the defocus in tilted micrographs varies depending on the position in the field, often forcing users to restrict the particle selection only to regions in the vicinity of the tilt axis. However, tilting establishes geometrical relations between different projections of the same particle, unambiguously allowing for robust determination of an initial 3D model and the handedness of the quaternary structure of the complex.

Electron microscope images can be either recorded on the film and subsequently converted to digital format, or they can be recorded using a charge-coupled device (CCD) camera in a digital format directly on a microscope. In either case, it is necessary to select the magnification of the microscope and the eventual pixel size of the digitized data before the data-collection session. High magnification can potentially yield high-resolution data, but at the same time it decreases the yield of particles. Lower magnification values can be used when images are recorded on film, which does not attenuate high spatial frequencies to the same extent as CCD cameras tend to do.

The pixel size has to be adjusted according to the expected resolution of the final structure. Although it is tempting to adopt a small pixel size (in the hope of achieving high resolution of the results), in most cases this is counterproductive, as it results in very large computer files that are difficult to handle and in excessively long data-processing times. Theoretically, the optimum pixel size is tied to the maximum frequency present in the data by Shannon's sampling theorem, which states that no information is lost if the signal is sampled at twice the maximum frequency present in the signal, and no additional information is gained by sampling using higher frequency. Thus, if the expected resolution is 12 Å, it should be sufficient to use a pixel size (on the specimen scale) of 6 Å. In practice, various image-processing operations performed during alignment of the data and 3D reconstruction of the complex significantly lower the range of useful frequencies. This is because in currently available single-particle reconstruction software packages rather unsophisticated interpolation schemes are employed, which were selected mainly for the speed of calculations. Therefore, it is advisable to oversample the data by a factor of 1.5 or even 3.0. For an expected resolution of 12 Å this corresponds to pixel sizes of 4 and 2 Å, respectively.

The windowed particles have to be normalized to adjust the image densities to a common framework of reference. The reason for this step is that microscopy conditions are never exactly the same and also within the same micrograph field the background densities can vary by a significant margin due to uneven ice thickness and other factors. A sensible approach to normalization is to assume that the statistical distribution of noise in areas surrounding particles should be the same (Boisset *et al.*, 1993). Hence a large portion of one of the micrographs from the processed set is selected and a reference histogram of its pixel values is generated. Next, assuming a linear transformation of pixel values, the two parameters of this transformation are found in such a way that the histogram of the transformed pixel values surrounding the particle optimally matches the reference histogram using χ^2 statistics as a discrepancy measure.

2.5.7.4. Assessment of the data quality and estimation of the image formation parameters

The initial assessment of the quality of the micrographs is usually performed during the data collection and in most cases before the micrographs are digitized. The micrographs are examined visually and those that have noticeable drift, astigmatisms, noticeable contamination or simply too low a number of particles to justify further analysis are simply discarded. After digitization of the accepted micrographs, the first step is estimation of the power spectrum, which will be examined for the presence of Thon rings (thus confirming that the micrograph is indeed usable) and astigmatism.

The method of averaged overlapping periodograms (Welch, 1967) is commonly used in EM to calculate the power spectrum. It is designed to improve the statistical properties of the estimate by taking advantage of the fact that when K identically distributed independent measurements are averaged, the variance of the average is decreased with respect to the individual variance by the ratio $1/K$. Thus, instead of calculating a periodogram (squared moduli of the discrete Fourier transform) of the entire micrograph field, one subdivides it into much smaller windows, calculates their periodograms and averages them. Typically, one would chose a window size of 512 × 512 pixels and an overlap of 50%, which will result in the reduction of the variance of the estimate to few percent with respect to the variance of the periodogram of the entire field (Fernandez *et al.*, 1997; Zhu *et al.*, 1997). Further reduction of the variance is achieved by rotational averaging of the 2D power-spectrum estimate. The resulting one-dimensional (1D) profile is finally used in the third step of our procedure.

For a set of micrographs the power spectra can be evaluated either visually or computationally in an automated fashion. Of main concern are the presence of Thon rings, the astigmatism and the extent to which Thon rings can be detected. Although in principle astigmatic data could be used in subsequent analysis (in fact, astigmatism could be considered advantageous, as particles from the same micrograph would contain complementary information in Fourier space), in practice they are discarded as currently there is no software that can process astigmatic data efficiently. The extent of Thon rings indicates the 'resolution' of the data, *i.e.*, the maximum frequency to which information in the data can be present.

A number of well established programs can assist the user in the calculation of power spectra and automated estimation of defocus and astigmatism (Huang *et al.*, 2003; Mindell & Grigorieff, 2003; Sander *et al.*, 2003; Mallick *et al.*, 2005). Given the analytical form of the CTF [(2.5.7.4)], the problem is solved by a robust fitting of the CTF parameters such that the analytical form of the CTF matches the power spectrum of the micrograph. Usually, the steps employed are: (1) robust estimation of the power spectrum; (2) calculation of the rotational average of the power spectrum; (3) subtraction from this rotational average of the slowly decreasing background [roughly corresponding to P_B in (2.5.7.5)]; (4) fitting of the defocus value Δf_n using known settings of the microscope (voltage, spherical aberration constant, ...) and usually assuming a constant and known value of the amplitude contrast ratio q (for cryo-EM data, q should be in the range 0.02–0.10); and (5) using the established defocus value Δf_n, analysis of the 2D power spectrum and fitting of the astigmatism amplitude and angle while refining the defocus. As long as the defocus value is not too small and there are at least two detectable zeros of the CTF, all available programs give very good and comparable results.

In some single-particle packages, the automated calculation of defocus is integrated with the estimation of additional characteristics of the image-formation parameters that are required for advanced application of a Wiener filter [(2.5.7.18)] (Saad *et al.*, 2001; Huang *et al.*, 2003), *i.e.*, the power spectra of two noise distributions P_S and P_B and the envelope function of the microscope E_n for each micrograph. A possible approach is to select slowly varying functions and fit their parameters to match the estimates of P_S, P_B and P_d obtained from the data. Finally, it is necessary to have a description of the 1D rotationally averaged power spectrum of the complex P_f. One possibility is to carry out X-ray solution scattering experiments (Gabashvili *et al.*, 2000; Saad *et al.*, 2001) that yield a 1D power spectrum of the complex in solution. However, these experiments require large amounts of purified sample and the accuracy of the results in terms of the overall fall-off of the power spectrum can be disputed. For the purpose of cryo-EM, a simple approximation of the protein power spectrum by analytical functions is satisfactory.