cation second' approach. It is possible to reverse this order by using invariants with the supporting rationale that once approximately homogeneous classes of images were found, it should be easy to align them subsequently as within each class they will share the same motif.

A practical approach to reference-free alignment known as *alignment by classification* (Dube *et al.*, 1993) is based on the observation that for a very large data set and centred particles one can expect that although the in-plane rotation is arbitrary, there is a high chance that at least some of the similar images will be in the same rotational orientation. Therefore, in this approach the images are first (approximately) centred, then subjected to classification, and subsequently aligned.

In its simplest form, the *multireference alignment* belongs to the class of *supervised classification* methods: given a set of templates (*i.e.*, reference images; these can be selected unprocessed particle projections, or class averages that resulted from preceding analysis, or projections of a previously determined EM structure, or projections of an X-ray crystallographic structure), each of the images from the available data sets is compared (using a selected discrepancy measure) with all templates and assigned to the class represented by the most similar one. Equally often multireference alignment is understood as a form of *unsupervised classification*, more precisely *K*-means classification, even if the description is not formalized in terms of the latter. Given a number of initial 2D templates, the images are compared with all templates and assigned to the most similar one. New templates are calculated by averaging images assigned to their predecessors and the whole procedure is repeated until a stable solution is reached.

2.5.7.7. Initial determination of 3D structure using tilt experiments

The 2D analysis of projection images provides insight into the behaviour of the protein on the grid in terms of the structural consistency and the number and shape of projection images. In order to obtain 3D information, it is necessary to find geometrical relations between different observed 2D images. The most robust and historically the earliest approach is based on tilt experiments. By tilting the stage in the microscope and acquiring additional pictures of the same area of the grid it is possible to collect projection images of the same molecule with some of the required Eulerian angles determined accurately by the setting of the goniometer of the microscope.

In random conical tilt (RCT) reconstruction (Radermacher *et al.*, 1987), two micrographs of the same specimen area are collected: the first one is recorded at a tilt angle of ~50° while the second one is recorded at 0° (Fig. 2.5.7.3). If particles have preferred orientation on the support carbon film (or within the amorphous ice layer, if no carbon support is used), the projections of particles in the tilted micrographs form a conical tilt series. Since in-plane rotations of particles are random, the azimuthal angles of the projections of tilted particles are also randomly distributed; hence the name of the method. The untilted image is required for two reasons: (i) the particle projections from the untilted image are classified, thus a subset corresponding to possibly identical images can be selected ensuring that the projections originated from similar and similarly oriented structures; and (ii) the in-plane rotation angle found during alignment corresponds to the azimuthal angles in three dimensions (one of the three Eulerian angles needed). The second Eulerian angle, the tilt, is either taken from the microscope setting of the goniometer or calculated based on geometrical relations between tilted and untilted micrographs. The third Eulerian angle corresponds to the angle of the tilt axis of the microscope stage and is also calculated using the geometrical relations between two micrographs. In addition, it is necessary to centre the particle projections selected from tilted micrographs; although various correlation-based schemes have been proposed,

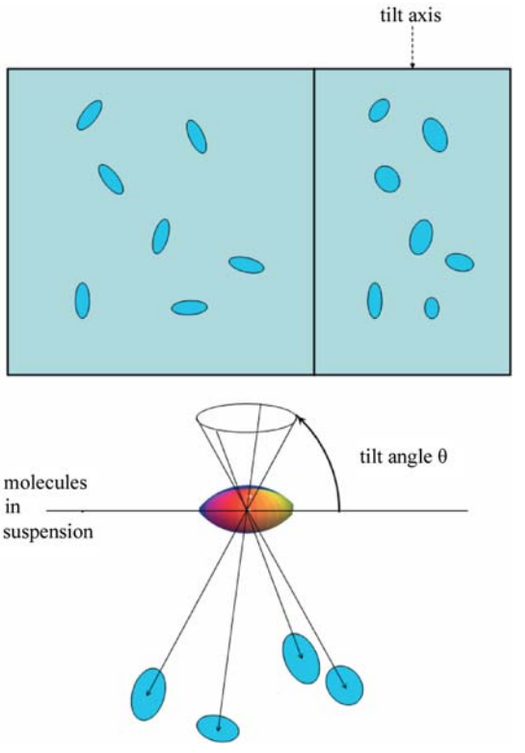

Fig. 2.5.7.3. Principle of random conical tilt reconstruction. A tilt pair of images of the same grid area is collected. By aligning the particle images in the untilted micrograph (left), the Eulerian angles of their counterparts in the tilted micrograph (right) are established. The particle images from the tilted micrograph are used for 3D reconstruction of the molecule (bottom). The set of projections form a cone in Fourier space; information within the cone remains undetermined.

the problem is difficult as the tilt data tend to be very noisy and have very low contrast.

Given three Eulerian angles and centred tilted projections, a 3D reconstruction is calculated. There are numerous advantages of the RCT method. (i) Assuming the sign of the tilt angle is read correctly (it can be confirmed by analysing the defocus gradient in the tilted micrographs), the method yields a correct hand of the structure. (ii) With the exception of the in-plane rotation of untilted projections, which can be found relatively easily using alignment procedures, the remaining parameters are determined by the experimental settings. Even if they are not extremely accurate, the possibility of a gross error is eliminated, which positively distinguishes the method from the *ab initio* computational approaches that use only untilted data. (iii) The computational analysis is entirely done using the untilted data, which have high contrast. (iv) The RCT method is often the only method of obtaining 3D information if the molecule has strongly preferential orientation and only one view is observed in untilted micrographs. The main disadvantage is that the conical projection series leaves a significant portion of the Fourier space undetermined. This follows from the central section theorem [equation (2.5.6.8) of Section 2.5.6]: as the tilt angle is less than 90°, the undetermined region can be thought to form a cone in three dimensions and is referred to as the missing cone. The problem can be overcome if the molecule has more than one preferred orientation. Subsets of particles that have similar untilted appearance (as determined by clustering) are processed independently and for each a separate 3D structure is calculated. If the preferred orientations are sufficiently different, *i.e.*, the orientations of the original particles in three dimensions are sufficiently different in terms of their angles with respect to the z axis, the 3D structures can be aligned and merged, all but eliminating the problem of the missing cone and yielding a robust, if resolution-limited, initial model of the molecule (Penczek *et al.*,

1994). It should be noted that RCT by itself almost never results in a high-resolution 3D model of the molecule. This is due to a variety of reasons, the main ones being the already mentioned poor quality of high-tilt data and difficulties with the collection of large numbers of high-quality tilted micrographs (they are often marred by drift).

In cases when the molecule does not have well defined preferred orientations, it is possible to use electron tomography to obtain the initial model. In this method, a single-axis tilt series of projection images of the same specimen area is collected using an angular step of ~2° and a maximum tilt angle not exceeding 60° (Crowther, DeRosier & Klug, 1970). The single-axis tilt data-collection geometry yields worse coverage of the Fourier space than the RCT method, leaving missing wedges uncovered (Penczek & Frank, 2006). This results in severe artifacts in real space, which make smaller objects virtually unrecognizable. The situation can be largely rectified using so-called double-axis tomography, in which a second single-axis tilt series of data are collected after rotating the specimen grid in-plane by 90° (Penczek *et al.*, 1995). This reduces the undetermined region to a missing pyramid and makes the resolution almost isotropic in the *xy* plane.

The tomographic projection data have to be aligned. This is done using either correlation techniques that enforce pairwise alignment of images (Frank & Mcewen, 1992) or by taking advantage of fiducial markers and enforcing their consistency with respect to a 3D model (Lawrence, 1992; Penczek *et al.*, 1995). In the application to single-particle work, it is possible to use locations of protein in the micrographs as markers. After the 3D reconstruction is calculated, regions collecting individual molecules are windowed from the volume and all molecules are aligned in three dimensions (Walz *et al.*, 1997). While generally robust, the procedure is labour- and computer-intensive. Unlike RCT, where only two exposures of the same field are required, electron tomography may require over one hundred images, raising serious concerns about radiation damage. Moreover, most of the data have to be collected at high tilt angle, thus are of lower quality. Particularly troublesome is alignment of 3D molecules deteriorated by the missing wedge/pyramid artifacts, with the directions of artifacts different for each object. However, when successful, electron tomography yields a very good initial model of the molecule, free from missing-Fourier-space-related artifacts and with defined handedness.

2.5.7.8. Ab initio 3D structure determination using computational methods

The experiment-based methods of initial 3D structure determination (RCT and electron tomography) are quite powerful, but rather challenging to employ in practice. Particularly frustrating is the fact that a large volume of difficult-to-record tilt data have to be collected, even though they cannot be used for subsequent high-resolution work. Therefore, whenever possible, preference is given to computational methods in which 3D geometrical relations between particle projections are established using various mathematical approaches using only untilted data.

The most straightforward approach and historically the earliest is based on the central section theorem [equation (2.5.6.8) of Section 2.5.6]: because Fourier transforms of 2D projections of a 3D object are the central section of the 3D Fourier transform of this object, it is a straightforward consequence that Fourier transforms of any two projections intersect along a line, henceforth called a *common line*. (Two trivial exceptions are the case of projections in the same direction, in which their Fourier transforms coincide with possible differences in in-plane rotation, and the case of projections in opposite directions, in which they are mirror versions of each other.) This fact was originally used by Crowther, DeRosier & Klug (1970) to solve the structure of viruses with icosahedral (60-fold) symmetry. In this case, the Fourier transform of each projection intersects itself (or rather the symmetry-related copies of itself) 37 times with the exception of degenerate cases of projections in directions of one of three symmetry axes, in which cases the number of common lines is less. Thus, it is possible to find the orientation of a single projection with respect to the chosen system of symmetry axes.

For asymmetric objects a set of three projections that do not intersect along the same line (which would correspond to the single-axis tilt geometry) uniquely determine their respective orientations (with the exception of the overall rotation, which remains arbitrary, and the handedness of the solution, which remains undetermined). Indeed, three projections span three common lines, and each common line yields two angles: for each of the intersecting sections it is the angle between the *x* axis in the system of coordinates of this section and the common line in the plane of this section. Thus, we have a total of six angles. At the same time, by arbitrarily setting the orientation of the first projection in 3D space to three Eulerian angles equal to zero (or the corresponding rotation matrix $\mathbf{R}_1 = \mathbf{I}$), we need to determine two rotation matrices $\mathbf{R}_2$ and $\mathbf{R}_3$ (or two sets of three Eulerian angles) for the remaining two projections, respectively. So, given six in-plane angles we have to find a solution for six Eulerian angles. Let the angle of the common line between the *i*th and *j*th projection in the plane of the *i*th projection be α_{ij} and the corresponding unit vector in the plane of *i*th projection be

$$\mathbf{n}_{ij} = [\cos\alpha_{ij} \quad -\sin\alpha_{ij} \quad 0]^T, \quad i, j = 1, 2, 3, \quad i \neq j, \tag{2.5.7.12}$$

where we added the third coordinate for convenience. The orientations of unit vectors $\mathbf{n}_{ij}$ in 3D space have to be related by rotation matrices; for example, vector $\mathbf{n}_{21}$ (the direction of the common line between the first and second projection in the plane of the second projection) should coincide with vector $\mathbf{n}_{12}$ (the direction of the same common line, but in the plane of the first projection) upon rotation by (unknown) matrix $\mathbf{R}_2$. All possible relations are

$$\begin{aligned} \mathbf{R}_2\mathbf{n}_{21} &= \mathbf{n}_{12} \\ \mathbf{R}_2\mathbf{n}_{23} &= \mathbf{R}_3\mathbf{n}_{32} \\ \mathbf{R}_3\mathbf{n}_{31} &= \mathbf{n}_{13}. \end{aligned} \tag{2.5.7.13}$$

Equations (2.5.7.13) have two solutions corresponding to two different hands of the molecule [for details see Farrow & Ottensmeyer (1992)].

The common-lines method works very well in the absence of noise. However, even a modest amount of noise can yield quite erroneous results or no results at all. The reason is that the solution of (2.5.7.13) is highly nonlinear with respect to the six given angles α_{ij} and small errors in the location of peaks in cross-correlation functions can lead to quite large discrepancies from the correct solution. The main difficulty with the application of the common-lines method is that the analytical solution in the form of (2.5.7.13) exists only for three projections (Goncharov *et al.*, 1987).

As a working approach to *ab initio* structure determination, the common-lines method has been implemented under the name of *angular reconstitution* in *IMAGIC* (van Heel, 1987*a*; van Heel *et al.*, 1996). In order to reduce sensitivity to noise, the method is applied not to individual projection images, but to class averages resulting from the multireference alignment of input data (van Heel *et al.*, 2000). In order to overcome the problem of the lack of solution for the larger than three number of projections, the user has to begin with selection of three judiciously chosen class averages, obtain the solution using (2.5.7.13) and subsequently include (angle) additional class averages using a brute-force approach, in which the Eulerian angles of the new projection are