Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

On optimal placement of molecules in the unit cell

Zbigniew Dauter

Synchrotron Radiation Research Section, MCL, National Cancer Institute, Argonne National Laboratory, Argonne, IL 60439, USA

Correspondence e-mail: dauter@anl.gov

Received 19 December 2012
Accepted 28 January 2013

Dedicated to Professor Zofia Kosturkiewicz, who taught the author the principles of crystallography.

There are currently no rules for a unified, standard way of placing macromolecular structures in the crystal lattice. An analysis of all possible symmetry-equivalent representations of molecular structures in various space groups leads to the concept of the anti-Cheshire symmetry and suggests that the center of a unique structural motif can always be placed within the selected asymmetric unit of the anti-Cheshire cell. The placement of structures according to this suggestion will ensure uniformity of presentation of all structurally equivalent Protein Data Bank models and will therefore diminish the possibility of confusing less crystallographically knowledgeable users of the PDB. The anti-Cheshire cells and their asymmetric units are defined and tabulated for all 65 space groups relevant to macromolecular crystallography that exhibit only rotational symmetry operations.

1. Introduction

Owing to the periodic and symmetric nature of crystals, a full description of their structure requires only the definition of the contents of the asymmetric unit (ASU), which encompasses the smallest unique structural motif that is repeated throughout the entire crystal according to its space-group symmetry. The limits of the ASU for each space group are defined in Volume *A* of *International Tables for Crystallography* (2005) but, in general, the ASUs can be selected in various ways. Usually they are presented as parallelepipeds, although in cubic space groups their shapes are more complicated. In molecular crystals such as those of macromolecules, each molecule or, more generally, each unique structural motif with its immediate surroundings forms the ASU (Fig. 1). Richards (1974) analyzed the volume occupied by several protein molecules in their crystals using the Voronoi tessellation algorithm (Voronoi, 1908), in which all points with the closest proximity to (any atom of) a reference molecule form an irregular ASU reflecting the molecular shape. All versions of the ASU have the same volume, which is equal to the appropriate fraction of the cell volume, *i.e.* the unit-cell volume divided by the number of general symmetry positions.

From a strictly crystallographic point of view, it does not matter which asymmetric unit is represented as unique and how it is related to the lattice origin. Obviously, it is beneficial if all atoms of the model are grouped to form individual chemical molecules, even if the coordinates of some atoms extend outside the standard ASU. Such a (bio)chemically sensible strategy makes stereochemical interpretations easier and all structures in the Protein Data Bank (PDB; Berman *et al.*, 2000) are represented as individual molecules. However, so

far there are no standard rules for the selection of an asymmetric unit, and very often similar (isomorphous) structures in the PDB are presented in various, inconsistent ways (Dauter, 2013). This inconsistency may confuse those who are less fluent in symmetry transformations of crystallographic space groups. It would therefore be beneficial to introduce clear guidelines suggesting how to locate unique molecules in the unit cell in a uniform and standard way. The purpose of this paper is to formulate such guidelines and to propose standards for presenting macromolecular structures in a unified way.

The variability in the placement of molecules results from the fact that at early stages of structure solution, either starting from a molecular-replacement search or from the location of heavy or anomalous atoms, various programs place the solved models in different ways (even if equivalent) in relation to the origin of the unit cell. The permissible locations of a search model are governed by the so-called Cheshire symmetry, first introduced by Hirshfeld (1968) and later elaborated as the symmetry of Euclidean normalizers of the space groups (Fischer & Koch, 1983; Koch & Fischer, 2006; see Part 15 of Volume *A* of *International Tables for Crystallography*, 2005).

2. Cheshire symmetry

There are usually several locations where the cell origin can be fixed, all characterized by the same relation to the symmetry elements of the space group. For example, in all *P*-, *C*- and *I*-centered orthorhombic groups the fractional coordinates of all atoms can be shifted by $\pm 1/2$ in any axial direction, leading to entirely equivalent descriptions of the structure. Such shifts do not change the amplitudes of the structure factors, although the phases are changed depending on the parity group of their indices. It is therefore completely arbitrary in relation to which of the permitted origins the structure is described. As a consequence, the placement of the initial atomic model may be restricted only to the appropriate part of the whole unit cell. In addition, in some space groups, such as $P2_1$ or $P1$, the origin is 'floating' in one or more directions and thus could be fixed arbitrarily.

Analysis of the unique parts of the unit cell sufficient to place the search atom or molecule led Hirshfeld (1968) to the concept of a related cell, usually smaller than the original cell, possessing the space-group symmetry corresponding to the graph of symmetry elements of the original parent space group but without any atoms. In other words, this would be the symmetry of the graph presented in *International Tables for Crystallography* Volume *A* for a given space group. Hirshfeld used the term 'Cheshire cell', in analogy to the famous Cheshire cat from *Alice in Wonderland* (Carroll, 1865), who was able to disappear (like the atoms), leaving only its grin (the symmetry elements). From the mathematical point of view, the Cheshire symmetry corresponds to the Euclidean normalizer of a particular space group (Fischer & Koch, 1983).

The Cheshire cells and symmetry for all space groups are tabulated in Part 15 of Volume *A* of *International Tables for Crystallography* (2005). They have some characteristic properties. For those space groups where the origin is 'floating' and can be fixed anywhere along one or more axes, as in polar groups (classes 2, 3, 4, 6 and *m*) or in *P*1, the corresponding parameters of the Cheshire cell are reduced to infinitesimally small dimensions and, following the nomenclature of Hirshfeld, the (formally primitive) Cheshire cell acquires the Z^1, Z^2 or Z^3 type. The majority of the Cheshire groups are centrosymmetric, except for space groups containing only one hand of a chiral screw (3_1, 3_2, 4_1, 4_3, 6_1, 6_2, 6_4 and 6_5), since the symmetry graphs of symmorphic space groups or those containing 2_1, 4_2 or 6_3 axes or glide planes are centrosymmetric, and also when screw axes with opposite chirality are present simultaneously ($I4_1xx$ or $F4_132$). Most Cheshire groups are primitive, of the *P* or Z^n type, but some are *I*-centered or *R*-centered. The volume of the Cheshire cell is usually smaller than of the parent unit cell, except for certain cubic symmetries. Fig. 2 illustrates the relation between the space group $P4_12_12$ and its corresponding Cheshire cell and symmetry.

Hirshfeld concluded that

> the asymmetric unit of the Cheshire cell is the region to be scanned by the rotation and translation parameters.

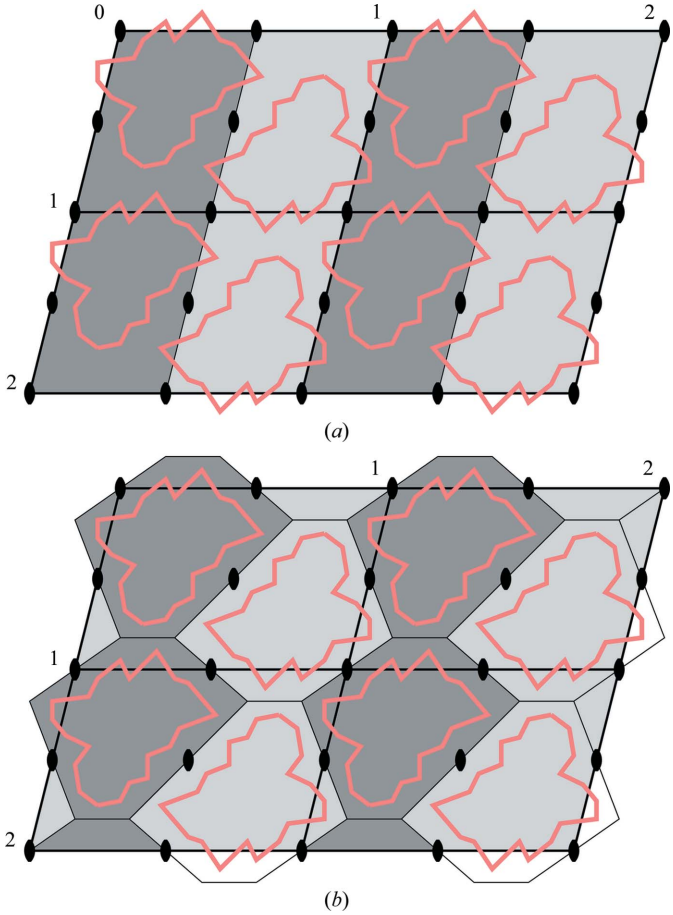

Figure 1
Molecules (represented as an irregular pink contour) in four unit cells of a structure in space group *P*2 shown in a projection along *b*. (*a*) The standard asymmetric units, as defined in *International Tables for Crystallography* ($0 \le x < 1$; $0 \le y < 1$; $0 \le x \le 1/2$), are marked as parallelograms (*y*-projected parallelepipeds). (*b*) An alternative choice of the asymmetric unit, nonstandard but encompassing the whole molecule.

This formulation can be elaborated in more detail. Hirshfeld (being a small-molecule, not a macromolecular, crystallographer) neglected the problem of chirality, as shown by the following passage from his paper:

> enantiomorphic structures are, in fact, hardly ever distinguishable in the trial-structure stage of a crystallographic study.

This is true for the location of heavy atoms by Patterson or direct methods, but is not applicable to molecular replacement (MR). In the interpretation of a Patterson synthesis, two inversion-related locations of the first heavy atom are equally plausible and, if the structure is noncentrosymmetric, its chirality has to be selected from the measured anomalous signal or from the *a priori* known chirality of the investigated chemical molecule, as is applicable for proteins and nucleic acids. However, in the macromolecular MR approach, the search model has fixed (*a priori* known) chirality and the eventuality of inverted handedness of the model does not need to be taken into account.

The location of the first single atom in the interpretation of a Patterson synthesis or the constellation of heavy or anomalous substructure obtained from direct methods may be restricted to one of the asymmetric units of the appropriate Cheshire cell. As pointed out above, when the Cheshire cell is centrosymmetric, selecting the first atom in one half of its asymmetric units leads to one chirality and selecting the first atom in the other half leads to the opposite chirality of the final structure. However, if the trial model to be located is chiral, the rules of its positioning are different.

3. Anti-Cheshire symmetry

A molecule can be equally well presented in any one of the orientations resulting from each of the rotational symmetry elements of the space group. Hirshfeld also assumed that it may be represented in either of the two possible chiralities. In conclusion, each orientation in each of the two chiral versions of the model may be placed in one of the asymmetric units of the Cheshire cell.

However, proteins and nucleotides are chiral and crystallize in one of the 65 space groups with only proper (*i.e.* rotational) symmetry elements (except when deliberately synthesized with inverted handedness to obtain racemic crystals; Yeates & Kent, 2012). Their location is therefore not permitted in an arbitrarily selected asymmetric unit of the Cheshire cell, since half of these units would require the opposite chirality of the search model.

Moreover, most of the MR programs first solve the rotation function, leading to one of the possible orientations of the search molecule, and only then perform a translation search for such a properly oriented molecule. At this stage it is known that the oriented molecule can be placed in one of the asymmetric units of the Cheshire cell, but it is not known in which. As a consequence, the MR translation search must encompass the entire Cheshire cell. When the Cheshire cell is not primitive, but is *I*- or *R*-centered, it is obviously sufficient to search only the appropriate fraction (1/2 or 1/3) of the Cheshire cell; in other words, only the 'primitive' part of the Cheshire cell.

When the structure is already known, it is possible to present it in one of the possible orientations and chiralities in a selected asymmetric unit of the Cheshire cell. When the

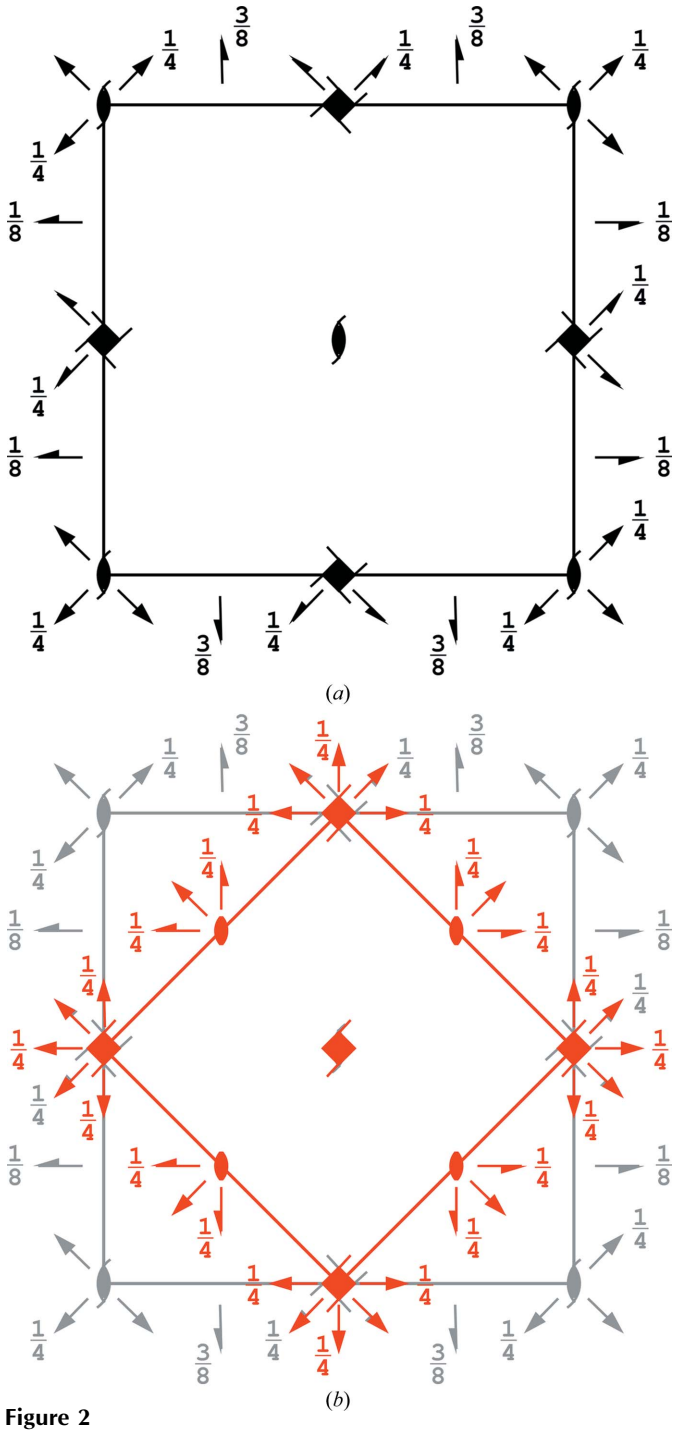

Figure 2
(*a*) Diagram of symmetry elements in space group $P4_12_12$. (*b*) The same diagram with the corresponding Cheshire cell and symmetry marked. Note that the square base of the unit cell is smaller and is rotated 45°. Since the *c* dimension of the Cheshire cell is halved, the relative level of the horizontal twofold axes is 1/4 (or 3/4) instead of 1/8 and 3/8 (or 5/8 and 7/8) and the translation along the fourfold screw axis becomes 1/2 of the *c* parameter, so that the Cheshire symmetry is $P4_222$.

chirality is specified, as in the case of macromolecules, the selected region for each differently oriented molecule is two times larger than a single asymmetric unit of the Cheshire cell. It corresponds to the asymmetric unit of a cell equal to the Cheshire cell but with symmetry including only rotations, since rotoinversions have to be excluded.

The resulting symmetry can be obtained by a superposition of all structures resulting from translation of the unit-cell contents according to all permitted locations of the origin in the particular space group. This is exactly opposite to the removal of atoms in the derivation of the Cheshire symmetry, and the result of this opposite procedure can therefore be termed the 'anti-Cheshire' cell and symmetry. In addition, if the crystal symmetry is lower than the lattice symmetry (as in classes 3, 4, 6, 32 and 23) the symmetry elements can be accommodated in the lattice in more than one way, and this additionally increases the anti-Cheshire symmetry. In such cases, placement of molecules in the selected asymmetric unit may require re-indexing of the diffraction data, as specified in Table 1. The anti-Cheshire symmetry groups are equivalent to the 'chirality-preserving Euclidean normalizers' tabulated by the Bilbao Crystallographic Server (Aroyo *et al.*, 2006; http://www.cryst.ehu.es/cryst/get_nor.html).

Table 1
Space groups for which placement of the structure in the optimal asymmetric unit of the anti-Cheshire cell may require re-indexing of the data.

Space groups	Re-indexing from h, k, l
$P4$, $P4_1$, $P4_2$, $P4_3$, $I4$, $I4_1$	$k, h, -l$
$P3$, $P3_1$, $P3_2$	$-h, -k, l$; $k, h, -l$; $-k, -h, -l$
$R3$	$k, h, -l$
$P312$, $P3_112$, $P3_212$	$-h, -k, l$
$P321$, $P3_121$, $P3_221$	$-h, -k, l$
$P6$, $P6_1$, $P6_2$, $P6_3$, $P6_4$, $P6_5$	$k, h, -l$
$P23$, $P2_13$, $I23$, $I2_13$, $F23$	$k, h, -l$

The analogous procedure of overlapping all structures resulting from all permissible shifts of the cell origin applied to space groups possessing not only pure rotations will reproduce the Cheshire groups. For space groups with symmetry elements of the second kind, the concepts of Cheshire and anti-Cheshire symmetries coincide.

The asymmetric unit of the anti-Cheshire cell is smaller than that of the original space group (except for $P1$). The whole model cannot therefore be completely accommodated inside

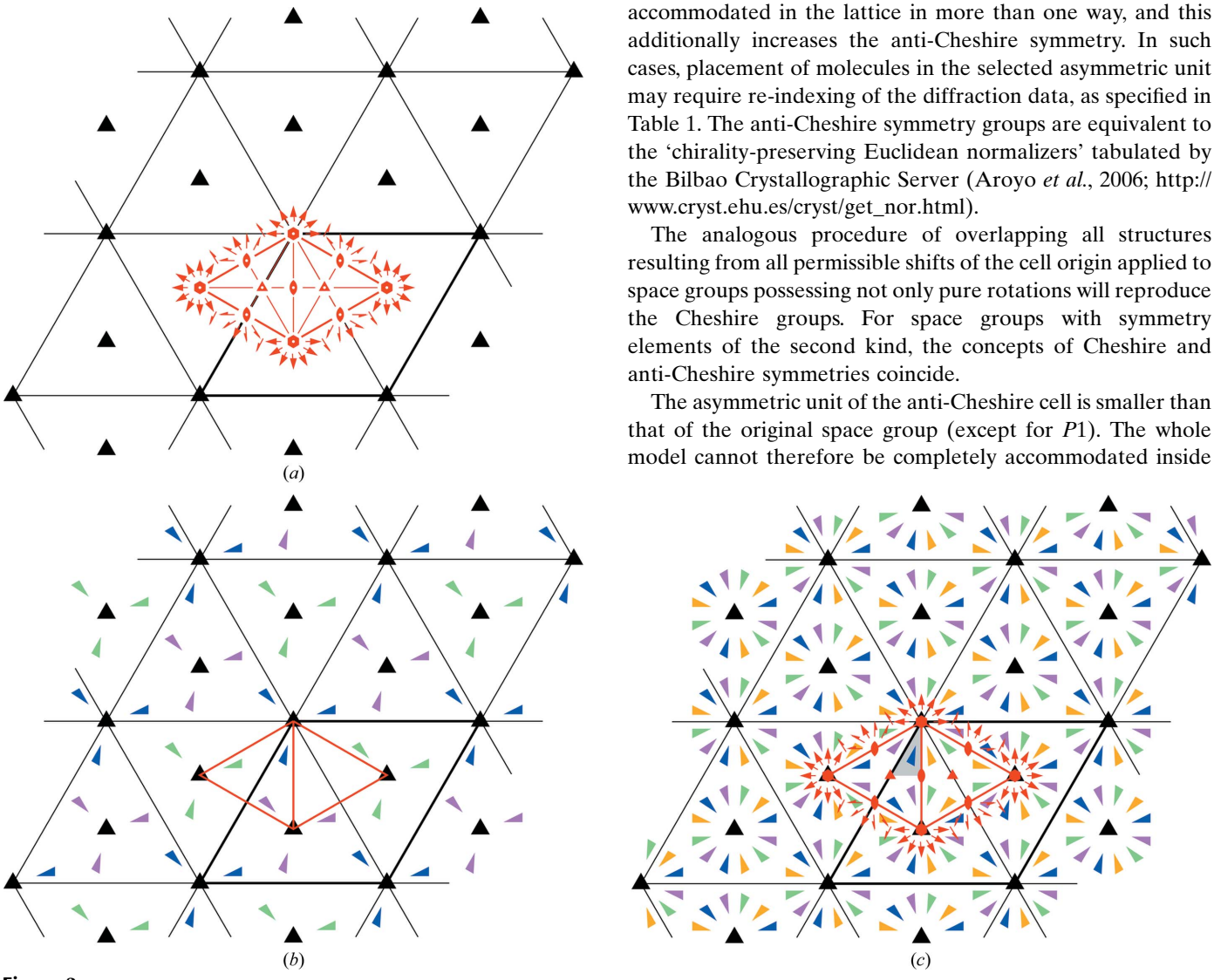

Figure 3
(*a*) Symmetry diagram of space group $P3$. The Cheshire cell and symmetry (indicated in red) is $Z^1 6/mmm$, with the c cell dimension reduced to the infinitesimally small value ε (note that in $P3$ the origin is not fixed along z), shortened a and b cell parameters and the directions of the x and y axes rotated appropriately. (*b*) The same symmetry diagram with positions of chiral molecules (represented by triangles) resulting from the superposition of the whole structure translated relative to the two additional possible cell origins at 1/3, 2/3, 0 and 2/3, 1/3, 0. Each translated constellation of molecules is shown in a different color. (*c*) A diagram with three additional sets of molecules (each in a different color) resulting from the permitted re-indexing of the data according to the three different twofold axes present in the metric (lattice) symmetry but not in the parent space-group symmetry. The resulting anti-Cheshire cell is marked in red and its symmetry is $Z^1 622$. The asymmetric unit of the anti-Cheshire cell is shaded in gray. When expressed in terms of the coordinates in the parent unit cell, its limits are $0 \le x \le 1/3$; $0 \le y \le x/2$; $z = 0$. It is always possible to represent the original structure by one molecule (or rather a unique structural motif) in the appropriate orientation with its center located in this asymmetric unit.

Table 2
Space groups possessing only rotational symmetry elements with the corresponding Cheshire symmetry and cell, limits of the translational molecular-replacement search of the oriented molecule, anti-Cheshire symmetry and limits for the optimal positioning of a molecule in the appropriate orientation.

The MR search limits cover the entire primitive (anti-)Cheshire cell or an appropriate fraction of the *I*- or *R*-centered cell. The last column specifies the limits of the positioning of the molecular center of gravity; this region corresponds to the asymmetric unit of the anti-Cheshire cell. x, y and z are fractional coordinates in the original cell. The symbol ε relates to an infinitesimally small cell dimension in a particular direction.

Space group						
No.	Symbol	Cheshire symmetry	Cheshire cell	MR search limits	Anti-Cheshire symmetry	Molecule location limits
1	*P*1	$Z^3\bar{1}$	εa, εb, εc	$x, y, z = 0$	$Z^3 1$	$x, y, z = 0$
3	*P*2	$Z^1 2/m$	$a/2$, εb, $c/2$	$0 \le x, z < 1/2$; $y = 0$	$Z^1 2$	$0 \le x < 1/4$; $0 \le z < 1/2$; $y = 0$
4	$P2_1$	$Z^1 2/m$	$a/2$, εb, $c/2$	$0 \le x, z < 1/2$; $y = 0$	$Z^1 2$	$0 \le x < 1/4$; $0 \le z < 1/2$; $y = 0$
5	*C*2	$Z^1 2/m$	$a/2$, εb, $c/2$	$0 \le x, z < 1/2$; $y = 0$	$Z^1 2$	$0 \le x < 1/4$; $0 \le z < 1/2$; $y = 0$
16	*P*222	*Pmmm*	$a/2$, $b/2$, $c/2$	$0 \le x, y, z < 1/2$	*P*222	$0 \le x, y < 1/4$; $0 \le z < 1/2$
17	$P222_1$	*Pmmm*	$a/2$, $b/2$, $c/2$	$0 \le x, y, z < 1/2$	*P*222	$0 \le x, y < 1/4$; $0 \le z < 1/2$
18	$P2_12_12$	*Pmmm*	$a/2$, $b/2$, $c/2$	$0 \le x, y, z < 1/2$	*P*222	$0 \le x, y < 1/4$; $0 \le z < 1/2$
19	$P2_12_12_1$	*Pmmm*	$a/2$, $b/2$, $c/2$	$0 \le x, y, z < 1/2$	*P*222	$0 \le x, y < 1/4$; $0 \le z < 1/2$
20	$C222_1$	*Pmmm*	$a/2$, $b/2$, $c/2$	$0 \le x, y, z < 1/2$	*P*222	$0 \le x, y < 1/4$; $0 \le z < 1/2$
21	*C*222	*Pmmm*	$a/2$, $b/2$, $c/2$	$0 \le x, y, z < 1/2$	*P*222	$0 \le x, y < 1/4$; $0 \le z < 1/2$
22	*F*222	*Immm*	$a/2$, $b/2$, $c/2$	$0 \le x, y < 1/2$; $0 \le z < 1/4$	*P*222	$0 \le x, y, z < 1/4$
23	*I*222	*Pmnm*	$a/2$, $b/2$, $c/2$	$0 \le x, y, z < 1/2$	*P*222	$0 \le x, y < 1/4$; $0 \le z < 1/2$
24	$I2_12_12_1$	*Pmnm*	$a/2$, $b/2$, $c/2$	$0 \le x, y, z < 1/2$	*P*222	$0 \le x, y < 1/4$; $0 \le z < 1/2$
75	*P*4	$Z^1 4/mmm$	$(a-b)/2$, $(a+b)/2$, εc	$0 \le x < 1/2$; $0 \le y < 1$; $z = 0$	$Z^1 422$	$0 \le x \le 1/4$; $0 \le y \le 1/2 - x$; $z = 0$
76	$P4_1$	$Z^1 422$	$(a-b)/2$, $(a+b)/2$, εc	$0 \le x < 1/2$; $0 \le y < 1$; $z = 0$	$Z^1 422$	$0 \le x \le 1/4$; $0 \le y \le 1/2 - x$; $z = 0$
77	$P4_2$	$Z^1 4/mmm$	$(a-b)/2$, $(a+b)/2$, εc	$0 \le x < 1/2$; $0 \le y < 1$; $z = 0$	$Z^1 422$	$0 \le x \le 1/4$; $0 \le y \le 1/2 - x$; $z = 0$
78	$P4_3$	$Z^1 422$	$(a-b)/2$, $(a+b)/2$, εc	$0 \le x < 1/2$; $0 \le y < 1$; $z = 0$	$Z^1 422$	$0 \le x \le 1/4$; $0 \le y \le 1/2 - x$; $z = 0$
79	*I*4	$Z^1 4/mmm$	$(a-b)/2$, $(a+b)/2$, εc	$0 \le x < 1/2$; $0 \le y < 1$; $z = 0$	$Z^1 422$	$0 \le x \le 1/4$; $0 \le y \le 1/2 - x$; $z = 0$
80	$I4_1$	$Z^1 4/nbm$	$(a-b)/2$, $(a+b)/2$, εc	$0 \le x < 1/2$; $0 \le y < 1$; $z = 0$	$Z^1 422$	$0 \le x \le 1/4$; $0 \le y \le 1/2 - x$; $z = 0$
89	*P*422	*P*4/*mmm*	$(a-b)/2$, $(a+b)/2$, $c/2$	$0 \le x < 1/2$; $0 \le y < 1$; $0 \le z < 1/2$	*P*422	$0 \le x \le 1/4$; $0 \le y \le 1/2 - x$; $0 \le z < 1/2$
90	$P42_12$	*P*4/*mmm*	$(a-b)/2$, $(a+b)/2$, $c/2$	$0 \le x < 1/2$; $0 \le y < 1$; $0 \le z < 1/2$	*P*422	$0 \le x \le 1/4$; $0 \le y \le 1/2 - x$; $0 \le z < 1/2$
91	$P4_122$	$P4_222$	$(a-b)/2$, $(a+b)/2$, $c/2$	$0 \le x < 1/2$; $0 \le y < 1$; $0 \le z < 1/2$	$P4_222$	$0 \le x \le 1/4$; $0 \le y \le 1/2 - x$; $0 \le z < 1/2$
92	$P4_12_12$	$P4_222$	$(a-b)/2$, $(a+b)/2$, $c/2$	$0 \le x < 1/2$; $0 \le y < 1$; $0 \le z < 1/2$	$P4_222$	$0 \le x \le 1/4$; $0 \le y \le 1/2 - x$; $0 \le z < 1/2$
93	$P4_222$	*P*4/*mmm*	$(a-b)/2$, $(a+b)/2$, $c/2$	$0 \le x < 1/2$; $0 \le y < 1$; $0 \le z < 1/2$	*P*422	$0 \le x \le 1/4$; $0 \le y \le 1/2 - x$; $0 \le z < 1/2$
94	$P4_22_12$	*P*4/*mmm*	$(a-b)/2$, $(a+b)/2$, $c/2$	$0 \le x < 1/2$; $0 \le y < 1$; $0 \le z < 1/2$	*P*422	$0 \le x \le 1/4$; $0 \le y \le 1/2 - x$; $0 \le z < 1/2$
95	$P4_322$	$P4_222$	$(a-b)/2$, $(a+b)/2$, $c/2$	$0 \le x < 1/2$; $0 \le y < 1$; $0 \le z < 1/2$	$P4_222$	$0 \le x \le 1/4$; $0 \le y \le 1/2 - x$; $0 \le z < 1/2$
96	$P4_32_12$	$P4_222$	$(a-b)/2$, $(a+b)/2$, $c/2$	$0 \le x < 1/2$; $0 \le y < 1$; $0 \le z < 1/2$	$P4_222$	$0 \le x \le 1/4$; $0 \le y \le 1/2 - x$; $0 \le z < 1/2$
97	*I*422	*P*4/*mmm*	$(a-b)/2$, $(a+b)/2$, $c/2$	$0 \le x < 1/2$; $0 \le y < 1$; $0 \le z < 1/2$	*P*422	$0 \le x \le 1/2$; $0 \le y < 1/2$; $x + y \le 1/2$; $0 \le z < 1/4$
98	$I4_122$	$P4_2/nnm$	$(a-b)/2$, $(a+b)/2$, $c/2$	$0 \le x < 1/2$; $0 \le y < 1$; $0 \le z < 1/2$	$P4_222$	$0 \le x \le 1/2$; $0 \le y < 1/2$; $x + y \le 1/2$; $0 \le z < 1/4$
143	*P*3	$Z^1 6/mmm$	$(2a+b)/3$, $(-a+b)/3$, εc	$0 \le x + y < 1$; $0 \le -x + 2y < 1$; $z = 0$	$Z^1 62$	$0 \le x \le 1/3$; $0 \le y \le x/2$; $z = 0$
144	$P3_1$	$Z^1 622$	$(2a+b)/3$, $(-a+b)/3$, εc	$0 \le x + y < 1$; $0 \le -x + 2y < 1$; $z = 0$	$Z^1 622$	$0 \le x \le 1/3$; $0 \le y \le x/2$; $z = 0$
145	$P3_2$	$Z^1 622$	$(2a+b)/3$, $(-a+b)/3$, εc	$0 \le x + y < 1$; $0 \le -x + 2y < 1$; $z = 0$	$Z^1 622$	$0 \le x \le 1/3$; $0 \le y \le x/2$; $z = 0$
146	*R*3	$Z^1\bar{3}1m$	$(2a+b)/3$, $(-a+b)/3$, εc	$0 \le x, y < 1/3$; $z = 0$	$Z^1 312$	$0 \le x \le 1/3$; $0 \le y \le x$; $z = 0$
149	*P*312	*P*6/*mmm*	$(2a+b)/3$, $(-a+b)/3$, $c/2$	$0 \le x + y < 1$; $0 \le -x + 2y < 1$; $0 \le z < 1/2$	*P*622	$0 \le x \le 1/3$; $0 \le y \le x/2$; $0 \le z < 1/2$
150	*P*321	*P*6/*mmm*	$a, b, c/2$	$0 \le x, y < 1$; $0 \le z < 1/2$	*P*622	$0 \le x \le 2/3$; $0 \le y \le x/2$; $y \ge 2x - 1$; $0 \le z < 1/2$
151	$P3_112$	$P6_222$	$(2a+b)/3$, $(-a+b)/3$, $c/2$	$0 \le x + y < 1$; $0 \le -x + 2y < 1$; $0 \le z < 1/2$	$P6_222$	$0 \le x \le 1/3$; $0 \le y \le x/2$; $0 \le z < 1/2$
152	$P3_121$	$P6_222$	$a + b, -a, c/2$	$0 \le x, y < 1$; $0 \le z < 1/2$	$P6_222$	$0 \le x \le 2/3$; $0 \le y \le x/2$; $y \ge 2x - 1$; $0 \le z < 1/2$
153	$P3_212$	$P6_422$	$(2a+b)/3$, $(-a+b)/3$, $c/2$	$0 \le x + y < 1$; $0 \le -x + 2y < 1$; $0 \le z < 1/2$	$P6_422$	$0 \le x \le 1/3$; $0 \le y \le x/2$; $0 \le z < 1/2$
154	$P3_221$	$P6_422$	$a + b, -a, c/2$	$0 \le x, y < 1$; $0 \le z < 1/2$	$P6_422$	$0 \le x \le 2/3$; $0 \le y \le x/2$; $y \ge 2x - 1$; $0 \le z < 1/2$
155	*R*32	$R\bar{3}m$	$-a, -b, c/2$	$0 \le x, y < 1$; $0 \le z < 1/6$	*R*32	$0 \le x \le 1/3$; $0 \le y < 1/3$; $0 \le z \le 1/4$
168	*P*6	$Z^1 6/mmm$	$a, b, \varepsilon c$	$0 \le x, y < 1$; $z = 0$	$Z^1 622$	$0 \le x \le 2/3$; $0 \le y \le x/2$; $y \ge 2x - 1$; $z = 0$
169	$P6_1$	$Z^1 622$	$a, b, \varepsilon c$	$0 \le x, y < 1$; $z = 0$	$Z^1 622$	$0 \le x \le 2/3$; $0 \le y \le x/2$; $y \ge 2x - 1$; $z = 0$
170	$P6_5$	$Z^1 622$	$a, b, \varepsilon c$	$0 \le x, y < 1$; $z = 0$	$Z^1 622$	$0 \le x \le 2/3$; $0 \le y \le x/2$; $y \ge 2x - 1$; $z = 0$
171	$P6_2$	$Z^1 622$	$a, b, \varepsilon c$	$0 \le x, y < 1$; $z = 0$	$Z^1 622$	$0 \le x \le 2/3$; $0 \le y \le x/2$; $y \ge 2x - 1$; $z = 0$
172	$P6_4$	$Z^1 622$	$a, b, \varepsilon c$	$0 \le x, y < 1$; $z = 0$	$Z^1 622$	$0 \le x \le 2/3$; $0 \le y \le x/2$; $y \ge 2x - 1$; $z = 0$
173	$P6_3$	$Z^1 6/mmm$	$a, b, \varepsilon c$	$0 \le x, y < 1$; $z = 0$	$Z^1 622$	$0 \le x \le 2/3$; $0 \le y \le x/2$; $y \ge 2x - 1$; $z = 0$
177	*P*622	*P*6/*mmm*	$a, b, c/2$	$0 \le x, y < 1$; $0 \le z < 1/2$	*P*622	$0 \le x \le 2/3$; $0 \le y \le x/2$; $y \ge 2x - 1$; $0 \le z < 1/2$
178	$P6_122$	$P6_222$	$a, b, c/2$	$0 \le x, y < 1$; $0 \le z < 1/2$	$P6_222$	$0 \le x \le 2/3$; $0 \le y \le x/2$; $y \ge 2x - 1$; $0 \le z < 1/2$
179	$P6_522$	$P6_422$	$a, b, c/2$	$0 \le x, y < 1$; $0 \le z < 1/2$	$P6_422$	$0 \le x \le 2/3$; $0 \le y \le x/2$; $y \ge 2x - 1$; $0 \le z < 1/2$
180	$P6_222$	$P6_422$	$a, b, c/2$	$0 \le x, y < 1$; $0 \le z < 1/2$	$P6_422$	$0 \le x \le 2/3$; $0 \le y \le x/2$; $y \ge 2x - 1$; $0 \le z < 1/2$
181	$P6_422$	$P6_222$	$a, b, c/2$	$0 \le x, y < 1$; $0 \le z < 1/2$	$P6_222$	$0 \le x \le 2/3$; $0 \le y \le x/2$; $y \ge 2x - 1$; $0 \le z < 1/2$
182	$P6_322$	*P*6/*mmm*	$a, b, c/2$	$0 \le x, y < 1$; $0 \le z < 1/2$	*P*622	$0 \le x \le 2/3$; $0 \le y \le x/2$; $y \ge 2x - 1$; $0 \le z < 1/2$
195	*P*23	*Im*3*m*	a, b, c	$0 \le x, y < 1$; $0 \le z < 1/2$	*I*432	$0 \le x \le 1/4$; $x \le y$, $z \le 1/2 - x$
196	*F*23	*Im*3*m*	$a/2$, $b/2$, $c/2$	$0 \le x, y < 1/2$; $0 \le z < 1/4$	*I*432	$0 \le x \le 1/8$; $x \le y$, $z \le 1/4 - x$
197	*I*23	*Im*3*m*	a, b, c	$0 \le x, y < 1$; $0 \le z < 1/2$	*I*432	$0 \le x \le 1/4$; $x \le y$, $z \le 1/2 - x$
198	$P2_13$	*Ia*3*d*	a, b, c	$0 \le x, y < 1$; $0 \le z < 1/2$	$I4_132$	$-3/8 < x < 1/8$; $-1/8 < y < 1/8$; $\max(x, y, y - x - 1/8) < z < y + 1/4$
199	$I2_13$	*Ia*3*d*	a, b, c	$0 \le x, y < 1$; $0 \le z < 1/2$	$I4_132$	$-3/8 < x < 1/8$; $-1/8 < y < 1/8$; $\max(x, y, y - x - 1/8) < z < y + 1/4$
207	*P*432	*Im*3*m*	a, b, c	$0 \le x, y < 1$; $0 \le z < 1/2$	*I*432	$0 \le x \le 1/4$; $x \le y$, $z \le 1/2 - x$
208	$P4_232$	*Im*3*m*	a, b, c	$0 \le x, y < 1$; $0 \le z < 1/2$	*I*432	$0 \le x \le 1/4$; $x \le y$, $z \le 1/2 - x$

Table 2 (continued)

Space group No.	Symbol	Cheshire symmetry	Cheshire cell	MR search limits	Anti-Cheshire symmetry	Molecule location limits
209	$F432$	$Im\bar{3}m$	$a/2, b/2, c/2$	$0 \le x, y < 1/2; 0 \le z < 1/4$	$I432$	$0 \le x \le 1/8; x \le y, z \le 1/4 - x$
210	$F4_132$	$Pn\bar{3}m$	$a/2, b/2, c/2$	$0 \le x, y < 1/2; 0 \le z < 1/4$	$I432$	$0 \le x \le 1/8; x \le y, z \le 1/4 - x$
211	$I432$	$Im\bar{3}m$	a, b, c	$0 \le x, y < 1; 0 \le z < 1/2$	$I432$	$0 \le x \le 1/4; x \le y, z \le 1/2 - x$
212	$P4_332$	$I4_132$	a, b, c	$0 \le x, y < 1; 0 \le z < 1/2$	$I4_132$	$-3/8 < x < 1/8; -1/8 < y < 1/8;$ $\max(x, y, y - x - 1/8) < z < y + 1/4$
213	$P4_132$	$I4_132$	a, b, c	$0 \le x, y < 1; 0 \le z < 1/2$	$I4_132$	$-3/8 < x < 1/8; -1/8 < y < 1/8;$ $\max(x, y, y - x - 1/8) < z < y + 1/4$
214	$I4_132$	$Ia\bar{3}d$	a, b, c	$0 \le x, y < 1; 0 \le z < 1/22$	$I4_132$	$-3/8 < x < 1/8; -1/8 < y < 1/8;$ $\max(x, y, y - x - 1/8) < z < y + 1/4$

the asymmetric unit of the anti-Cheshire cell, but the center of gravity of the model can always be shifted to such a region. In fact, the volume of the asymmetric unit of the anti-Cheshire cell is equal to the fraction of the parent unit cell obtained by division by the number of general positions, the number of possible origin locations and the number of different indexing possibilities in the parent space group.

Fig. 3 illustrates space group $P3$ and the Cheshire and anti-Cheshire cells derived for this space group. Both of them are equal and their ab face is three times smaller than for the parent cell. Since the origin in $P3$ is floating along the threefold axis, the c cell dimension of both the Cheshire and anti-Cheshire cells is infinitesimally small. The asymmetric unit of the anti-Cheshire cell can always be selected to accommodate the center of the whole parent structural motif, as illustrated in Fig. 4 for one of the structures from the PDB.

4. Optimal location of molecules in the unit cell

The above property of anti-Cheshire cells suggests the possibility of introducing a standard convention for the presentation of structural models of molecular crystals. The molecule can always be transformed by one of the space-group

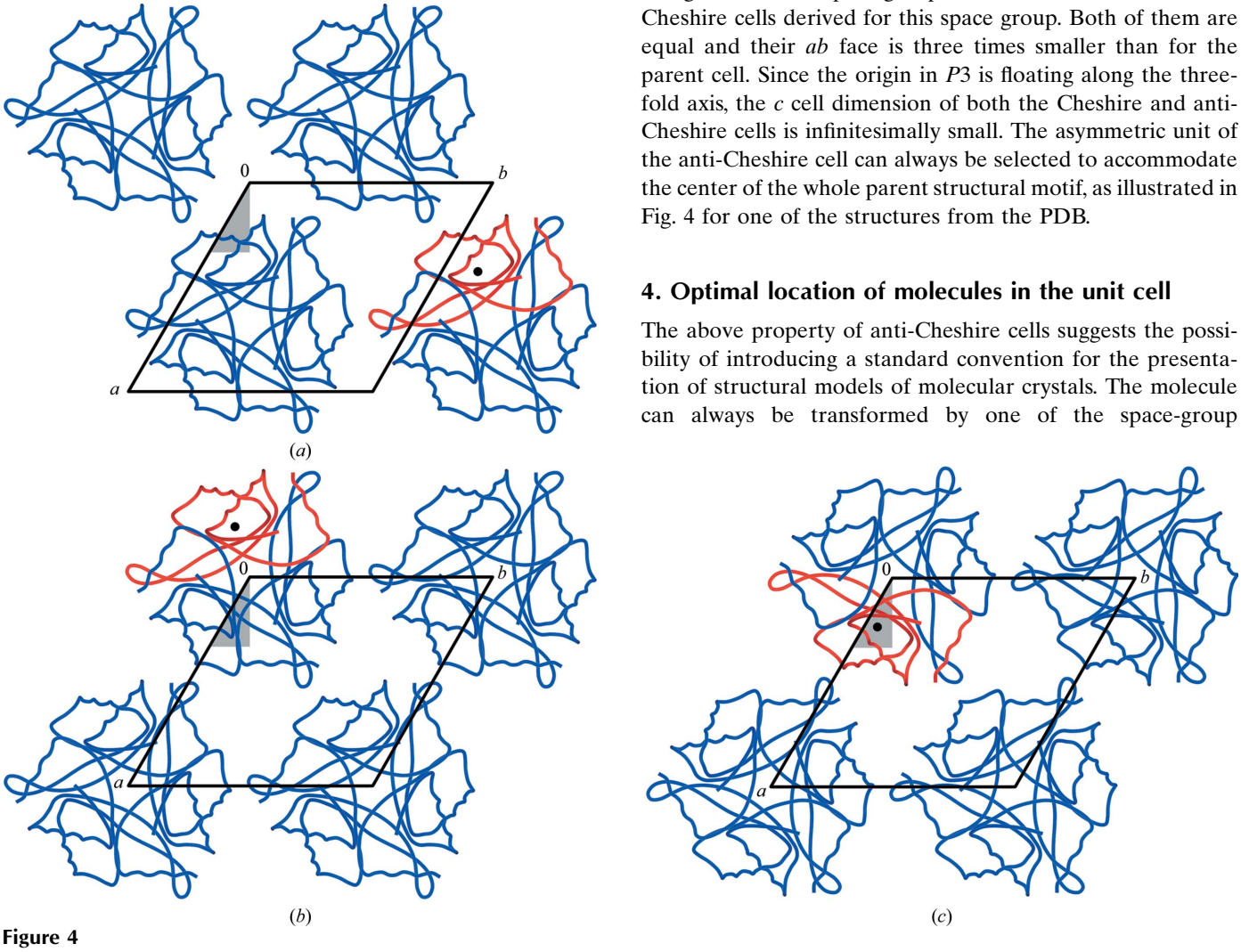

Figure 4
Packing of molecules in PDB entry 1o5j (Joint Center for Structural Genomics, unpublished work). Symmetric homotrimers are formed around the threefold axis of space group $P3$, with the unique protomer deposited in the PDB represented in red and all symmetry-equivalent protomers shown in blue. The center of the unique molecule is marked by a black dot and the selected asymmetric unit of the anti-Cheshire cell is shaded gray. (*a*) The original location of molecules corresponding to the PDB deposition. (*b*) An equivalent constellation shifted by $-2/3, -4/3, 0$ to an alternative cell origin. (*c*) Rotation of the whole structure by 180° around the twofold axis parallel to y, with coordinate transformation $-x, y - x, -z$, which also requires reindexing of the diffraction data according to $k, h, -l$, ensures that the center of the unique (red) protomer lies within the designated asymmetric unit of the anti-Cheshire cell.

operators and translated relative to the allowed cell origins to place its center within the selected asymmetric unit of the appropriate anti-Cheshire cell. Optimally, the selected asymmetric unit should be located close to the origin of the parent unit cell. The limits of such regions are specified in Table 2 for each of the proper-symmetry space groups relevant to macromolecular crystallography. If multiple symmetry-independent molecules are present in the structure, they should be first transformed to create the most compact oligomer or aggregate and then transformed as a single entity to the selected region of the anti-Cheshire cell. If the oligomer is generated by crystallographic symmetry, it can also be treated as one structural entity, but its center will be placed at a special position at the border of the anti-Cheshire asymmetric unit (note the 'inclusive' limits '$\leq$' in the definitions in Table 2). The unique protomer (subunit) representing such an oligomer should also lie within the selected asymmetric unit. If the anti-Cheshire cell has an infinitesimally small dimension, such as for example in $Z^{1}422$, the molecular center should be shifted, by convention, to the zero level along that direction.

It would benefit all users of the PDB if deposited crystal structures were presented according to unified standard rules such as those proposed above.

References

Aroyo, M. I., Perez-Mato, J. M., Capillas, C., Kroumova, E., Ivantchev, S., Madariaga, G., Kirov, A. & Wondratschek, H. (2006). *Z. Kristallogr.* **221**, 15–27.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.

Carroll, L. (1865). *Alice's Adventures in Wonderland*, ch. 6. London: Macmillan.

Dauter, Z. (2013). *Acta Cryst.* D**69**, 2–4.

Fischer, W. & Koch, E. (1983). *Acta Cryst.* A**39**, 907–915.

Hirshfeld, F. L. (1968). *Acta Cryst.* A**24**, 301–311.

International Tables for Crystallography (2005). *Vol. A*, edited by Th. Hahn. Heidelberg: Springer.

Koch, E. & Fischer, W. (2006). *Z. Kristallogr.* **221**, 1–14.

Richards, F. M. (1974). *J. Mol. Biol.* **82**, 1–14.

Voronoi, G. (1908). *J. Reine Angew. Math.* **134**, 198–287.

Yeates, T. O. & Kent, S. B. (2012). *Annu. Rev. Biophys.* **41**, 41–61.