*International Tables for Crystallography* (2006). Vol. F, Section 2.1.1, pp. 45–46.

# 2. BASIC CRYSTALLOGRAPHY

## 2.1. Introduction to basic crystallography

BY J. DRENTH

### 2.1.1. Crystals

It is always amazing to see how large molecules, such as proteins, nucleic acids and their complexes, order themselves so neatly in a crystalline arrangement. It is surprising because these large molecules have irregular surfaces with protrusions and cavities, and hydrophilic and hydrophobic spots. Nevertheless, they pack themselves into an orderly arrangement in crystals of millimetre sizes.

Crystals of biological macromolecules are, like most other crystals, not ideal. The X-ray diffraction pattern fades away at diffraction angles corresponding to lattice-plane distances between 1 and 2 Å or even worse. This is not so surprising, since protein crystals are relatively soft. The interaction energy between protein molecules in crystals is of the order of $63 \times 10^{-21}$ J per protein molecule, or approximately 15 $kT$ (Haas & Drenth, 1995). This corresponds to about ten hydrogen bonds, four salt bridges, or a 400 $\text{Å}^2$ buried hydrophobic surface. Although this energy might not be very different from crystalline interactions between small molecules, the large size of the protein molecules or macromolecular assemblies makes the crystals much more sensitive to distorting forces. Irregularities in the crystal lattice can also stem from the incorporation of impurities – either foreign substances or slightly denatured molecules from the parent protein. Moreover, some molecules may be incorrectly oriented, because the difference in interaction energy between different orientations is rather small. Also, amino-acid side chains assume more than one conformation. These are static irregularities. In addition, dynamic disorder exists: parts of the macromolecule are flexible and affect the X-ray diffraction pattern just as the temperature does.

By neglecting distortions caused by lattice imperfections, crystals are found to have a repeating unit, the unit cell, with basis vectors **a**, **b** and **c**, and angles $\alpha$, $\beta$ and $\gamma$ between them (Fig. 2.1.1.1). The enormous number of unit cells in a crystal are stacked in three dimensions, in an orderly way, with the origins of the unit cells forming a grid or lattice. In Fig. 2.1.1.2, part of a crystalline lattice containing $5 \times 3 \times 3$ unit cells is drawn.

It is customary to call the direction along the unit-cell vector **a** the $x$ direction in the lattice; similarly, $y$ is along **b**, and $z$ along **c**.

Crystallographers use a simple system to indicate the planes in a crystal lattice. For instance, the plane containing the unit-cell vectors **a** and **b** is called (001), and the plane containing the vectors **b** and **c** is called (100). The plane (010) contains the vectors **a** and **c**. It should be pointed out that these planes are not limited to one unit cell, but extend through the entire crystal. Moreover, each of these three planes is only one member of a set of parallel and equidistant planes: the set (001), the set (100) and the set (010). For each set, the lattice planes pass through all lattice points, where the lattice points are at the corners of the unit cells (see Fig. 2.1.1.2). Besides the sets of planes (001), (100) and (010), many more sets of parallel and equidistant planes can be drawn through the lattice points. In Fig. 2.1.1.3, this is done for a two-dimensional lattice. Lattice planes always divide the unit-cell vectors **a**, **b** and **c** into a number of equal parts. If the lattice planes divide the **a** vector of the unit cell into $h$ equal parts, the first index for this set of planes is $h$. The second index, $k$, is related to the division of **b**, and the third index, $l$, to the division of **c**. If the set of lattice planes is parallel to a basis unit-cell vector, the corresponding index is 0. Indices for lattice planes are given in parentheses. They should not be confused with directions of vectors connecting lattice points; these are given in square brackets: [$uvw$], where $u$ is the coordinate in the **a** direction expressed as the number of **a**'s, $v$ in the **b** direction expressed as the number of **b**'s and $w$ in the **c** direction expressed as the number of **c**'s. $u$, $v$ and $w$ are taken as the simplest set of whole numbers. For instance, [100] is along **a**; [200] has the same direction, but [100] is used instead. [111] points from the origin to the opposite corner of the unit cell.

The choice of the unit cell is not unique and, therefore, guidelines have been established for selecting the standard basis vectors and the origin. They are based on symmetry and metric considerations:

(1) The axial system should be right-handed.

(2) The basis vectors should coincide as much as possible with directions of highest symmetry.

(3) The cell taken should be the smallest one that satisfies condition (2).

(4) Of all lattice vectors, none is shorter than **a**.

(5) Of those not directed along **a**, none is shorter than **b**.

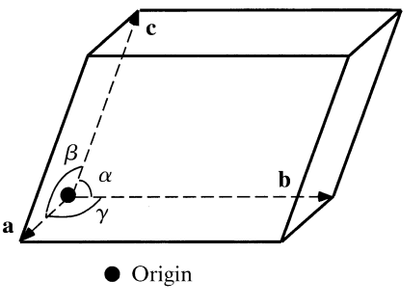



Fig. 2.1.1.1. One unit cell with axes **a**, **b** and **c**. The angles between the axes are $\alpha$, $\beta$ and $\gamma$. Note that the axial system is right-handed. Reproduced with permission from Drenth (1999). Copyright (1999) Springer-Verlag.

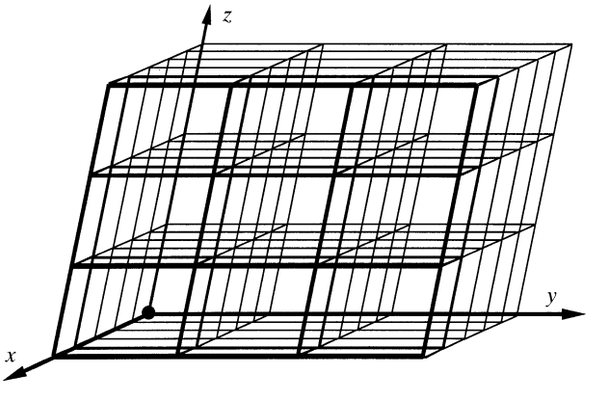
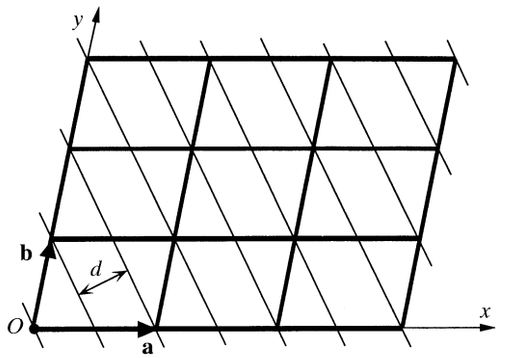
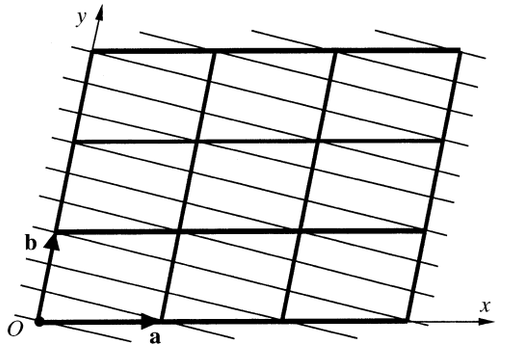



Fig. 2.1.1.2. A set of $5 \times 3 \times 3$ unit cells. The points where the lines intersect are called lattice points. The axes $x$ and $y$ form a (001) plane, which is one member of the set of parallel and equidistant (001) planes; $y$ and $z$ form a (100) plane, and $z$ and $x$ a (010) plane. Reproduced with permission from Drenth (1999). Copyright (1999) Springer-Verlag.

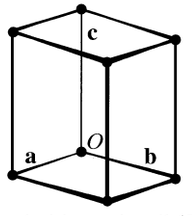



(*a*)

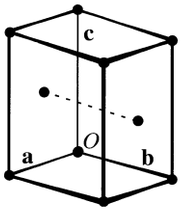



(*b*)

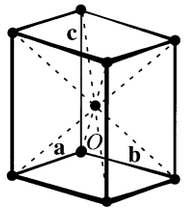
Fig. 2.1.1.3. A two-dimensional lattice with 3 × 3 unit cells. In both (*a*) and (*b*), a set of equidistant parallel lattice planes is drawn. They pass through all lattice points. Lattice planes always divide the unit-cell axes into a whole number of equal parts – 1, 2, 3 *etc*. For instance, in (*a*), the vector **a** of the unit cell is cut into two parts, and the vector **b** into only one part. This set of planes is then given the indices $h = 2$ and $k = 1$. In three dimensions, there would be a third index, *l*. In (*b*), the set of lattice planes has the indices $h = 1$ and $k = 3$. In general, lattice planes have the indices (*hkl*), known as Miller indices. If a set of lattice planes is parallel to an axis, the corresponding index is 0. For instance, (001) is the set of planes parallel to the unit-cell vectors **a** and **b**. Note that the projection of **a**/*h* on the line normal to the lattice plane is equal to the lattice-plane distance *d*. This is also true for **b**/*k*. Reproduced with permission from Drenth (1999). Copyright (1999) Springer-Verlag.



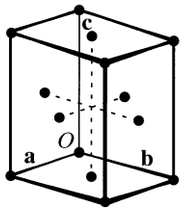

A primitive unit cell (*P*)




A unit cell centred in the (010) planes (*B*)




A body-centred unit cell (*I*)




A face-centred unit cell (*F*)

Fig. 2.1.1.4. Non-centred and centred unit cells. Reproduced with permission from Drenth (1999). Copyright (1999) Springer-Verlag.

(6) Of those not lying in the *ab* plane, none is shorter than **c**.

(7) The three angles between the basis vectors **a**, **b** and **c** are either all acute ($< 90°$) or all obtuse ($\geq 90°$).

It should be noted that the rules for choosing **a**, **b** and **c** are not always obeyed, because of other conventions (see Section 2.1.3). Condition (3) sometimes leads to a centred unit cell instead of a primitive cell. Primitive cells have only one lattice point per unit cell, whereas non-primitive cells contain two or more lattice points. They are designated *A*, *B* or *C* if opposite faces of the cell are centred: *A* for *bc* centring, *B* for *ac* centring and *C* for *ab* centring. If all faces are centred, the designation is *F*, and if the cell is body-centred, it is *I* (Fig. 2.1.1.4).

### 2.1.2. Symmetry

A symmetry operation can be defined as an operation which, when applied, results in a structure indistinguishable from the original one. According to this definition, the periodic repetition along **a**, **b** and **c** represents translational symmetry.

In addition, rotational symmetry exists, but only rotational angles of 60, 90, 120, 180 and 360° are allowed (*i.e*. rotation over 360/*n* degrees, where *n* is an integer). These correspond to *n*-fold rotation axes, with $n = 6, 4, 3, 2$ and 1 (identity), respectively. Rotation axes with $n = 5$ or $n > 6$ are not found as crystallographic symmetry axes, because translations of unit cells containing these axes do not completely fill three-dimensional space. Another type of rotational symmetry axis is the screw axis. It combines a rotation with a translation. For a twofold screw axis, the translation is over 1/2 of the unit-cell length in the direction of the axis; for a threefold screw axis, it is 1/3 or 2/3 *etc*. In this way, the translational symmetry operators can be obeyed. The requirement that translations are 1/2, 1/3, 2/3 *etc*. of the unit-cell length does not exist for individual objects that are not related by crystallographic translational symmetry operators. For instance, an $\alpha$-helix has 3.6 residues per turn.

Besides translational and rotational symmetry operators, mirror symmetry and inversion symmetry exist. Mathematically, it can be proven that not all combinations of symmetry elements are allowed, but that 230 different combinations can occur. They are the space groups which are discussed extensively in *IT* A (1995). The graphical and printed symbols for the symmetry elements are also found in *IT* A (pp. 9–10).

Biological macromolecules consist of building blocks such as amino acids or sugars. In general, these building-block structures are not symmetrical and the mirror images of the macromolecules do not exist in nature. Space groups with mirror planes and/or inversion centres are not allowed for crystals of these molecules, because these symmetry operations interchange right and left hands. Biological macromolecules crystallize in one of the 65 enantio-

Table 2.1.2.1. *The most common space groups for protein crystals*

Situation as of April 1997; data extracted from the Protein Data Bank and supplied by Rob Hooft, EMBL Heidelberg.

| Space group | Occurrence (%) |
|---|---|
| $P2_12_12_1$ | 23 |
| $P2_1$ | 11 |
| $P3_221$ | 8 |
| $P2_12_12$ | 6 |
| $C2$ | 6 |



references