Scanning Microscopy Vol. 11, 1997 (Pages 109-129) 0891-7035/97$5.00+.25
Scanning Microscopy International, Chicago (AMF O'Hare), IL 60666 USA

AN EVALUATION OF NOISY IMAGES OF HELICAL STRUCTURES AND THE DETECTION OF HIGH RESOLUTION DATA

David Gene Morgan* and David J. DeRosier

Rosenstiel Basic Medical Sciences Research Center, Department of Biology,
Brandeis University, Waltham, MA

Abstract

Helical structures are more disordered and diffract more weakly than two-dimensional crystals, and images of them are inherently noisier. Automated techniques allowed us to correct, align and merge data from hundreds of images. To extend from ~1.0 nm resolution to the near atomic level (~0.4 nm), we needed to know whether our handling of images could be improved. For example, curved particles after correction may or may not produce data as good as particles that need no correction.

We also wanted to evaluate the accuracy with which we predict the locations of layer lines hidden by noise and determine if all images contribute equally to higher resolution data. For this purpose, we developed an algorithm (the sniffer) which evaluates regions of Fourier transforms that contain signal buried in noise.

We examined both the image handling procedures and the sniffer using images of bacterial flagellar filaments embedded in glucose and phosphotungstic acid. Although there was a correlation between phase residual and such factors as tilt, curvature and disorder, these defects explained only ~5% of the variance. This suggested that helical particles following correction contribute almost as much as particles requiring no correction. In addition, positions of high resolution layer lines appeared accurately predicted from lower resolution layer lines. We also found that the signal-to-noise ratio in low resolution layer lines only weakly correlated with that of high resolution layer lines.

Key Words: Electron cryomicroscopy, image processing, helical structure, three-dimensional reconstruction.

*Address for correspondence:
David Gene Morgan
Department of Biophysics, Boston Univ. Sch. Medicine
Boston, MA 02118, and Department of Cell Biology, Harvard Medical School, Boston, MA 02115
Telephone number: (617) 638-4088
FAX number: (617) 638-4041
E-mail: dgm@med-phobos.bu.edu

Introduction

Although a helical object was the first biological structure investigated using electron microscopy and three-dimensional (3-D) reconstruction (DeRosier and Klug, 1968), no one has yet attained resolutions comparable to those attained for two-dimensional crystalline arrays (Henderson *et al*., 1990). One reason is that high resolution data from helices is generally much weaker than that from comparably sized crystalline specimens (Morgan and DeRosier, 1992). Moreover, a single filament has fewer subunits, and hence, less signal than a two-dimensional crystal. Thirdly, helical structures are generally more disordered than crystalline objects: helical structures tend to flex in solution, and subunit positions within a filament may vary from the average value. Unless such factors are corrected, the amplitude of the high resolution data is reduced. By averaging together many corrected images, it is possible to obtain these high resolution data.

To this end, we developed methods (Morgan and DeRosier, 1992) that allow for the analysis of hundreds of images of helical objects. The algorithms use correlation methods to automate many steps of the analysis and correction, and thus require little or no human intervention. With large numbers of images for analysis, we were able to collect data to ~1.0 nm resolution (Morgan *et al*., 1995). Using various approaches, others also obtained 1.0 nm data for helical structures (Jeng *et al*., 1989; Unwin, 1993; Mimori *et al*., 1995).

To see if we could improve our averaged data set, we wanted to know whether some images are better than others:

(1) Even though we correct the images of filaments for curvature (Egelman, 1986) and tilt out of the plane normal to the direction of the electron beam (DeRosier and Moore, 1970), are the data obtained from the corrected images as good as those from images that require no correction?

(2) Do layer line positions vary from particle to particle, that is, is there really variable twist or are the variations simply due to noise in the images?

(3) Do images with good low resolution information tend to have good high resolution information?

Since we were particularly interested in high reso-

lution Fourier coefficients, we needed to be able to detect them when they were hidden in the noise. Consequently, we developed an algorithm we call the sniffer (Morgan and DeRosier, 1993). We describe herein the method and its application to images of bacterial flagellar filaments embedded in a mixture of glucose and phosphotungstic acid (glc/PTA). We are exploring this method of embedment since it preserves the filament structure to near-atomic resolution (Ruiz *et al*., 1994).

Materials and Methods

Microbiology and Biochemistry

Flagellar filaments were isolated from *Salmonella typhimurium* strain SJW1660 (obtained from Dr R. Kamiya, National Institute for Basic Biology, Okasaki, Japan). This strain contains a single point mutation in the flagellin gene (*fli*C) which causes a Gly to Ala mutation at residue 426 in the flagellin protein (Hyman and Trachtenberg, 1991; Kanto *et al*., 1991). This single amino acid substitution causes the normally corkscrew-shaped filaments to be straight, and therefore, more amenable to image analysis than wild type filaments.

Bacterial cultures were grown to mid log phase, and the cells were isolated by centrifugation (14000 g for 20 minutes at 4°C) (Ruiz *et al*., 1993; Morgan *et al*., 1995). The cells were resuspended in several volumes of ice-cold phosphate buffer (pH 7.0). Filaments were sheared from the cells by vortex mixing using two slightly different procedures. The sample referred to below as glc/PTA I was initially resuspended in 20 mM sodium phosphate (pH 7.0) and was vortex mixed for 60 seconds (s). The other sample (glc/PTA II) was resuspended using 20 mM potassium phosphate buffer (pH 7.0), and the filaments were sheared by vortex mixing for 30 s, diluting approximately two-fold with ice-cold potassium phosphate buffer and vortex mixing an additional 30 s. This latter procedure increases the efficiency of shearing.

Intact cells and any filaments remaining attached to them were removed by centrifugation (14000 g for 20 to 30 minutes at 4°C), and the isolated filaments were concentrated by centrifugation (145000 g for 120 minutes at 4°C) into a volume of approximately 0.1 mL, as described previously (Ruiz *et al*., 1993; Morgan *et al*., 1995). Isolated filaments were stored on ice at 4°C and remain usable for many weeks.

Specimen preparation and electron microscopy

The initial electron micrographs used in this study (referred to below as glc/PTA I) were obtained while one of us (D.J.D.) was on sabbatical leave at the Laboratory for Molecular Biology, MRC, Cambridge, U.K. The remainder of the work described here was performed at Brandeis University, Waltham, MA.

(1) glc/PTA I: The filament preparation was diluted using distilled water until filaments were well spaced on carbon coated electron microscope grids which had been glow discharged for 2 minutes in a methylamine atmosphere. Typically, the dilution was 1:10. A 2 μL drop of filament suspension was placed onto the grid, blotted after 1-2 minutes, then embedded and lightly stained using a 2 μL drop of 2% glucose (pH 7.0) followed by a 2 μL drop of 2% phosphotungstic acid. The combined drops were blotted after 1-2 minutes. Grids were loaded into a cryoholder designed by R. Henderson (Henderson *et al*., 1991) and cooled to -150°C inside a Philips (Philips Electronic Instruments, Mahwah, NJ) 420 electron microscope. Images at a defocus of approximately 400 nm underfocus were recorded onto Kodak (Eastman Kodak Co., Rochester, NY) SO-163 film at a nominal magnification of 55,000x using 100 kV electrons from a standard tungsten electron source. The total electron dose was limited to less than 1000 electrons/nm^2, and films were developed for 10 minutes using full strength Kodak D 19 developer.

(2) glc/PTA II: The filament preparation was diluted 200-fold using MilliQ filtered water (Millipore Corp., Bedford, MA) and immediately vortex mixed. Freshly diluted filaments (5 μL aliquots) were vigorously mixed with 15 μL of a 2:1 mixture of 2% glucose (pH 7.0) and 2% PTA (pH 7.0). Drops (4-5 μL) of this mixture were applied to freshly prepared continuous carbon foils and directionally blotted using Whatman (Whatman Inc., Clifton, NJ) #40 filter paper after 4 minutes. Grids were immediately loaded into either a Gatan (Gatan Inc., Pleasanton, CA) model 626 cryoholder or a prototype Oxford Instruments (Oxford Instruments Inc., Concord, MA) single tilt cooling holder, loaded into a Philips CM12 electron microscope (Philips Electron Optics, Eindhoven, The Netherlands), and cooled to -180 or -174°C, respectively. The Philips microscope was equipped with both the standard Philips anticontaminator and a Gatan model 651N anti-contaminator held at -180°C. Images in several defocus ranges (~200, ~500, ~700 and ~900 nm underfocus) were recorded onto Kodak SO-163 film at a nominal magnification of 60000x using 120 kV electrons from a Denka (Denka Kagaku Kogyo Kaishi, Shibukawa, Japan) M3 LaB_6 electron source. The total electron dose was limited to less than 1000 electrons/nm^2, and films were developed for 12 minutes using full strength Kodak D19 developer at 20°C.

Digitization

Micrographs were screened visually and images containing well separated, relatively straight filaments with no visible defects were included for analysis. We further selected only those with clear Thon rings indicating little or no astigmatism or drift. Some micrographs from the sample

referred to as glc/PTA I were digitized at a stepsize corresponding to 0.35 nm using a flat bed microdensitometer (Arndt *et al*., 1969). The remaining micrographs were digitized at a 0.33 nm stepsize using an Eikonix (Eastman Kodak) #1412 linear CCD (charge-coupled device) densitometer (Owen and DeRosier, 1993). We digitized 240 to 500 nm long lengths of filaments. In addition, large areas from all the micrographs were digitized using the Eikonix device. These areas were used to estimate defocus and astigmatism (Morgan *et al*., 1995).

Image analysis

An initial analysis of the glc/PTA II data and a detailed description of the analysis of vitreous ice embedded flagellar filaments have been presented (Morgan and DeRosier, 1992; Morgan *et al*., 1995). Images from the samples referred to as glc/PTA I and II were analyzed using the procedures described in Morgan *et al*. (1995). The following steps briefly describe the image analysis. The cross-correlation steps require a reference which is in the form of a layer line data set.

(1) Estimation of helix axis. Each filament image (240-500 nm in length) was divided into short segments of filament (16.5 to 24.8 nm in length), and the equatorial data from the transforms of each segment were cross-correlated with the equatorial data from the reference. The peak in this cross-correlation map gave the approximate location of the helix axis for each segment of filament in the image. The helix axis for an entire filament was estimated by smoothing these peak positions using a spline algorithm (Egelman, 1986).

(2) Refinement of helix axis and determination of layer line positions. Each filament image was again divided into segments, but in this case, segments were longer (57.8 to 99 nm). Non-equatorial layer line data from the transform of each segment (selected as depicted in Morgan *et al*., 1995) were cross-correlated with the reference. The peaks in these cross-correlation maps gave a better estimate of the location of the helical axis.

We also used the peak positions to predict the exact layer line positions for the helical diffraction data in each image. This helical cross-correlation approach (described in detail in Morgan and DeRosier, 1992) worked as follows: if the reference and a filament being analyzed had exactly the same helical symmetry, knowing the orientation of one of the 60-100 nm segments from a 240-500 nm long filament image, we were able to predict exactly the position and orientation of neighboring segments. If the symmetry of the test filament differed from the reference, the positions and orientations of neighboring segments deviated from the predictions in a systematic fashion. Such systematic variations were used to determine the exact helical parameters, and hence, layer line positions for each particle. In addition, the non-systematic variations provided a measure of the axial and angular disorder in each filament.

(3) Filament straightening and layer line collection. The filaments were straightened using the refined estimates for the helix axis. A spline fitting algorithm similar to the one described by Egelman (1986) but constrained to pass exactly through the refined positions (Press *et al*., 1988) provided the estimates for the axis at all positions along each filament. The filaments were straightened and tightly reboxed (width ~30 nm), apodized (Stewart *et al*., 1981), floated (DeRosier and Moore, 1970) into a 512 by 4096 array and Fourier transformed. The 85 layer lines, with Bessel orders $|n| \leq 60$ and with a meridional resolution less than 0.68 nm, were extracted. In addition, we extracted 170 lines of noise which together with the 85 real layer lines uniformly sampled reciprocal space.

(4) Alignment and merging of layer line data. Selected sub-sets of layer line data from each filament (selected as described in Morgan *et al*., 1995) were cross-correlated with the reference data set used in steps 1 and 2. The position of the peak in these cross-correlation maps gave the x and z shifts required to align the two data sets (DeRosier and Moore, 1970; Morgan and DeRosier, 1992). Out-of-plane tilt was determined by further minimization of the phase residual as a function of tilt applied to the layer line data from each image (Amos and Klug, 1975). Data sets with a phase residual of $< 90^\circ$ and at least a 10° difference in the phase residual in the up versus the down orientation (Amos and Klug, 1975) were merged. The merged layer line data set was used as reference data in the subsequent round of alignment and merging. After a good average was generated, we repeated the entire process starting with the determination of layer line positions and re-extraction of data. Three to five iterations seemed sufficient to fix the alignment (e.g., tilt) of each image.

(5) Three-dimensional reconstruction Lines of noise and layer lines with no statistically significant signal (Morgan *et al*., l995) were removed from the merged data sets. The steps of near- and far-side data extraction, Fourier Bessel inversion and three-dimensional reconstruction were performed using standard methods and have been described previously (DeRosier and Moore, 1970; Trachtenberg and DeRosier, 1987; Owen and DeRosier, 1993; Morgan *et al*., l995; Owen *et al*., 1996).

During steps 1 through 4, we collected a set of parameters that describe in detail the images themselves and the operations performed on them (see Table 1). We calculated linear regression analyses of pairs of parameters to look for correlations, for example, phase residuals *versus* initial amount of curvature. The coefficient of determination (r^2) from the linear regression measures the amount of variation explained by the correlation while the statistical significance of the correlation is measured using the Student's t-test:

Table 1. Examples of parameters collected during image processing.

microscope factors:
defocus/astigmatism
accelerating voltage
magnification
image qualities:
dimensions
minimum density
maximum density
image processing characteristics:
pixel size
overall curvature
in-plane tilt
out of plane tilt
angular disorder
axial disorder
helical lattice parameters
alignment parameters
orientation
amplitude weighted phase residuals
up/down phase differences
run-time parameters for various programs used during analysis

$$t = \frac{\sqrt{r^2(N-2)}}{(1-r^2)} \qquad (1)$$

where N is the number of pairs of observations.

Contrast transfer function (CTF) correction

To combine data from images taken at different degrees of defocus, we corrected each individual layer line data set using values for defocus and astigmatism (determined as described in Morgan *et al*., l995). For the results presented here, we corrected only the phase information for the effects of the CTF.

Statistical analysis

Statistical evaluation of the aligned and merged layer line data sets has been described previously (Morgan *et al*., 1995). Evaluation of the final three-dimensional reconstructions was performed as described in Trachtenberg and DeRosier (1987) and Morgan *et al*. (1995).

The sniffer algorithm

We have devised a simple method to quantitate the amount of signal present in noisy regimes (Morgan and DeRosier, 1993). Our proposed measure is based on the cosine of the phase differences between corresponding pairs of Fourier coefficients, one phase from a layer line pixel of an individual image and the other from the corresponding position from an average of a set of images. The phase difference for any one pair of Fourier coefficients will depend upon the signal-to-noise ratio: for Fourier coefficients with very large signal-to-noise ratios, the expected phase difference will be close to zero (its cosine will approach 1), and for very small signal-to-noise ratios, the average phase difference will approach 90° (its cosine will approach 0). An estimate of signal based on one pair of Fourier coefficients will not be very reliable but averaging cosines from many pairs increases the reliability in a predictable fashion.

Our proposed measure of the signal present in the k^{th} individual image is Sg_k:

$$Sg_k = \langle \cos(\Delta\phi_k) \rangle = (1/N) \sum_{j=1}^{N} \cos(\varphi_{j,k} - \langle \varphi_k \rangle) \qquad (2)$$

where $\langle \cos(\Delta\varphi_k) \rangle$ is the average value of the cosine of the phase difference $\Delta\varphi$, $\Delta\varphi_k = \varphi_{j,k} - \langle \varphi_k \rangle$, N is the number of Fourier terms included in the evaluation, $\varphi_{j,k}$ is the phase of the j^{th} Fourier term of the k^{th} image, and $\langle \varphi_j \rangle$ is the corresponding phase derived from the average of all the images. We tried calculating an amplitude-weighted average of $\cos(\Delta\phi)$ but found it made little improvement. We included only terms corresponding to regions where there was statistically significant signal in the average. Alternative selection or weighting schemes may improve the sensitivity of Sg.

Noise in the image will contribute noise to the Fourier coefficients obtained from the image; in particular, if the noise or variance in the densities in the images is σ^2, then the variance in the Fourier coefficients derived from the images will also be σ^2. The variance will be equally partitioned into the real and imaginary components of the Fourier coefficients and should have a normal distribution according to the central limit theorem. We can therefore calculate the expected effects of noise on Sg (see Fig. 1):

$$Sg = \langle \cos(\Delta\varphi) \rangle = \frac{2}{\pi\sigma^2} \int\int \cos\{\tan^{-1}[I,(S+R)]\} \exp\frac{-2R^2}{\sigma^2} \exp\frac{-2I^2}{\sigma^2} dR dI \qquad (3)$$

where S is the amplitude of the signal which is defined for simplicity to have only a real component, σ^2 is the variance of the noise in either the image or its transform, and R and I are the real and imaginary parts of the Fourier transform of the noise. R and I should be normally distributed, i.e., $P(R) \propto \exp(-2R^2/\sigma^2)$ and $P(I) \propto \exp(-2I^2/\sigma^2)$.

We now wish to characterize the accuracy with which we can estimate Sg_k by calculating its variance, var(Sg), and

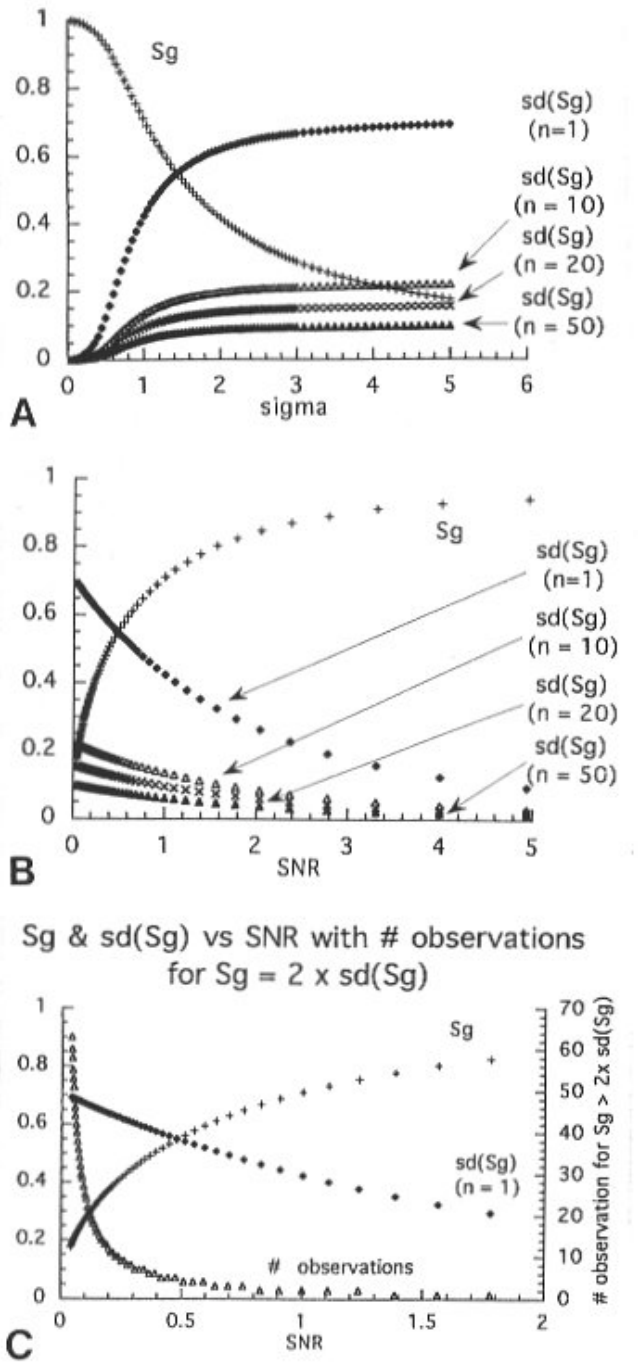

Figure 1. Expected results from sniffer algorithm. In A and B, Sg is shown using plus signs. sd(Sg) is shown using filled diamonds for one observation, open triangles for 10 observations, X's for 20 observations, and closed triangles for 50 observations; (**A**) Sg and sd(Sg) are plotted as functions of σ; (**B**) Sg and sd(Sg) are plotted as functions of the more familiar signal-to-noise ratio (SNR). Figure 1a shows the behavior of Sg and sd(Sg) as ó increases (as the noise increases) while Figure 1b shows the behavior of Sg and sd(Sg) as the SNR increases; (**C**) Sg (plus signs), sd(Sg) (filled diamonds) and the number of observations necessary for Sg to be larger than 2 x sd(Sg) (open triangles) are plotted as functions of SNR. The SNR covers the range 0 to 2.0 in order to show the behavior of these quantities around SNR = 1.0. In these plots, the signal is fixed at a value of 1.00.

standard deviation, sd(Sg):

$$\mathrm{var}(Sg) = (1/n)(\langle \cos^2(\Delta\varphi)\rangle \ \ \langle \cos(\Delta\varphi)\rangle^2) \qquad (4)$$

$$sd(Sg) = \sqrt{var(Sg)} \qquad (5)$$

where N is the number of observations, and $\langle \cos^2\Delta\varphi)\rangle$ is the expectation value of the square of the cosine of the phase difference:

$$< \cos^2(\Delta\varphi) > = \qquad (6)$$
$$\frac{2}{\pi\sigma^2}\int\int \cos\{\tan^{-1}[I,(S+R)]\}\exp\frac{-2R^2}{\sigma^2}\exp\frac{-2I^2}{\sigma^2}dRdI$$

We evaluated Equations (2) to (6) using numerical methods. The results, assuming one, ten, twenty and fifty observations are plotted as a function of σ in Figure 1A. Figure 1B shows these same data replotted as a function of the more familiar quantity signal-to-noise ratio (SNR). We define SNR as $\langle S^2\rangle/\sigma^2$, the ratio of the average power of the signal S to the average power of the noise σ. Since we have arbitrarily set the signal to 1, SNR = $1/\sigma^2$.

The error in Sg needs to be sufficiently low for a single value of Sg to be considered significantly different from zero. If we pick Sg > (2 x sd(Sg)) as significant, we can use Equation (4) to calculate the number of observations necessary for Sg to be significant for any given value of σ or SNR. Figure 1c is a plot of the number of observations necessary for Sg to be statistically significant as a function of SNR.

A realistic situation is one in which signal varies while the expectation value of the noise is relatively constant. In such cases:

$$\langle Sg\rangle = \int P(S)\langle Sg(S)\rangle dS \qquad (7)$$

where P(S) is the distribution of the signal S and $\langle Sg(S)\rangle$ is the expectation value of Sg for a given S (assuming σ constant). The variance of $\langle Sg\rangle$ is

$$\mathrm{var}(\langle Sg\rangle) = \int P(S)\langle \mathrm{var}(Sg(S))\rangle dS \qquad (8)$$

To illustrate the effect of varying signal, we elected to let the signal $S = J_{10}(X)$, as might occur for some layer line of a simple helical structure where J_{10} is a Bessel function of

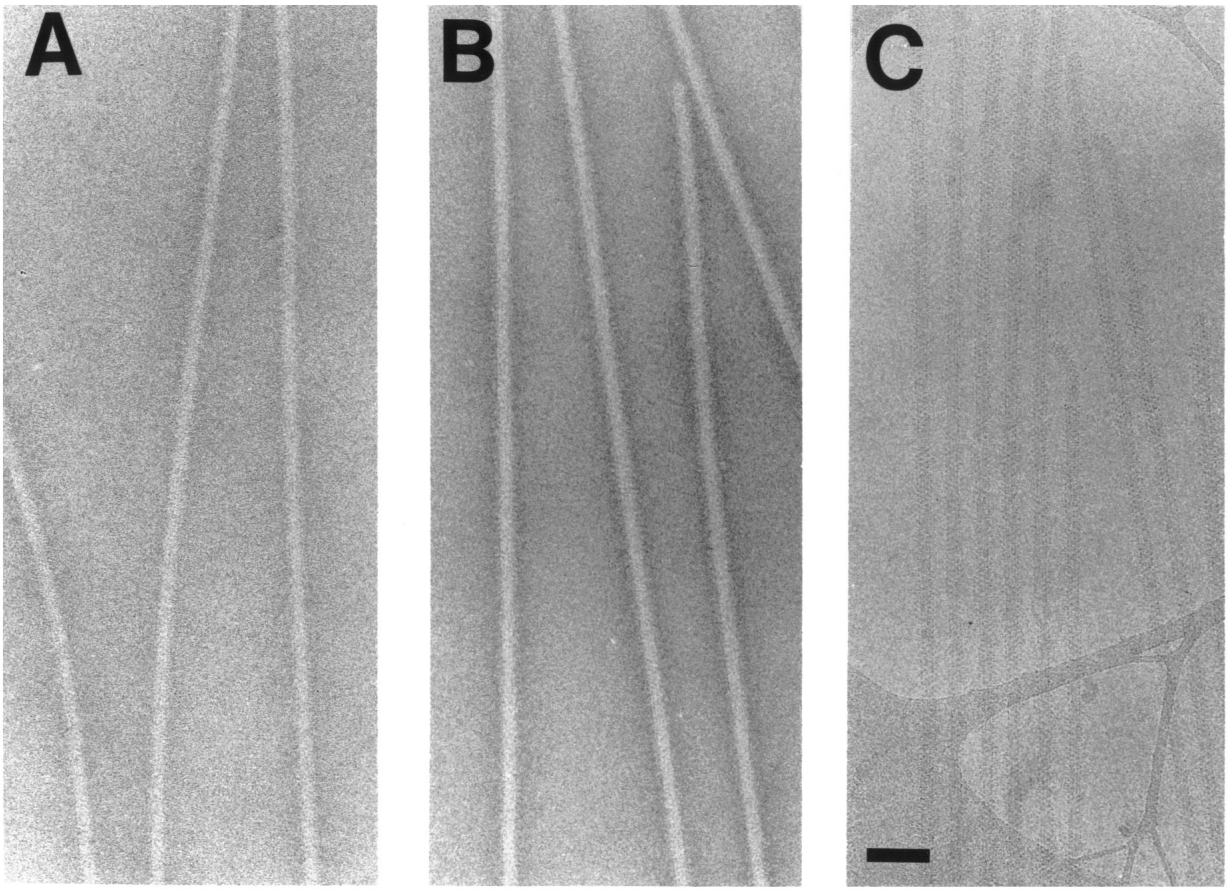

Figure 2. Images of bacterial flagellar filaments. Images from (**A**) glc/PTA I (400 nm underfocus), (**B**) glc/ PTA II (200 nm underfocus) and (**C**) vitreous ice embedded (~500 nm underfocus) samples. Bar = 40.0 nm (A, B and C are at the same magnification).

order 10. We evaluated Equations (7) and (8) over the interval $6 < X < 15$, which might be a typical range over which significant values are found in the average of a set of noisy images. We compared the values obtained to those which would apply if the signal were constant and equal to <S>, the average value of S, obtained from J_{10} over the interval $6 < x < 15$. For this interval, $\langle Sg(S)\rangle$ is 0.481 and var($\langle Sg(S)\rangle$) is 0.0295. If this interval is sampled at 19 evenly spaced points (every 0.5 units), (Sg) is 0.476, $Sg(\langle S\rangle)$ is 0.547 and var($Sg(\langle S\rangle)$) is 0.0290. Since these values are fairly similar, we believe that we can obtain a reasonable estimate of signal even in regions of varying noise.

Computational requirements and times

The work described here was done using a variety of Digital Equipment Corporation (DEC; Maynard, MA) VAX computers running either VMS (3000 series VAXStations) or Open VMS (a DEC3000 Alpha). The Alpha runs our programs ~10x faster than the non-Alpha workstations. Estimation and refinement of the helix axis and layer line extraction takes < 5 minutes for each filament image on the Alpha. Alignment and averaging of the largest data set takes ~2 hours on the Alpha. The sniffer takes ~1 hour to process all the data from a set of 150 filaments.

Display of images

All the images were processed using Adobe Photoshop version 3.0 (Adobe Systems Inc., Mountain View, CA). Graphs were made either with Kaleidagraph version 3.0.2 (Synergy Software, Reading, PA) or Deltagraph version 2.0.3 (Delta Point, Inc., Monterey, CA).

Results

General characterization of images

Micrographs (Fig. 2) of flagellar filaments reveal long, often curved structures ~25 nm in diameter. The morphological features of filaments (clearest in Fig. 2C, but

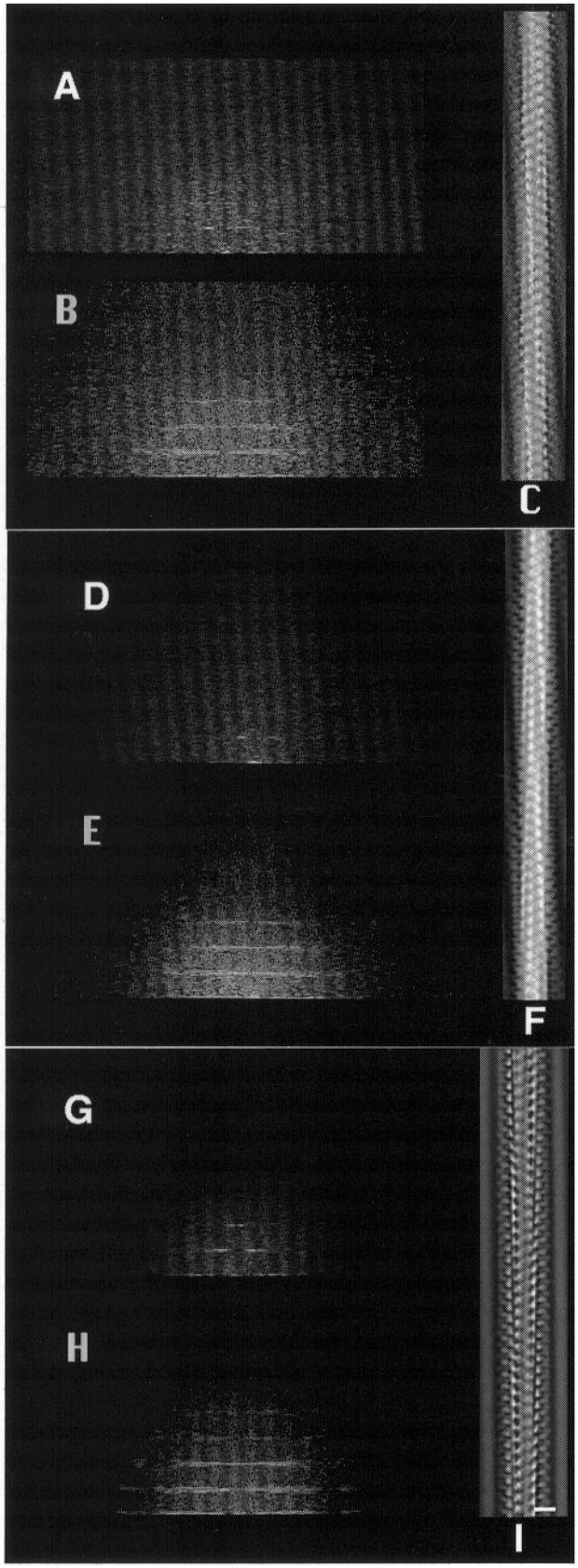

Figure 3. Amplitude data from Fourier transforms and average images. Panels **A**, **D** and **G** show the amplitude portion of the layer line data collected from images similar to those shown in Figure 2. Panels **B**, **E** and **H** show the amplitudes from the averaged layer line data of the three sets of images. Panels **C**, **F** and **I** are images generated by back Fourier transforming the layer line data from panels B, E and H. The scale bar in I represents 10.0 nm. The data used to generate panels E and I were phase corrected for effects of the CTF before aligning and merging. For the amplitude data (panels A, B, D, E, G and H), we eliminated the equatorial data. In panels B, E and H, we display $(\ln(1.0 + \text{amp}))^2$ in order to show the weak, high resolution data. We additionally adjusted the visibility to better reveal the high resolution information in all the amplitude data (panels A, B, D, E, G and H).

also present, albeit more weakly, in Figs. 2A and 2B) form rows characteristic of the 5-, 6- and 11-start helical lattice of the filament. Amplitudes from computed Fourier transforms (Figs. 3A, 3D, and 3G) show the layer lines corresponding to orders n = -11, 6, -5 and 1. These four layer lines provide data to a resolution of ~2.5 nm. There is no indication of higher resolution data in either the meridional or equatorial direction.

We aligned and merged data into three averages: the glc/PTA I average (Fig. 3C), which has data from the fewest images and is the noisiest of the averages; the glc/PTA II average (Fig. 3F) and the vitreous ice embedded average (Fig. 3I). This last average shows most clearly the striations due to the subunit organization although such striations are present in all the average images. The averaged sets of Fourier coefficients (Figs. 3B, 3E and 3H) corresponding to the images show striking differences when compared to the transforms of single images; namely, more layer lines are visible in the low resolution region of the transforms, and this low resolution information is stronger and extends further from the meridian. More importantly for the current study, higher resolution layer lines are visible. These higher resolution data extend to a meridional resolution of 1.0 nm and are strongest in the glc/PTA II sample.

Alignment and merging of layer line data

Using the described correlation based methods, we processed 64 images from the glc/PTA I sample, 336 images from the glc/PTA II sample and 223 images from the vitreous ice embedded sample (Table 2).

The sixty-four images from the glc/PTA I set were taken between 400 and 500 nm underfocus. We did not CTF correct any of these data but restricted the layer line data to a resolution of 1.4 nm (before the first node in the CTF).

Table 2. Parameters from alignment and merging of data sets.

	number of images (included) (total) (# subunits)		phase residual[a]	polarity[a]	Z_{-11}[b] (std. dev.)	Z_{-5}[b] (std. dev.)	axial[c] disorder	angular[d] disorder
glc/PTA I								
	62 (45k)	64	68	22	0.01088	0.19255	0.013	0.10
glc/PTA II								
-200 nm	129* (99k)	149	61	20	0.01107 (0.00089)	0.19541 (0.00172)	0.018 (0.020)	0.13 (0.10)
-500 nm	65 (50k)	68	52	27	0.01081 (0.00084)	0.19561 (0.00158)	0.013 (0.011)	0.11 (0.07)
-700 nm	25 (21k)	25	52	28	0.01068 (0.00090)	0.19568 (0.00101)	0.021 (0.020)	0.14 (0.09)
-900 nm	91 (71k)	92	43	26	0.01089 (0.00048)	0.19366 (0.00115)	0.013 (0.010)	0.10 (0.07)
after CTF correction	309 (240k)	336	56	20				
vitreous ice								
	169 (126k)	223	56	29	0.01109 (0.00181)	0. 19682 (0.00024)	0.037 (0.037)	0.15 (0.14)

[a]in degrees; [b]in nm^{-1}; [c]in nm/subunit; [d]in degrees/subunit; std. dev. = standard deviation.
*Most of the images which could not be aligned and merged were from a single micrograph. Upon closer inspection of this micrograph, it was determined that its CTF exhibited unusual features and these data should not have been included in this analysis.

For the purpose of checking the quality of the images taken at different degrees of defocus, images from the glc/PTA II set were initially processed according to defocus groups: ~200 nm underfocus, ~500 nm underfocus, ~700 nm underfocus and ~900 nm underfocus. The layer line positions and amounts of axial and angular disorder in these sets of images are very similar (Table 2). As expected, layer line data from the further-from-focus images, which have stronger low resolution Fourier coefficients, aligned and merged with lower phase residuals and greater differences in polarity than that from images recorded closer to focus. Satisfied with the images, we then produced a proper CTF weighted data set. To do so, we corrected for CTF induced phase reversals in each of the 336 layer line data sets, aligned and merged the data, and trimmed the average layer line data to a resolution of 0.7 nm. Layer line data from the closest-to-focus images were also separately aligned and merged, and the average trimmed to 0.7 nm.

Images of vitreous ice embedded filaments at defocuses ranging from ~400 nm to ~2500 nm underfocus were processed as described previously (Morgan *et al*., 1995). After correction for phase reversals due to the CTF, these images were merged, and the average data trimmed to a resolution of 0.7 nm.

Despite differences in embedding media, microscopy and digitization, the three data sets are all similar with respect to layer line position and strength and angular and axial disorder. Although we were able to identify only the layer lines corresponding to Bessel orders -11, 6, -5 and 1 in transforms of single images, we collected Fourier coefficients along 85 real layer lines. This was done in order to collect all possible high resolution data for our averaging procedures. After averaging, 34 of the 85 layer lines had statistically significant signal (using the q = 2.0 statistical

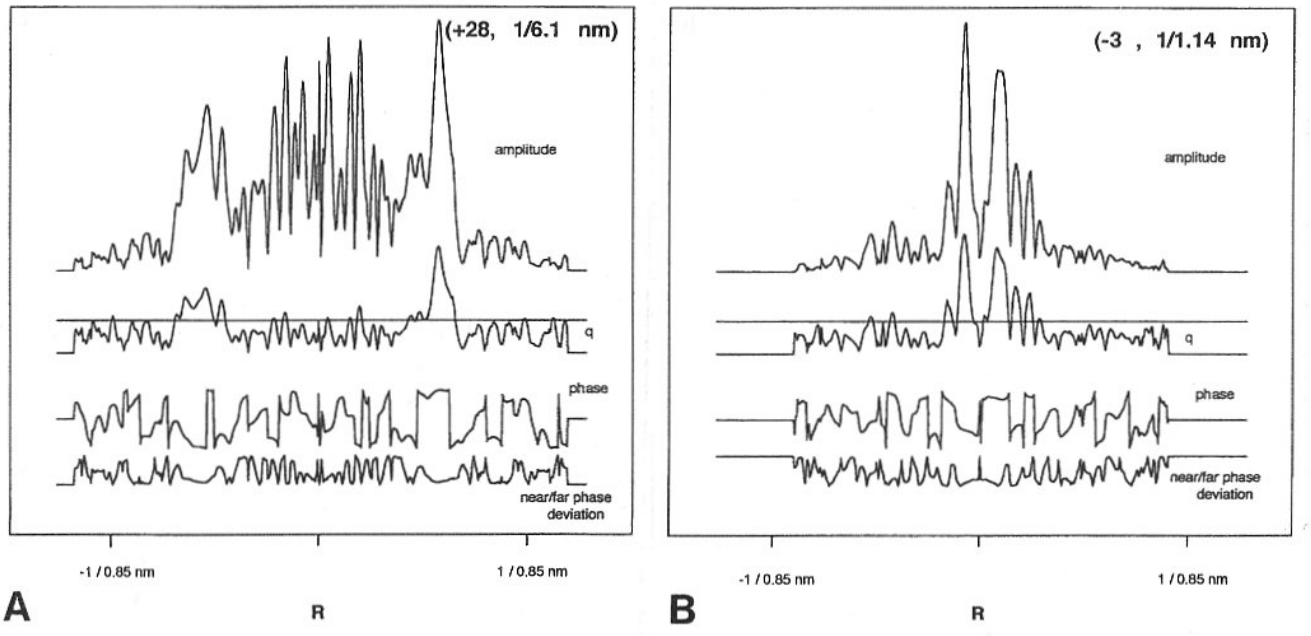

Figure 4. Selected layer lines after CTF correction, alignment and merging. Panels **A** and **B** show the n = 28 and n = -3 layer lines, respectively, for data from the glc/PTA II embedded filaments. Each panel contains (from top to bottom) the amplitude, the q statistic (including a line at the q = 2.0 or 95.7% confidence level), the phase and the near-side/far-side phase difference deviation from 0° or 180° *versus* distance, R, radius from the meridian. The layer line order and meridional resolution are listed in the upper right of each panel. Phase data range from -180° to 180° and the near side/far side phase difference ranges from 0° to 180°.

threshold described in Morgan *et al*., 1995) in at least two of the three data sets. In all three data sets reduced to 34 layer lines, the near side versus far side phase symmetry is good (i.e., the phases for the corresponding sides differ by less than ~30°). The glc/PTA II data set appeared best: A set of three layer lines (Bessel orders 8, -3 and -14) at a meridional resolution of ~1.0 nm have statistically significant, symmetrically shaped amplitude peaks with good near side/far side phase symmetry. In addition, there are a number of layer lines (e.g., Bessel orders 28, -26 and -20) with lower meridional resolution but with statistically significant, symmetrically shaped amplitude peaks and with good near side/far side phase symmetry. Figures 4A and 4B show these data for the Bessel order 28 and -3 layer lines, respectively.

We could align the average data sets from the glc/PTA embedded preparations to one another. However, when attempts were made to align and merge individual sets of layer line data from the two preparations, multiple cycles of alignment and merging resulted in rejection of the glc/PTA I data, and the resulting averaged data were dominated by the glc/PTA II data (see **Discussion**).

Variation in layer line positions

In the collection of data from different images, we used slightly different estimates for the layer line positions. Each image was used to obtain an estimate of the positions of the layer lines for the transform of that image. Is the apparent variation in layer line position real, or is it simply a consequence of noise in the images? From the closest-to-focus images, we collected the same set of 85 real layer lines, assuming constant positions for the layer lines. The constant positions were of course the average positions determined for the set of images. If the layer line positions really vary, the average generated using constant positions will be significantly worse than the average generated assuming variable positions, whereas if the positions are really constant, the reverse will be true. Figure 5 compares the average assuming a variable position with that assuming a constant position for a pair of layer lines: the strong n = +l layer line and the weak, high resolution n = -3 layer line. Clearly, the data obtained assuming variable position are better, especially at high resolution. Thus, the positions of the layer lines do vary, and the apparent variation in the twist of the filaments is real.

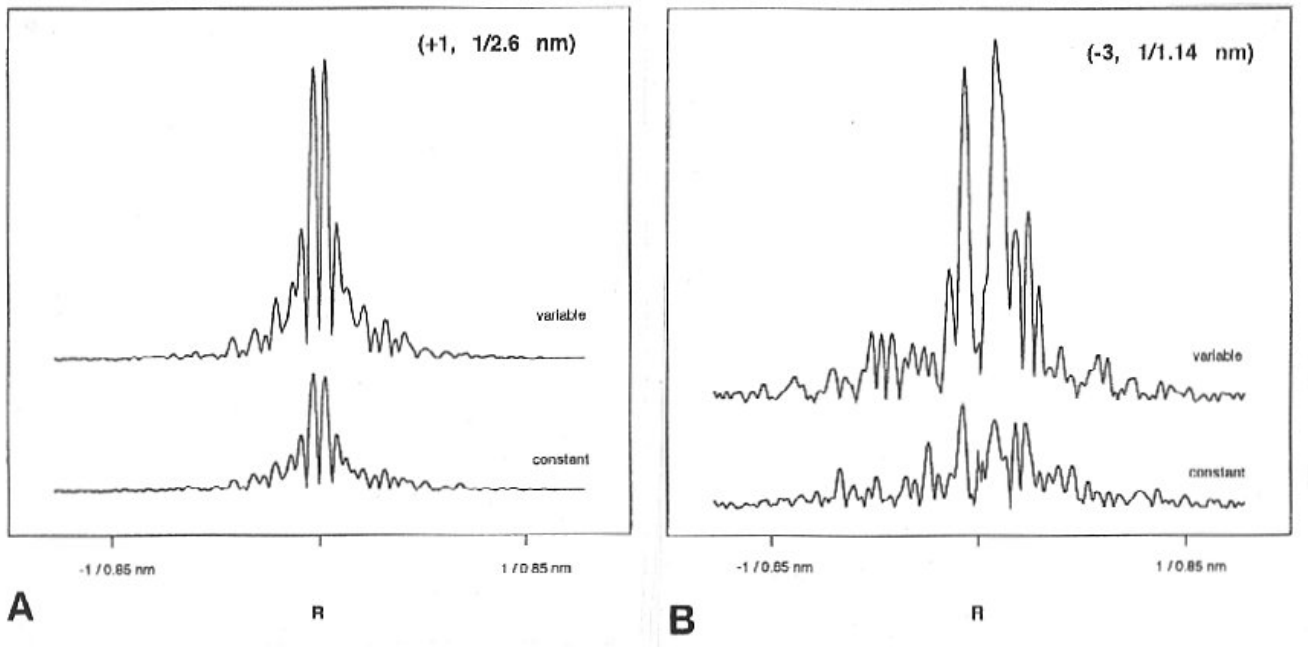

Figure 5. Comparison of averaged layer lines assuming constant or variable layer line positions. The results for a strong, low resolution layer line (n = +1, panel **A**) and for a weak, high resolution layer line (n = -3, panel **B**) are shown. The upper trace in each panel represents the average data using the measured position for the layer line in each image, while the lower trace represents the results using the average of that position from all the data sets. For both layer lines, assuming a variable position produces significantly better results. In panel A, although the difference in amplitudes is ~2x, the overall data are quite similar. In panel B, not only is the amplitude difference greater (~3-4x), but the signal near the meridian is much stronger, and the noise away from the meridian is much weaker. The layer line order and meridional resolution are listed in the upper right of each panel. Amplitude scales are identical for each trace within a panel.

Characterization of reconstructions

Figure 6 shows surface representations obtained from three averaged layer line data sets. The directions of the 11-, 6- and 5-start helical lattices are marked on panel c. All three reconstructions show a knobby structure whose outermost (D3) domains [using the nomenclature of Mimori *et al.* (1995) and Morgan *et al.* (1995)] slew from left to right. The reconstructions from glc/PTA embedded samples (Figs. 6A and 6B) show an extension (marked with arrows) between the base of D3 and the tip of D3 from a subunit along the 5-start helical lattice. This extension is not present in the reconstruction from the vitreous ice embedded preparation.

Thick and thin transverse slices through the reconstructions in Figure 6 are shown in Figure 7. Although the surface representations of the three preparations are similar, these transverse sections appear different. Common features are most easily seen in 5.2 nm thick slices (Figs. 7A, 7B and 7C). The central channel (diameter ~2.5 nm) is surrounded by a ring of density (r_i or D0) at a radius of ~2.0 nm. Surrounding r_i, there is an annular feature (r_o) in Figure 7F at a radius of ~4.5 nm. In the reconstructions from the glc/PTA embedded filaments, this feature is made up of peaks of density referred to as D1 in previous work (Trachtenberg and DeRosier, 1987; Namba *et al.*, 1989; DeRosier, 1992; Ruiz *et al.*, 1993). There are peaks of density (D2) at a radius of ~6.5 nm in all three reconstructions. The outermost feature (D3) occurs at a radius of ~9.5 nm and is much weaker in the glc/PTA II sample (Figs. 7B and 7E) than in the other two. The relative strengths of these features differ in the three different reconstructions, and it is the variations in strength which make these transverse sections appear so different.

We determined the significant features in these reconstructions using the significance map procedure of Trachtenberg and DeRosier (1987). Figure 8 shows the regions of the transverse thin sections (Figs. 7D, 7E and 7F) which are statistically significant at a 99% confidence level. At this confidence level, all the domains referred to above are significant in all three reconstructions, including the weak D3 region of the glc/PTA II structure (arrow, Fig. 8B). Small, statistically signifi-cant differences (Milligan and Flicker, 1987; Trachtenberg and DeRosier, 1987) between near side and far side reconstructions are associated primarily with the D3 domains (data not shown).

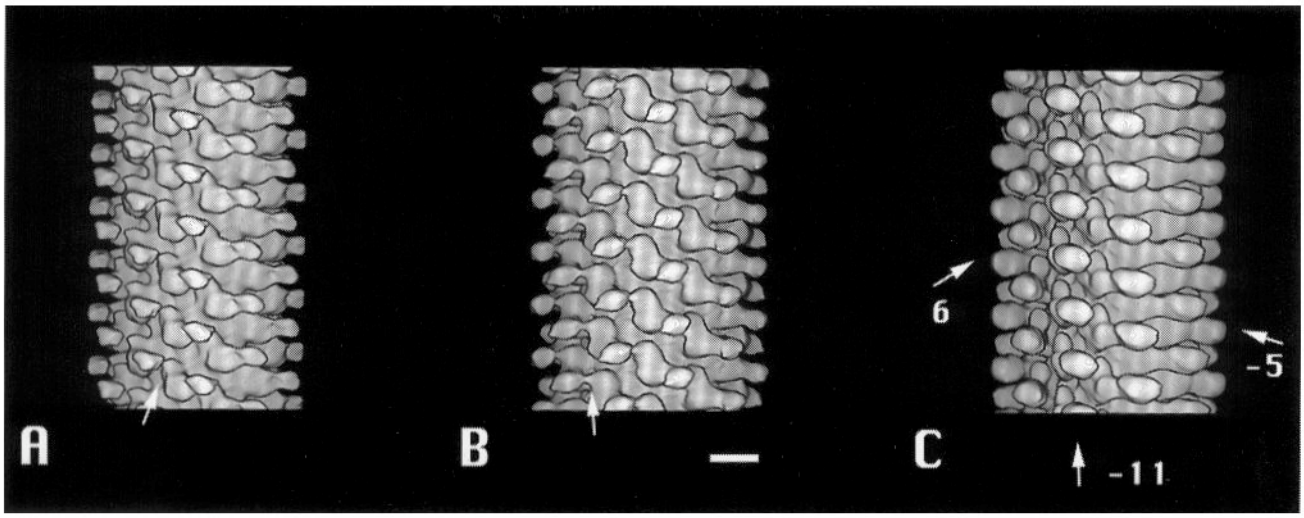

Figure 6. Surface representation of the three-dimensional reconstructions. Contour levels enclosing similar volumes were chosen for the glc/PTA I (**A**), glc/PTA II (**B**) and the vitreous ice embedded (**C**) samples. The cell proximal end of the flagellar filament is at the top, and the directions of the 11-, 6- and 5-start helical lattices are marked with arrows in panel C. Arrows in panels A and B mark an extension between the base of D3 and the tip of D3 from a subunit along the 5-start helix, which is not seen in the vitreous ice embedded reconstruction. The reconstructions shown in this and subsequent panels correspond to the far side data. Bar = 5.0 nm (for A, B and C).

Signal detection

The purpose of our proposed sniffer algorithm is to detect a signal in regions where it is masked by noise. By locating and measuring the signal, we tried to refine the positions of layer lines and to determine if some images in the data set were better than others. We report here on the use of this algorithm on the glc/PTA II data taken at ~200 nm underfocus. We chose this set because the layer line data from the glc/PTA II images are generally better, and the higher resolution signal should be strongest in the closest-to-focus data. In addition, no CTF correction was necessary for the interpretation of these data.

To estimate the SNR in the image, we used the positions at which the averaged amplitudes were significantly different from zero. We did this for the four strongest layer lines at the 99.9% level, and then at the 95% level, and for two high resolution layer lines, at the 95% level (Table 3). The value of Sg for the low resolution layer lines was 0.7, dropping to about 0.4 for the regions within the 95% confidence level. This corresponds to a change in the signal to noise ratio (SNR) from 1 to 0.25 (Fig. 1C). We found Sg of about 0.25 on two high resolution layer lines. This corresponds to a SNR of less than 0.1.

In Figure 9A, we show how Sg varies when the position of the layer line is varied about the position previously determined using helical cross-correlation methods. We show values of Sg for three layer lines obtained from a single image. We see that the peaks lie close to the measured positions of the layer lines. We can get a better estimate of layer line positions by increasing the number of points used to determine Sg. Thus, rather than trying to determine the position of each layer line individually, we can determine them jointly. Recall that in the Fourier transform of an image, the positions of all potential layer lines are linear combinations of the two helical lattice constants, a and b (Owen *et al*., 1996). To combine data from all or from any subset of layer lines, we varied a and b systematically about their calculated values, determined the corresponding positions of the layer lines and combined the layer line data from these positions to calculate Sg_k. The correct values of the constants a and b should maximize Sg. We tried this procedure varying just the high resolution layer lines, which should more accurately fix the layer line positions, and also just the low resolution layer lines. Figures 9B and 9C show typical plots for the low and high resolution data, respectively. The best values for the constants a and b derived from the high or low resolution data were not very different from each other or from those determined in the original processing of the data. Values of a and b typically varied by less than a reciprocal pixel. Changes in the predicted positions of layer lines increase due to changes in a and b with increasing layer line height. Given the measured changes in a and b, the positions of the highest resolution layer lines changed by less than 2 pixels, which is within the

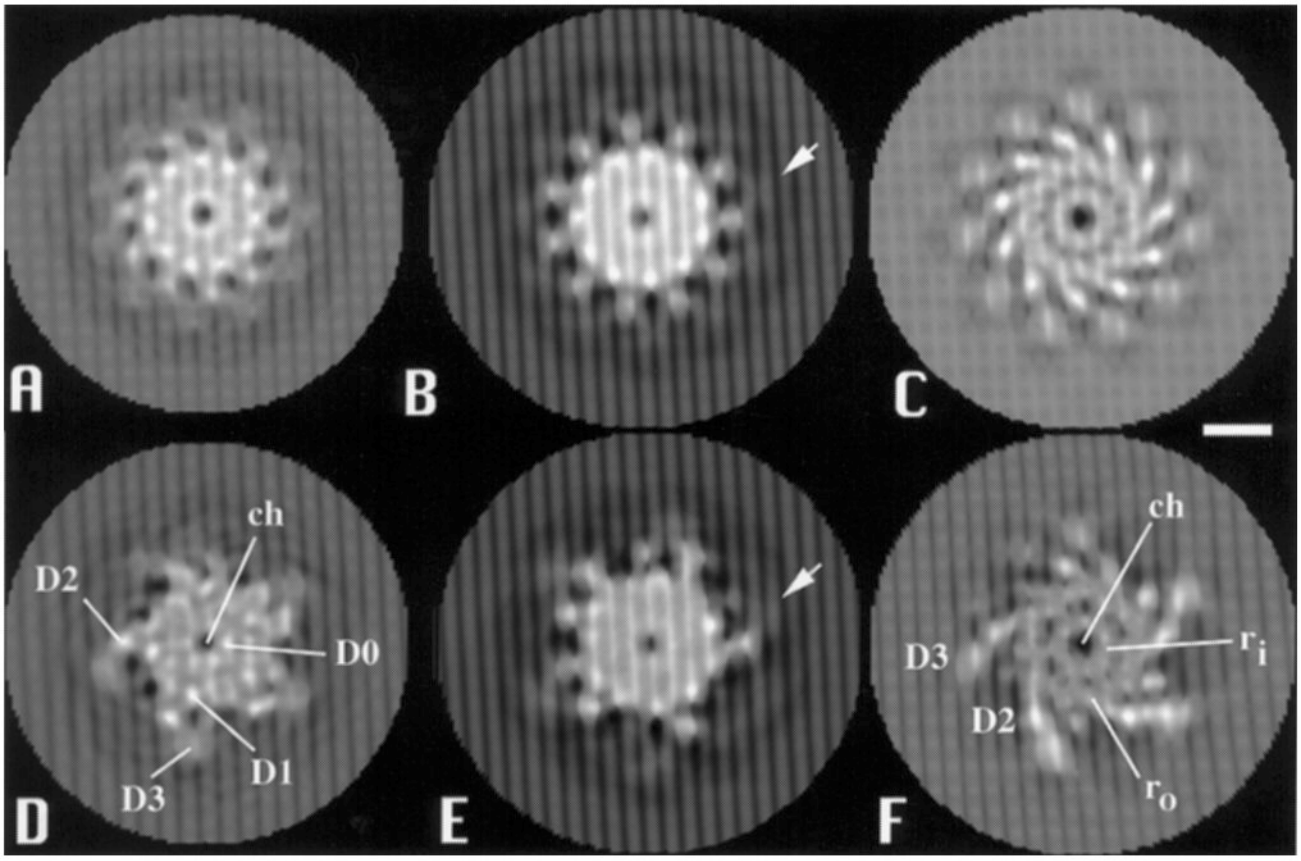

Figure 7. Thick and thin transverse sections taken from the three-dimensional reconstructions. The thick sections in panels **A**, **B** and **C** are 5.2 nm thick, showing an entire asymmetric unit along each 11-start helix. Panels **A** and **D** correspond to the glc/PTA I data shown in Figure 3B after trimming the resolution to 1.4 nm. Panels **B** and **E** correspond to the glc/PTA II data shown in Figure 3E and panels **C** and **F** correspond to the vitreous ice embedded data shown in Figure 3H. The resolution in these last two data sets was trimmed to 0.7 nm before each reconstruction was calculated. Domains D0, D1, D2 and D3 and the channel (ch) are marked in panel D, while D2, D3, r_O, r_i and ch are marked in panel F. An arrow marks the weak D3 domain in panels B and E. Bar = 5.0 nm (for 7A-7F).

layer line width. Moreover, the use of the changed positions to collect the layer line data for a subset of images did not result in an improvement of the phase residuals when sets were compared to the average (data not shown).

Correlation of various factors with "goodness" of individual images

The data collected from the glc/PTA II preparation represent the largest collection of data in this study. This method of sample preparation also appears to preserve the strongest high resolution/high Bessel order information. For these reasons, we used these data to determine the quality of individual images in the set. We used the phase residual for the alignment of individual images as an initial measure of overall image "goodness" and correlated it with parameters we collected during image processing (e.g., tilt out of the image plane, filament curvature, axial and angular disorder and Sg on several layer lines). Results of such comparisons are shown in Table 4. Although most of the correlations are statistically significant ($t > 2$), they account for only a small fraction (~5%) of the variance. We also looked for correlation between axial and angular disorder and between tilt out of the plane of the image and the positions of the two helical lattice constants. These correlations are statistically significant, but they account for only a small amount of the variance.

Discussion

We want to produce maps of filaments to better than 1.0 nm, ideally to about 0.4 nm resolution. As part of this effort, we are exploring the use of glc/PTA embedment. In glc/PTA, structural detail is preserved to 0.4 nm, and since

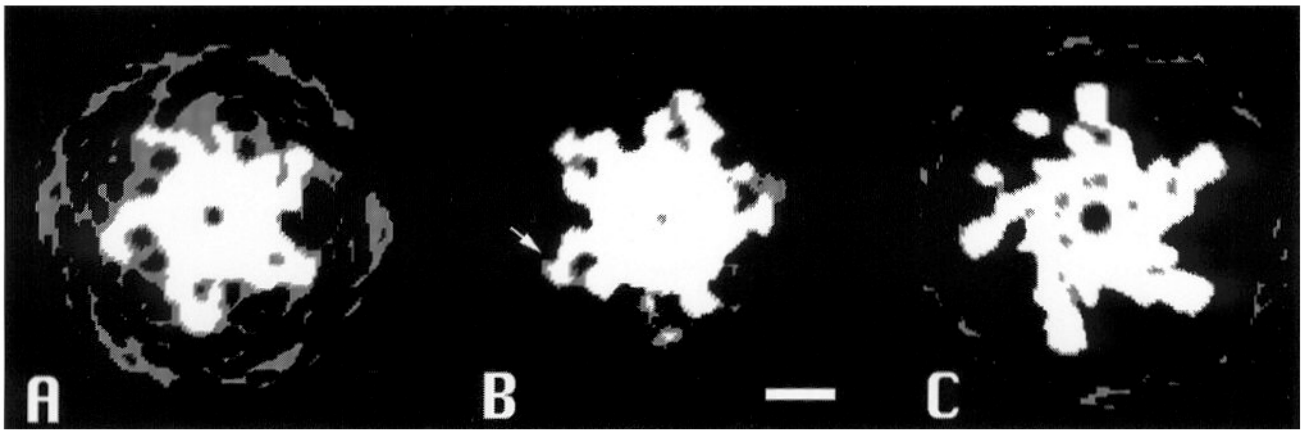

Figure 8. Statistical significance of reconstructions. Features significant at a 99% confidence level are shown for the glc/PTA I (**A**), glc/PTA II (**B**) and vitreous ice embedded (**C**) data. In these reconstructions, all the features marked in Figure 7 are significant, including the weak D3 domains in Figures 7B and 7E (marked with arrow in panel B). Positive significant density (relative to the density corresponding to the edge of the structure) is white, negative significant density is black, and density that is not statistically significant is gray. Bar = 5.0 nm (8A, 8B and 8C).

Table 3. Determination of Sg from closest-to-focus data set.

	Bessel order (# observations)	Sg(all images) average	range	Sg(data which align) average	range
q = 5.0					
	-11 (20)	0.68	(0.92 to 0.13)	0.71	(0.92 to 0.20)
	+6 (15)	0.67	(0.91 to -0.06)	0.70	(0.91 to 0.35)
	-5 (16)	0.67	(0.92 to 0.04)	0.72	(0.92 to 0.45)
	+1 (14)	0.61	(0.93 to -0.17)	0.67	(0.93 to 0.21)
q = 2.0					
	-11 (48)			0.44	(0.69 to 0.17)
	+6 (41)			0.40	(0.74 to 0.14)
	-5 (43)			0.43	(0.70 to -0.02)
	+1 (34)			0.43	(0.68 to 0.16)
	+2 (31) (at ~l/1.29 nm^{-1})			0.29	(0.66 to -0.07)
	-3 (19) (at ~l/1.03 nm^{-1})			0.24	(0.61 to -0.18)

stain embedment provides better contrast at low resolution than ice embedment, we can work closer to focus where the transfer of high resolution detail into the images is better. We began with a 1.0 nm resolution study of flagellar filaments embedded in glc/PTA.

Attempts to collect a large set of images

Since the glc/PTA I images appeared suitable for higher resolution structural studies, we set out to repeat the work, and therefore, to increase the number and defocus range of images of glc/PTA embedded flagellar filaments. The glc/PTA II images are the result of this effort. Although

Table 4. Correlations among image processing parameters*

parameter 1	parameter 2	coefficient of correlation	coefficient of determination	t value
phase residual	up/down difference	-0.82	0.673	25.3
	Sg (n = -5)	-0.42	0.176	5.2*
	Sg (n = +1)	-0.43	0.187	5.5*
	Σ Sg (n =-11, +6, -5, +1)	0.65	0.422	9.7*
	out of plane tilt	0.06	0.004	1.0
	in plane tilt	0.23	0.052	4.1
	curvature (l/r)	0.15	0.024	1.8
	angular disorder	0.22	0.048	4.0
	axial disorder	0.18	0.031	3.2
	Sg (n = +3)	0.18	0.031	2.0
	Sg (n = -8)	-0.30	0.092	3.6
out of plane tilt	position of lattice constant a	0.18	0.035	3.3
	position of lattice constant b	0.25	0.060	4.6
axial disorder	angular disorder	0.34	0.114	6.3
Sg (n = -5)	Sg (n = +1)	0.17	0.029	2.0
Sg (n = +1)	Sg (n = -3)	0.14	0.022	1.7
Sg (n = -5)	Sg (n = +8)	0.05	0.003	0.6
Sg (n = -3)	Sg (n = +8)	-0.02	0.004	0.2
ΣSg (n = -11, +6, -5, +1)	ΣSg (n = +8, -3, -14)	0.18	0.003	2.1

*Determined from the 200 nm underfocus data only.

we could align averages produced from the two sets of images, attempts to align and merge the individual images from both sets did not succeed: when data from individual images were iteratively aligned and merged, the procedures tended to eliminate most of the less numerous glc/PTA I images. We examined several possible causes for this failure.

One cause could be that the two sets of images were digitized using different devices and (slightly) different pixel sizes. We tested this possibility by processing approximately 30 additional glc/PTA I images exactly as described for the glc/PTA II images. This reconstruction (data not shown) appears virtually identical to that from the original glc/PTA I images (Figs. 7A and 7D), and has essentially identical statistically significant regions (Fig. 8A). We were able to align and merge these data with the layer line data from the original processing, but were still unable to align and merge the new data with that from the glc/PTA II images. Thus, processing of the micrographs did not seem to be the cause of the failure of the data sets to merge.

It seemed likely that the cause lay in the sample preparation prior to microscopy. Conditions for the glc/PTA II samples involved longer exposure to negative stain, which could disorder various features of the filaments. Since the glc/PTA II images generate good high resolution data, we do not believe this to be a significant factor. While the outside surfaces of the two reconstructions look the same, the features within the two have noticeable differences. What is particularly striking is the lack of internal detail within the region defined by the ring of D1 domains in the glc/PTA II map. It is as if there is less stain penetration in that preparation. In light of this observation, it is interesting that the amount of glucose in the glc/PTA II samples is greater than that in the glc/PTA I sample. Thus, the ratio of glucose and PTA may be important.

Comparison to the reconstructions from ice-embedded preparations

The surface representations (Fig. 6) of the three reconstructions are similar. Thick and thin sections through the reconstructions (Fig. 7) reveal that the outermost (D3) domains of the filament are weaker in the reconstructions

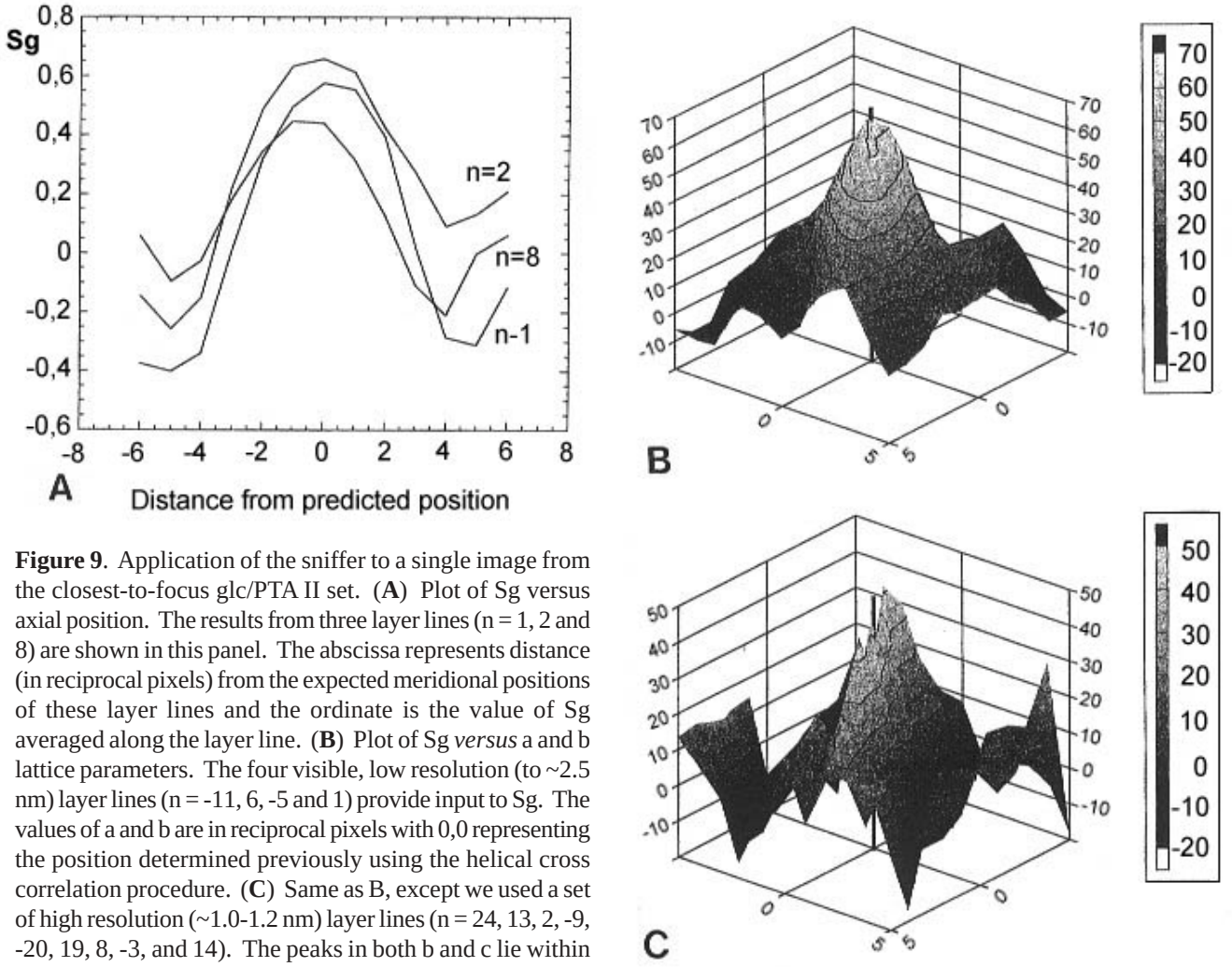

Figure 9. Application of the sniffer to a single image from the closest-to-focus glc/PTA II set. (**A**) Plot of Sg versus axial position. The results from three layer lines (n = 1, 2 and 8) are shown in this panel. The abscissa represents distance (in reciprocal pixels) from the expected meridional positions of these layer lines and the ordinate is the value of Sg averaged along the layer line. (**B**) Plot of Sg *versus* a and b lattice parameters. The four visible, low resolution (to ~2.5 nm) layer lines (n = -11, 6, -5 and 1) provide input to Sg. The values of a and b are in reciprocal pixels with 0,0 representing the position determined previously using the helical cross correlation procedure. (**C**) Same as B, except we used a set of high resolution (~1.0-1.2 nm) layer lines (n = 24, 13, 2, -9, -20, 19, 8, -3, and 14). The peaks in both b and c lie within one reciprocal pixel of 0,0.

based on glc/PTA embedded filaments than in the reconstruction from the vitreous ice embedded filaments. In general, the map derived from the ice-embedded filaments reveals better detail. While all three reconstructions show the same subdomain organization, and both the glc/PTA II and the vitreous ice embedded reconstructions have data to beyond 1.0 nm resolution, only the latter reconstruction reveals the 1.1 nm features within r_o (D1) and r_i (D0), which we believe correspond to axially-oriented α-helices. The most likely explanation for the failure of the glc/PTA embedded samples to show such features has to do with the reversal of con- trast caused by the stain. In negative stain, domains appear as low density holes and scatter generally out of phase with respect to the scattering from internal features of the domains. In ice, both the domains and their internal features scatter in phase. Thus, the stain shell surrounding the protein, which produces strong low resolution scattering at ~0.8 nm resolution, will weaken the 1.0 nm resolution scattering due to the protein itself, in particular the α-helical rods. The portion of the transform that arises from the α-helical rods is located near the equator at ~1.0 nm resolution. This region is strong in Figure 3H (ice data set) relative to that in Figure 3E (glc/PTA II data set).

We conclude from this low resolution work that with stain embedment, we can work close to focus yet preserve sufficient low resolution information by which to straighten and align images of the filament. We record data at 1.0 nm resolution even though it does not reveal the expected peaks corresponding to á-helices. This fine detail should, however, reappear at a resolution beyond 0.8 nm (i.e., the usual cut-off used by X-ray crystallographers to avoid reflections whose amplitudes are affected by solvent). Thus, the increase in the scattering due to stain should not

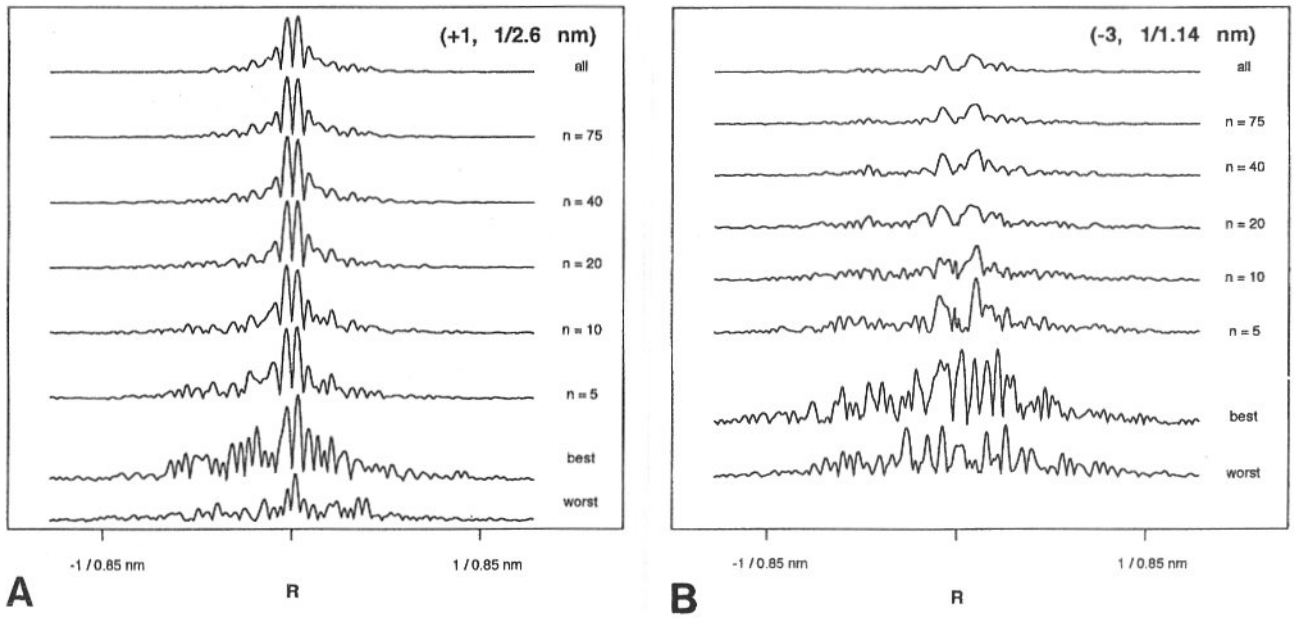

Figure 10. Results of averaging data. The results for averaging data from 5, 10, 20, 40, 75 and 129 (all) images for a strong, low resolution layer line (n = +1, panel (**A**) and for a weak, high resolution layer line (n = -3, panel **B**) are shown. Each panel contains the amplitude *versus* radius from the meridian along both sides of the meridian. At the bottom of each panel, the corresponding best and worst layer line are also shown. The worst in the two cases are from different images as are the best. In panel B, the increase in SNR is most obvious. The layer line order and meridional resolution are listed in the upper right of each panel. Amplitude scales are identical for each trace within a panel.

compromise our ability to collect data in the 0.8 to 0.4 nm region.

How Good are our Procedures for Extracting Data?

With the large data set glc/PTA II, we can ask whether we could improve our signal to noise ratio in the average by handling the set of images differently, for example, by eliminating "poor" images.

Determination of layer line positions

We showed that the layer line positions actually vary from particle to particle. This variation results from the flexibility of the filament much like the variation in particle curvature. To see if we had correctly located the high resolution layer lines, which are not visible in transforms of single images, we used the sniffer algorithm to redetermine their positions. We concluded that the layer line positions (at least to 1.0 nm resolution) were sufficiently accurate to require no further correction. However, Carragher *et al.* (1996) have reported that use of the sniffer produces a significant improvement in finding weak layer lines at ~2.5 nm resolution with images of actin decorated with the myosin S1 fragment.

Validity of corrections for curvature and tilt

If there are systematic errors in any of our image processing procedures, we might identify them by correlating various parameters (e.g., tilt out of the plane) with phase residual. If the correction for out-of-plane tilt is imperfect, there should be a correlation between high tilt angles (requiring large corrections) and high (bad) phase residuals. We might also identify other defects in particles (e.g., high axial and angular disorder), which we could correct in a way similar to the correction of images for curvature.

The correlations we examined (Table 4) are all statistically significant but account for < 5% of the variance between images. Thus, there is no reason to eliminate a subset of images (e.g., those that are tilted) since the loss of signal would far outweigh the loss of noise. It also seems unlikely that a weighted average would be much of an improvement. Thus, we conclude, at least for this data set, that our procedures for correction work well enough.

This kind of analysis can be useful in uncovering problems in the algorithms used for the corrections. We initially noticed a correlation between tilt in the plane of the image and phase residual that accounted for about 10% of

the variance. Since tilt in the plane is simply a function of orientation of the micrograph in the densitometer, it seemed to point to a problem in the spline- fitting algorithm we use to correct for curvature and tilt. Indeed, there was a defect and much of the correlation disappeared when the algorithm was corrected.

Assessment of the goodness of individual images

Is it possible that some of the images in the set are much worse than the others and that the "worst" of these are degrading the average, especially at high resolution? We examined this hypothesis by creating a series of averages which include data from increasing numbers of images. We used the data from the closest-to-focus glc/PTA II images and ordered the images based on increas-ing phase residual. The results of averaging 5, 10, 20, 40, 75 and all (129) data sets is shown in Figure 10 for both a strong, low resolution layer line (n = +1) and for a weak, high resolution layer line (n = -3). In both cases, the noise decreases with the square root of the number of filaments. The results are more dramatic for the n = -3 layer line: the increase from 75 to 129 data sets improves the near side/far side amplitude symmetry, and while it appears that the amplitudes of the maxima decrease somewhat between the average containing 20 data sets and the final average of 129 data sets, the SNR improves due to the large reduction in noise along the layer line. The weak fall-off in amplitude with increasing numbers of filaments suggests that there is real variation in the set of images but that the variation is modest.

We also examined a number of parameters collected during image processing to determine which other parameters are correlated with phase residual. These results are listed in Table 4. Polarity (up/down phase residual differences) shows the expected correlation with phase residual: the higher the phase residual, the smaller the up-down difference. This correlation accounts for almost 70% of the variation in the two measurements. Thus, either of these measures could be used as an indicator of overall image goodness. There are also strong correlations between the phase residual and Sg measured on two strong, low resolution layer lines (n = -5 and n = +1) which account for 15-20% of the variation. The correlation between phase residual and the sum of Sg along the four layer lines visible in all Fourier transforms of these filaments is even stronger, accounting for over 40% of the variation. These correlations are not surprising since in the calculation of an rms (root mean square) amplitude-weighted phase residual, the data along these layer lines accounts for a significant amount of the overall amplitude weighting.

We also investigated whether Sg at low resolution is a good predictor of Sg at high resolution because this might afford us a way to screen particles for inclusion in an average where we seek to include the best high resolution data. Our results (Table 4) indicate that the correlation between Sg measurements for a low resolution and a high resolution layer line is barely statistically significant (t ranges from 0.6 to 1.7) and accounts for only a few percent of the variance (0.3-2.2%) in the measurements. Correlations between pairs of low resolution layer lines or between pairs of high resolution layer lines are also weak (t ranges from 0.2 to 2.0) and account for small amounts (0.4-2.9%) of variance. We also correlated the sum of Sg from the 4 strong, low resolution layer lines (n = -11, +6, -5, +1) with the sum of Sg from three weak but statistically significant, high resolution layer lines (n = +8, -3, -14). Again, the correlation is barely significant (t = 2.1) but accounts for a small amount of the variance (0.3%). Thus, the strength of low resolution data is not a good predictor of the strength of high resolution data.

It is perhaps a surprising conclusion that measurements of signal in one region of layer line data do not better predict signal in other regions. Of course, we eliminated those images which do not align and merge, and thus, we are not making a statement about all the images but rather only those that have passed our selection procedure. From this analysis, we conclude that the selected images used in the average are all similar in quality and that variations among them are uncorrelated.

We are thus in a position to tackle the extension of data to 0.4 nm. The sniffer algorithm appeared to work and provided a check on the accuracy of layer line positions for 1.0 nm resolution layer lines. It may help in locating the 0.4-0.5 nm resolution layer lines. The sniffer, however, can be used for any structure: nothing in the formalism excludes its use for other symmetries.

Acknowledgments

The authors would like to thank Linda Melanson for her invaluable microscopy on the glc/PTA II and vitreous ice embedded samples. DJD expresses his thanks to R. Henderson for time spent in his laboratory. Figures based on analysis of the glc/PTA I work have been presented previously (DeRosier, 1992; Morgan *et al*., 1993; Ruiz *et al*., 1993), although the analysis has not. Ideas based on the sniffer have been presented earlier (Morgan and DeRosier, 1993). We acknowledge the support of the John Simon Guggenheim Foundation for a fellowship to DJD. This work was also supported by grants GM26357, GM35433 and GM49851 to DJD.

References

Amos LA, Klug A (1975) Three-dimensional reconstructions of the contractile tail of T4 bacteriophage. J Mol Biol **99**: 51-73.

Arndt UW, Barrington-Leigh J, Mallett JFW, Twinn KE (1969) A mechanical microdensitometer. J Phys E Sci Instrum **2**: 385-387.

Carragher B, Whittaker M, Milligan RA (1996) Helical processing using PHOELIX. J Struct Biol **116**: 107-112.

DeRosier DJ (1992) Whipping flagellin into shape. Curr Opin Struct Biol **2**: 280-285.

DeRosier DJ, Klug A (1968) Reconstruction of three-dimensional structures from electron micrographs. Nature **217**: 130-134.

DeRosier DJ, Moore PB (1970) Reconstruction of three-dimensional images from electron micrographs of structures with helical symmetry. J Mol Biol **52**: 355-369.

Egelman EH (1986) An algorithm for straightening images of curved filamentous structures. Ultramicroscopy **19**: 367-374.

Henderson R, Baldwin JM, Ceska TA, Zemlin F, Beckmann E, Downing KH (1990) Model for the structure of bacteriorhodopsin based on high resolution electron cryo-microscopy. J Mol Biol **213**: 899-929.

Henderson R, Raeburn C, Vigers G (1991) A side-entry cold holder for cryoelectron microscopy. Ultra-microscopy **35**: 45-53.

Hyman HC, Trachtenberg S (1991) Point mutations that lock *Salmonella typhimurium* flagellar filaments in the straight right-handed and left-handed forms and their relationship to filament superhelicity. J Mol Biol **220**: 7988-*last page*.

Jeng T-W, Crowther RA, Stubbs G, Chiu W (1989) Visualization of alpha helices in tobacco mosaic virus by cryo-electron microscopy. J Mol Biol **205**: 251-257.

Kanto S, Okino H, Aizawa S, Yamaguchi S (1991) Amino acids responsible for flagellar shape are distributed in terminal regions of flagellin. J Mol Biol **219**: 471-480.

Milligan RA, Flicker PF (1987) Structural relationships of actin, myosin, and tropomyosin revealed by cryo-electron microscopy. J Cell Biol **105**: 29-39.

Mimori Y, Yamashita I, Murata K, Fujiyoshi Y, Yonekura K, Toyoshima C, Namba K (1995) The structure of the R-type straight flagellar filament of Salmonella at 9 Å resolution by electron cryomicroscopy. J Mol Biol **249**: 69-87.

Morgan DG, DeRosier D (1992) Processing images of helical structures: A new twist. Ultramicroscopy **46**: 263-285.

Morgan DG, DeRosier D (1993) Detection of high resolution signal in electron micrographs of helices. Biophys J **64**: A243.

Morgan DG, Macnab RM, Francis N, DeRosier DJ (1993) Domain organization of the subunit of the *Salmonella typhimurium* flagellar hook. J Mol Biol **229**: 79-84.

Morgan DG, Owen C, Melanson LA, DeRosier DJ (1995) Structure of bacterial flagellar filaments at 11 Å resolution: Packing of the á-helices. J Mol Biol **249**: 88-110.

Namba K, Yamashita I, Vonderviszt F (1989) Structure of the core and central channel of bacterial flagella. Nature **342**: 648-654.

Owen C, DeRosier DJ (1993) A 13-Å map of the actin-scruin filament from the Limulus acrosomal process. J Cell Biol **123**: 337-344.

Owen CH, Morgan DG, DeRosier DJ (1996) Image analysis of helical objects: The Brandeis helical package. J Struct Biol **116**: 167-175.

Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1988) Numerical Recipes in C. Cambridge University Press, Cambridge, UK. pp 96-97.

Ruiz T, Francis N, Morgan DG, DeRosier DJ (1993) Size of the export channel in the flagellar filament of Salmonella typhimurium. Ultramicroscopy **48**: 417-425.

Ruiz T, Ranck J-L, Diaz-Avalos R, Caspar DLD, DeRosier DJ (1994) Electron diffraction of helical particles. Ultramicroscopy **55**: 383-395.

Stewart M, Kessler MW, Levine RJC (1981) Structure of Limulus telson muscle thick filaments. J Mol Biol **153**: 781-790.

Trachtenberg S, DeRosier DJ (1987) Three-dimensional structure of the frozen-hydrated flagellar filament: The left-handed filament of *Salmonella typhimurium*. J Mol Biol **195**: 581-601.

Unwin N (1993) Nicotinic acetylcholine receptor at 9 Å resolution. J Mol Biol **229**: 1101-1124.

Discussion with Reviewers

B. Carragher: It would be helpful to provide the number of images involved in reaching 1.0 nm resolution for the flagellar filaments and also to estimate how many would be required to reach still higher resolutions.

Authors: While we used more than 100 images in our reconstruction of vitreous ice embedded flagellar filaments, significant high resolution data on either the near- or the far-side are obtained with about 20 images (see Fig. 10B for data based on glucose/PTA embedment). The additional images increase the signal to noise (and thus our confidence) in the final reconstruction. Approximately 20 images for a 1.0 nm resolution structure is in rough agreement with the results both of Mimori *et al*. (1995), who averaged data from both sides of 16 images to attain 0.9 nm resolution for a different form of the bacterial flagellar filament, and of Jeng *et al*. (1989), who averaged data from both sides of 12 images of tobacco mosaic virus (TMV) to attain an ~1.0 nm resolution structure.

We did not attempt to look for 0.45 nm data (where the electron diffraction is strong), but we did not see anything significant between ~1 .0 nm and 0.7 nm resolution (the digitization limit of this study which happens to

correspond to where the electron diffraction is very weak). We may require 1000 or more images to attain ~0.4 nm resolution. For example, the amplitudes of the 1.0 nm data are reduced by an order of magnitude relative to the low resolution (< 5.0 nm) data, and if the fall-off is logarithmic with resolution, then one would roughly need $10^{1.0/0.4} \cong 300$ times more images or ~300 x 20 ≅ 6000. This estimate may be too pessimistic, but until we try, we have no way of accurately determining the number of images.

B. Carragher: Why are CTF corrections made only for phases and not amplitudes (or temperature factors)? Is it the case that groups of images in similar defocus ranges are averaged together prior to CTF correction? **Authors**: We have attempted CTF correction with simple phase reversal and with a more rigorous, multiple-image-based Wiener filter approach [*cf*. Morgan *et al*. (1995) for a description], which attempts to correct phase and amplitude. Amplitudes at the lowest resolutions, especially on the equator, appear to be incorrectly amplified by the Wiener filter in that the density modulations across the interior of our reconstructions are too flat. We explored the use of temperature factors to correct our higher resolution amplitudes, but since we have no knowledge of the fall-off in amplitudes, these corrections are hard to justify.

In all our CTF correction work, we determined the defocus and astigmatism for each micrograph and applied those parameters to the individual layer line data sets from single flagellar filaments. This allows us to account for any azimuthal relationship between the major axis of astigmatism in the micrographs and the filament axis. We did, however, as a way of checking on image quality, try aligning and merging our data in sets where all members of the set were recorded with similar defocus. We did not use those preliminary averages further.

K. Namba: Although it is stated that the scattering due to stain should not compromise the ability to collect data in the 0.8 to 0.4 nm region, the significant differences between the maps shown in Figures 7D and 7E, which were calculated from the two sets of data collected from two different preparations of stained samples (especially the differences in the core region of the filament, where the á-helices are densely packed) seem to suggest that the distributions of penetrated stain can vary from preparation to preparation. This implies possible variations and modifications of the density distributions by stain, which may make the interpretations of the map rather difficult, even if the structure itself is well preserved by the sample-preparation method and can be deduced at higher resolution.

M. Schmid: The difference between the two data sets in glucose appears to be that the glc/PTA I has generally less contrast between the layer lines and the background, both individually and in the average. Is this due to a lower real space contrast arising from different embedding of the filaments in the two experiments? Is then the observation of weaker density at the outside of the filaments in the glc/PTA II reconstruction perhaps due to a higher level of staining at the periphery of the filament?

Authors: The differences in the two maps appear due in part to differences in stain penetration. Note that the glc/PTA II specimen, which shows less stain penetration, was exposed to 0.7% PTA, while the more stain penetrated glc/PTA I specimen was exposed to 1.0% PTA. On the other hand, the glc/PTA II was suspended in the stain solution, placed onto a grid and blotted, while the glc/PTA I specimen was only briefly exposed to stain after being deposited on the grid. In addition to differences in stain penetration, the outer domain (D3) of the glc/PTA II preparation appears weaker. The glucose/stain layer in the glc/PTA II preparation is thinner (compare Figs. 2A and 2B) and perhaps, interactions of the surfaces of the stain layer with the filaments have disordered the outer domains of the structure.

Whether or not the glucose/stain mixture used here alters the subunit conformation is unknown. We do know that the axial á-helices are present in both the glc/PTA and vitreous ice embedded samples, but we do not know that either of these methods preserves the subunits in their native states. Fortunately, there is X-ray fiber diffraction data available, and ultimately, this may permit us to judge how well the structure determined by microscopy accounts for the X-ray diffraction results.

M. Schmid: How could one tell that one needs the "sniffer"? If the data appear to end past a certain resolution, are there always more data there, waiting to be extracted and averaged, or are some structures intrinsically limited in their resolution, and can the sniffer tell the difference?

Authors: The sniffer as we have used it does not find data where none has been detected in the average. Its use therefore, is in assessing the contribution of each filament to the average and in refining parameters for each filament (e.g., locating layer line positions). We found that our filaments contribute roughly equally to the average data and that our estimates of layer line position were good enough for 1.0 nm resolution. On the other hand, Carragher *et al*. (1996) found that an algorithm similar to the sniffer helped them in analysis at ~2.5 nm resolution of F-actin decorated with the S1 fragment of myosin. It seems, therefore, worth using, even if only to judge whether one is doing as well as possible for a given set of images. It could also be used to assess new images which one is thinking of adding to an existing data set. More exotic applications of the sniffer are also possible (e.g., the sniffer should detect anti-correlation in regions of phase reversal and might

therefore be useful for CTF correction, although our initial attempts at this have been less than satisfying).

With regard to the question of data which disappear beyond a certain resolution, it may well be that there are always data there, but that the signal-to-noise ratio drops below a useful level. The sniffer relies on the comparison of a high signal-to-noise data set (an average in the cases described here) to individual data sets: if one had such a high signal-to-noise data set (e.g., an atomic resolution structure based on modeling rather than averaging data), it might be possible to use the sniffer to detect comparable signal from individual images.

M. Schmid: How good does one's initial estimate of helical parameters (of n's, and of the a and b cell axes) have to be for the sniffer (or the helical cross-correlation) to work properly?
Authors: To do any sort of three-dimensional work on helices, one must know the values of n, the order of the layer line. Estimates of a and b must be accurate enough so that data is collected within the axial width of the layer line (which is inversely proportional to the length of the image transformed). One might use the sniffer to refine estimates made at low resolution for higher resolution work.

M. Schmid: Any weighting scheme that could be applied to the described phase correlation methods would likely upweight lower resolution data at the expense of weaker, noisier high resolution data. After the decision to include a piece of data in the first place, it would appear that the most sensitive weighting scheme would be to treat all data with equal weight. Do you agree?
Authors: We experimented with various weighting schemes for both the sniffer and for the alignment and merging algorithm. The sniffer algorithm, which is intended to answer a yes/no question concerning the presence of a signal along given tracks through Fourier space, appears to work best when all data are weighted equally. This makes intuitive sense in that the best possible answer should occur when the maximum number of positions are most phase coherent.

For purposes of alignment of images to a reference, we have tried amplitude based weighting of all data, log (amplitude) based weighting of all data, weighting based only on reference amplitudes, equal weighing of all data and IQ-like estimates of either the signal-to-noise or statistical significance of each Fourier coefficient (i.e., weights based on classifying data into a limited number of "quality ranked" groups). With the closest to focus images from the glc/PTA II data set described here, there is very little difference between averages generated using an alignment based on any of the amplitude-related weighting schemes or a scheme which uses the statistical significance of each Fourier coefficient. When these same data are aligned using only phase coherence (weighing all data equally), the resulting average is similar to the other averages referred to above except that the number of images which align drops by ~60%, and the average is appropriately noisier.

M. Schmid: A signal observable at 1.0 nm resolution is quite remarkable in negative stain, where the resolution limit is normally about 1.5 nm. Is the glucose/ PTA embedding medium described here more properly considered "high contrast glucose"? Do you have an estimate of the mass density of the embedding medium?
Authors: Clearly, the resolution described here is higher than that usually cited. If by "high contrast glucose," you mean something like auro-thio-glucose, which has been shown to preserve atomic detail in certain specimens, it is not yet clear that such data exist in the glucose/PTA embedded flagellar filaments although electron diffraction patterns reveal near meridional layer line data at 0.4-0.5 nm (Ruiz *et al*., 1994).

In X-ray diffraction experiments, variation due to solvent density extends down to about 0.8 nm resolution, and in this sense, negative stains appear to act like high density mother liquor. It is not perhaps surprising, therefore, that comparable resolution is obtained from the disordered "solvents" (stain and mother liquor) in the two cases.

We have not measured the density of the dried glucose/PTA layer.

N. Unwin: The analysis of the same structure to the same resolution has previously been reported (Morgan *et al*., 1995). In that paper, the authors "present details of improved methodology to extract and evaluate the original data and also to assess the statistical significance of features in the three-dimensional map." One aspect of the present paper not discussed thoroughly in the earlier work is a description/ analysis of the sniffer algorithm. This algorithm represents a methodological advance that should benefit other people in the field.
Authors: In this paper, we are attempting to present some results on glucose/PTA embedment which might, in some cases, be preferred to the use of vitreous ice embedment. In analyzing these data, we developed the sniffer algorithm and tested it on the glucose/PTA embedded samples which have a bit stronger signal-to-noise ratio than the ice embedded samples, and for which we have a large set of images recorded close enough to focus that the CTF does not affect the data until beyond the disappearance of all useful signal. Although the concept of the sniffer has been mentioned before (Morgan and DeRosier, 1993), it was not used in our analysis of vitreous ice embedded filaments (Morgan *et al*., 1995).

A. Steven: Based on our limited but decidedly negative experience with data digitized on an Eikonix scanner, I found it surprising that the glucose/PTA I data scanned with an Eikonix scanner appeared to be as good as data from the same series of micrographs scanned on a Joyce Loebl densitometer at the Laboratory for Molecular Biology, MRC. Could the authors comment on that?

Authors: The Eikonix scanner does not measure densities as well as the Joyce Loebl scanner. We think the Eikonix degrades both moderate and high resolution amplitudes but not phases, so that although both the signal and the noise amplitudes are suppressed, averaging allows us to extract useful signal from the Eikonix scanned data. Our comparison of images scanned with the Joyce-Loebl and the Eikonix was not used to provide information about the quality of data from the two scanners, but rather to show that the glc/PTA I and glc/PTA II images are inherently different.