

2. RECIPROCAL SPACE IN CRYSTAL-STRUCTURE DETERMINATION

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 $\sim 2^\circ$ 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

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

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

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

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, 1987a; 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

2.5. ELECTRON DIFFRACTION AND ELECTRON MICROSCOPY IN STRUCTURE DETERMINATION

calculated using a similarity measure based on common lines with the already-angled set serving as a reference.

Some of the disadvantages of the angular reconstitution were addressed in the common-lines-based method for determining orientations for $N > 3$ particle projections simultaneously (Penczek *et al.*, 1996). In this method, the problem is formulated in terms of minimization of the variance of the 3D structure, as expressed in terms of common-lines discrepancy between N projections. In a sense, the design of the method is the exact opposite of the ‘standard’ common-lines approach: instead of trying to determine the Eulerian angles (rotation matrices \mathbf{R}_i) based on angles α_{ij} of common lines in the planes of the projections, one assumes that rotation matrices \mathbf{R}_i are known, finds the set of angles α_{ij} of common lines and computes the overall discrepancy along these lines. For a pair of projections i and j , the in-plane angles of common lines are found by solving the system of equations

$$\mathbf{R}_i \mathbf{n}_{ij} = \mathbf{R}_j \mathbf{n}_{ji} \quad (2.5.7.14)$$

for α_{ij} and α_{ji} . The discrepancy minimized in the method is the variance of the 3D structure that, by analogy to the 2D case (2.5.7.10) and (2.5.7.11) is

$$L_{\text{cl}}(\{F\}, \{\mathbf{R}\}) = \sum_{m=0}^M \sum_{k=1}^N \|F_k(u_m, \alpha; \mathbf{R}_k) - \langle F \rangle_k\|^2 u_m^2 \Delta u \Delta \Omega_{kl},$$

$$\langle F \rangle_k = \frac{1}{N-1} \sum_{l=1, l \neq k}^N F_l(u_m, \alpha; \mathbf{R}_l), \quad (2.5.7.15)$$

where L_{cl} is written in Fourier 3D polar coordinates and $F_k(u_m, \alpha; \mathbf{R}_k)$ is the Fourier transform of the k th projection in 2D polar coordinates (u_m, α) with the orientation in 3D Fourier space given by the rotation matrix \mathbf{R}_k . All Fourier planes F_k are considered to have zero thickness, so all discrepancies are calculated only along common lines and the ‘partial average’ $\langle F \rangle_k$ is in fact an arrangement of $N-1$ Fourier planes in 3D space. An approximation to $u_m^2 \Delta u \Delta \Omega_{kl}$ is calculated by equating the values of $\Delta \Omega_{kl}$ to the areas of the Voronoi diagram cells constructed on a unit sphere for points of intersection of common lines with this unit sphere (see Section 2.5.6.6). Generally the method performs very well, particularly if the projection images cover 3D angular space evenly.

Some macromolecules, particularly those that have an elongated barrel-like shape, will have a strongly preferential orientation with respect to the direction of electron beam showing only what are often called ‘side views’, *i.e.*, projections perpendicular to rotation along one axis corresponding to single-axis tilt data-collection geometry. These orthoaxial projections form single-axis tilt reconstruction geometry. In this case, Fourier transforms of all projections share only one common line, the line coinciding with the rotation axis, and clearly the common-lines-based method is not applicable to the *ab initio* structure determination. To cope with this situation, a method termed *Sidewinder* was developed (Pullan *et al.*, 2006). It is based on the observation that a Fourier transform of a finite object with a diameter D can be considered to have a nonzero thickness $1/D$ (Fig. 2.5.6.5). Thus, if the angle between two central sections of the 3D Fourier transform of this object, as derived from 2D Fourier transforms of its projections, is not too large, then these two sections will share information in Fourier space that is proportional to the amount of overlap of the two ‘slabs’ in Fourier space. Using this observation, the general idea employed in *Sidewinder* is to calculate pairwise cross-correlation coefficients (CCCs) between class averages of side views and to use this information to deduce the values of the azimuthal Eulerian angles using the Monte Carlo minimization method (Fishman, 1995).

For structures that have reasonably high symmetry and for those for which it is possible to collect high-quality EM data, it is sometimes possible to determine the initial structure using the 3D projection alignment method, which will be described in the next section. However, the approach is extremely computationally intensive and it is virtually impossible to try the method repeatedly to verify that the approach converges to more-or-less the same 3D structure, as is recommended for other *ab initio* methods described in this section. When the method is successful it is quite powerful, as an intermediate resolution structure can be obtained without going through intermediate and quite laborious steps of analysis of the data. A word of caution is warranted: with the direct method, unless there is external evidence that the obtained structure is correct, it is possible to obtain a self-consistent but entirely incorrect model of the molecule.

In the absence of reliable objective measures of the correctness of the structure, one can apply common sense in order to spot definitely improbable 3D maps. Given the mass of the complex it is possible to calculate the corresponding volume, and thus the threshold at which the map should be examined (Section 2.5.7.11). If at this threshold the mass density is discontinuous or there are pieces of mass surrounding the structures, the map is most likely to be incorrect. Similarly, strong directional artifacts appearing as streaks permeating the structure indicate that either the collected projection images are dominated by one or two views of the structure or that the angular assignment is incorrect. In addition, the 3D map should be centred in the window box; although the centring is not strictly speaking a mathematical requirement for a successful reconstruction, all single-particle structure-determination software packages take advantage of the fact that for centred objects orientation searches are easier to perform. So, if the map is not centred it is a clear indication of the failure of the procedure. Finally, for symmetric structures there should be no large pieces of mass on the symmetry axes.

2.5.7.9. Refinement of a 3D structure

Given an initial low-resolution model of the 3D structure and the data set of 2D projection images of the complex that have Fourier-space information extending beyond the resolution of the model, it is possible to refine the structure such that the full extent of the resolution information in the data will be utilized. In some cases, it is also possible to use as an initial structure in the refinement procedure a structure of a homologous protein, thus avoiding the process of *ab initio* structure determination altogether. The goal of the refinement is to find such orientation parameters for each of the particle projections for which (2.5.7.6) is minimized. There exist various implementations of the structure-refinement strategy and they can be roughly divided into those that perform exhaustive searches for all five orientation parameters (two translations and three Eulerian angles per 2D projection image) and those that perform local searches, usually by employing gradient information. Finally, the strategies may differ in how the correction for the CTF is implemented.

The original 3D projection-matching strategy (Penczek *et al.*, 1994) is based on the observation that given an ideal structure f and the necessary parameters of the CTF and image-formation model, it is straightforward to find five orientation parameters for each projection image. One begins with the determination of the sufficient angular step: assuming the structure is properly sampled at the Nyquist frequency and has a real-space radius of r voxels, the angular step is given by

$$\delta\theta \cong \arctan(1/r). \quad (2.5.7.16)$$

Next, keeping in mind that projection directions are parametrized by two Eulerian angles (φ, θ) , one generates a set of projection directions quasi-uniformly distributed over half a unit