirz, J. (1998, August). X-ray Microscopy with Synchrotron Radiation. *Nature Structural Biology - Synchrotron Supplement* , 650.

Kennedy, C. J., & Wess, T. J. (2003). The Structure of Collagen within Parchment - A Review. *Restaurator , 24* (2), 61.

Kratky, O., & Serkora, A. (1943). *Journal of Makromolecular Chemistry , 1*, 113.

Libbert, J. e. (1997). Speckle in coherent x-ray reflectivity from Si(111) wafers. *Physical Review B , 56* (11), 6454.

McCallion, C. (2008). Private Communication.

McCallion, C. (2008). Unpublished Report - Investigation into the preparation and characterisation of collagen samples.

Meek, K. M., & Quantock, A. J. (2001). The Use of X-ray scattering Techniques to Determine Corneal Ultrastructure. *Progress in Retinal and Eye Research , 20* (1), 95.

Miao, J., Charalambous, P., Kirz, J., & Sayre, D. (1999). Extending the Methdology of X-ray Crystallography to Allow Imaging of Micrometre-sized Non-Crystalline Specimens. *Nature , 400*, 342.

Miao, J., Sayre, D., & Chapman, H. (1998). Phase Retrieval From the Magnitude of the Fourier Transforms of Nonperiodic Objects. *Journal of the Optical Society of America A , 15* (6), 1662.

Miller, A., & Wray, J. S. (1971). Molecular Packing in Collagen. *Nature , 230*, 437.

Morrison, G. R., & Browne, M. T. (1992). Dark Field Imaging with the Scanning Transmission X-ray Microscope. *Review of Scientific Instruments , 63*, 611.

North, A., Cowan, P. M., & Randall, J. T. (1954). Structural Units in Collagen Fibrils. *Nature , 174*, 1142.

Orgel, J. P., Wess, T. J., & Miller, A. (2000). The in situ Conformation and Axial Location of the Intermolecular Corss-Linked Non-Helical Teleopeptides of Type I Collagen. *Structure , 8* (2), 137.

Pfeiffer, F. e. (2008). Hard X-ray Dark-Field Imaging Using a Grating Interferometer. *Nature Materials , 7*, 134.

Read, R. J. (2005, June 7). *Fourier Transforms: Structure Factors, Phases and Electron Density*. Retrieved March 5, 2009, from Structural Medicine - Protein Crystallography Course - University of Cambridge: http://www-structmed.cimr.cam.ac.uk/Course/Fourier/Fourier.html

Robinson, I. e. (2008). SM549 - Project Log Book.

Rodenburg, J. (1992). The Theory of Super-Resolution Electron Microscopy via Wigner-Distribution Deconvolution. *Philosophical Transcripts of the Royal Society of London A , 339*, 521.

Sayre, D. (1952). Some Implications of a Theorem due to Shannon. *Acta Crystallography , 5*, 843.

Schmitt, F. O., Hall, C. E., & Jakus, M. A. (1942). Electron Microscope Investgations of the Structure of Collagen. *Journal of Cellular and Comparative Physiology , 20* (1), 11.

Shannon, C. (1949). Communication in the Presence of Noise. *Proceedings of the Institute of Radio Engineers , 37* (1), 10.

Stout, G., & Jensen, L. (1989). *X-ray Structure Determination.* New York: Wiley.

Stuhrmann, H. B., & Miller, A. (1978). Small-Angle Scattering of Biological Structures. *Journal of Applied Crystallography , 11*, 325.

Sutton, M. e. (1991). Observation of Speckle by Diffraction with Coherent X-rays. *Nature , 352*, 608.