electronic reprint

STRUCTURAL BIOLOGY

ISSN: 2059-7983
journals.iucr.org/d

Crystallographically correct but confusing presentation of structural models deposited in the Protein Data Bank

Zbigniew Dauter and Alexander Wlodawer

Acta Cryst. (2018). D**74**, 939–945

IUCr Journals
CRYSTALLOGRAPHY JOURNALS ONLINE

STRUCTURAL BIOLOGY
ISSN 2059-7983

Crystallographically correct but confusing presentation of structural models deposited in the Protein Data Bank

Zbigniew Dauter[a]* and Alexander Wlodawer[b]*

[a]Synchrotron Radiation Research Section, Macromolecular Crystallography Laboratory, NCI, Argonne National Laboratory, Argonne, IL 60439, USA, and [b]Protein Structure Section, Macromolecular Crystallography Laboratory, National Cancer Institute, Frederick, MD 21702, USA. *Correspondence e-mail: dauter@anl.gov, wlodawer@nih.gov

Received 14 March 2018
Accepted 9 July 2018

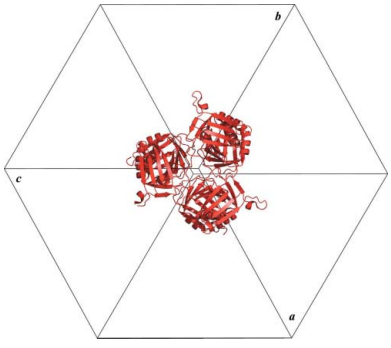
Edited by S. Wakatsuki, Stanford University, USA

Keywords: Protein Data Bank; unit cell; placement of coordinates; molecular symmetry.

The Protein Data Bank (PDB) constitutes a collection of the available atomic models of macromolecules and their complexes obtained by various methods used in structural biology, but chiefly by crystallography. It is an indispensable resource for all branches of science that deal with the structures of biologically active molecules, such as structural biology, bioinformatics, the design of novel drugs *etc.* Since not all users of the PDB are familiar with the methods of crystallography, it is important to present the results of crystallographic analyses in a form that is easy to interpret by nonspecialists. It is advisable during the submission of structures to the PDB to pay attention to the optimal placement of molecules within the crystal unit cell, to the correct representation of oligomeric assemblies and to the proper selection of the space-group symmetry. Examples of significant departures from these principles illustrate the potential for the misinterpretation of such suboptimally presented crystal structures.

1. Introduction

Since crystal structures are built from periodically repeating, identically arranged objects filling the unit cells, it is irrelevant from the strictly crystallographic point of view whether the whole structure is presented with various fragments dispersed in different unit cells, or even whether the structure is expressed with a symmetry lower than that for its true space group. However, for the proper interpretation of many chemical and biological phenomena, such as various intermolecular interactions or the formation of biologically relevant assemblies, the whole content of the crystal unit cell should be presented in as compact as possible arrangement of individual molecules, preferentially in their proper oligomeric state. Such an approach makes it easier for people who are less experienced in the application of crystallographic symmetry transformations to properly interpret the various biological and chemical properties and characteristics of the whole structure. After all, the Protein Data Bank (PDB; Burley *et al.*, 2018), as the principal repository of biologically relevant structures of macromolecules, serves not just crystallographers but chiefly researchers interested in the biological and biomedical aspects of the structures of macromolecules and their complexes.



This fact is not always taken into account by authors submitting their results to the PDB. A large number of structures are presented in a way that, despite being correct crystallographically, is highly confusing for noncrystallographers. Here, we include some examples of such structures.

electronic reprint

Table 1
Selected structures placed away from the unit cell.

(*a*) Structures with a single independent molecule placed more than five unit-cell lengths from the origin. n_a, n_b and n_c are the translations along each crystal axis that are needed to bring the center of mass of the molecule to reside within the unit cell.

PDB code	n_a	n_b	n_c	Space group
4erl	0	0	21	$P3_2$
3ozq	0	−12	0	$P2_1$
4gso	0	0	12	$I4$
4kek	0	−1	−8	$I4_1$
5t8u	4	6	2	$P3_221$
5f5e	3	6	2	$P3_221$
1yt7	0	1	6	$P3_2$
3ra6	0	−6	0	$P2_1$
5cu4	0	−4	6	$P2_1$
5cu6	0	−5	5	$P2_1$
5moh	0	−5	5	$P2_1$
5mov	0	−5	5	$P2_1$
5eqg	5	−1	4	$C2$
4pyp	5	−1	3	$C2$
5i9v	5	0	2	$P2_1$
5cdj	−5	0	1	$P1$

(*b*) All of the following 25 structures are described in a single paper (Beylkin *et al.*, 2017).

PDB code	n_a	n_b	n_c	Space group
5wdw	0	4	5	$I422$
5wap	−1	3	5	$I422$
5wa7	5	2	4	$I422$
5wdc	5	2	4	$I422$
5wg9	5	2	4	$I422$
5web	2	−4	5	$I422$
5wfm	2	−4	5	$I422$
5wfw	4	1	4	$I422$
5w92	1	1	3	$I422$
5wfz	1	1	3	$I422$
5wei	1	2	2	$I422$
5wf3	1	2	2	$I422$
5wb3	1	2	2	$I422$
5w73	1	1	2	$I422$
5w7u	1	1	2	$I422$
5w9g	−1	1	2	$I422$
5wa6 (HPUB)	?	?	?	$I422$
5wdn	−1	0	0	$P3_221$
5w3i	0	−1	0	$P3_221$
5we7	0	−1	0	$P3_221$
5wcs	1	1	0	$P6_422$
5we9	0	1	0	$P6_422$
5w44	0	1	0	$P6_422$
5wef	0	1	0	$P6_422$
5wct	0	0	0	$P6_422$

Since checking whether the atomic models of the submitted structures are properly placed in the unit cell is not part of the validation process, the job of assuring the most optimal presentations falls squarely on the authors.

2. Materials and methods

For the purposes of this analysis, we utilized the contents of the Protein Data Bank as of 13 January 2018. Our analysis of the placement of molecules at a distance from the unit cell was limited to structures with only a single molecule in the asymmetric unit, as judged from the information in the CRYST1 record of the PDB file header. Apart from the cell dimensions and the space group, this record also provides the number of molecules (or, rather, unique structural motifs) in the whole cell. The geometric center of all atoms within a molecule was computed and transformed to fractional coordinates, and its translational component was evaluated. The cases of molecules placed furthest from the unit cell are presented in Table 1. An analogous procedure was applied to all 500 tetragonal crystal structures of hen egg-white lysozyme and the results are included in Table 2.

Table 2
The approximate fractional coordinates of molecular centers and the number of such instances in 500 tetragonal crystal structures of HEWL (space group $P4_32_12$) deposited in the PDB.

x	y	z	No.
0	$\frac{1}{4}$	$\frac{1}{2}$	182
$-\frac{1}{4}$	0	$-\frac{1}{4}$	83
0	$-\frac{1}{4}$	0	70
$\frac{1}{4}$	0	0	30
$\frac{1}{4}$	0	$\frac{1}{4}$	19
0	$\frac{3}{4}$	0	19
$\frac{1}{4}$	0	$\frac{1}{2}$	18
$-\frac{1}{4}$	0	0	17
0	$\frac{1}{4}$	0	14
0	$\frac{1}{4}$	$\frac{1}{4}$	14
0	$-\frac{1}{4}$	$\frac{1}{4}$	12
1	$\frac{3}{4}$	1	5
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	2
0	$\frac{1}{4}$	$1\frac{1}{2}$	1
$-\frac{1}{4}$	0	$\frac{1}{4}$	1
0	$\frac{1}{4}$	2	1
0	$1\frac{1}{4}$	0	1
$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
$\frac{1}{4}$	$\frac{1}{2}$	$1\frac{1}{4}$	1
$\frac{1}{2}$	$\frac{3}{4}$	$\frac{1}{4}$	1
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{2}$	1
$\frac{3}{4}$	1	$\frac{1}{4}$	1
$\frac{3}{4}$	0	0	1
$\frac{3}{4}$	$\frac{1}{2}$	0	1
1	$-\frac{1}{4}$	0	1
1	$\frac{1}{4}$	$1\frac{1}{4}$	1
2	$1\frac{3}{4}$	1	1

Oligomeric assemblies were analyzed by us primarily for macromolecules crystallized in space group $P1$ since, in our experience, many models presented in this space group disregard the presence of internal symmetry of the molecular assemblies.

The following criteria were used for the analysis of crystal structures with potentially higher symmetry than that presented in the PDB. Only structures with more than one molecule in the asymmetric unit in the triclinic, monoclinic, orthorhombic and cubic space groups were inspected. In addition, structures that were declared as twinned and those for which the statistics of the diffraction data suggested the possibility of twinning were omitted from further analysis. The diffraction data for low-symmetry structures (taken from the PDB CIF files) with cell dimensions potentially giving rise to higher symmetry lattice metrics were submitted to

electronic reprint

POINTLESS (Evans, 2011) and/or *XPREP* (Sheldrick, 2015) and, if appropriate, were merged in the high-symmetry point group by *XPREP*. The structures were then re-solved in the high-symmetry space group with *MOLREP* (Vagin & Teplyakov, 2010) and refined with *REFMAC*5 (Murshudov *et al.*, 2011). The model refinement was not comprehensive, consisting of only ten cycles of automatic isotropic refinement with no attempt to interpret the structures using computer graphics. All models were stripped of water molecules, and for structures with a data resolution of 2.5 Å or better, the solvent was acquired after every refinement cycle by *ARP*/*wARP* (Perrakis *et al.*, 1999); water molecules were not included in lower resolution structures.

References to the appropriate original publications can be found in the headers of the PDB files. We must stress that we did not aim to provide a comprehensive analysis of the whole contents of the PDB, but only searched for a sufficiently large number of specific examples to illustrate the points to be made below.

3. Results and discussion

3.1. Placement of molecules away from the unit cell

In many crystal structures in the PDB the molecules are located outside the unit cell. Such placement may sometimes make it difficult to inspect the intermolecular interactions, since the crystallographic transformations bringing various molecules close together then require the application of multiple translations. For example, if a molecule at position x, y, z in space group $P2$, forming a functional dimer with its mate at $-x, y, -z$, is instead located at $2 + x, y, z$, its mate requires transformation by $4 - x, y, -z$. Such a transformation may be trivial for a crystallographer, but not even all

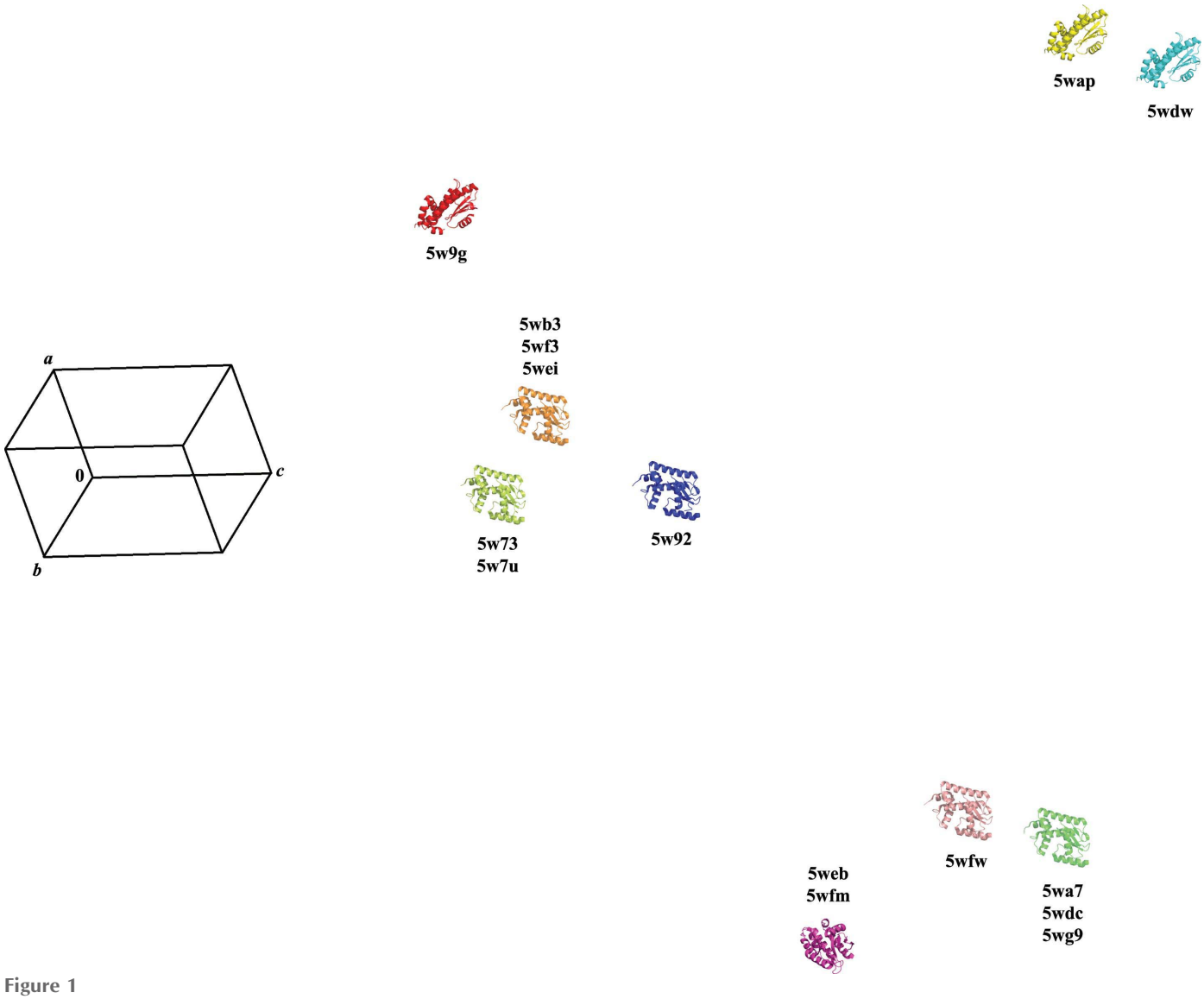


Figure 1
Positioning of the molecules in 15 isomorphous *I*422 structures quoted in Table 1(*b*) in nine different locations, all of them at a distance from the unit cell. The view is a projection down an arbitrary direction that does not correspond to any crystal axis and the molecules are drawn to scale. This being a projection, significant shifts in the direction perpendicular to the viewing axis cannot be seen.

electronic reprint

crystallographic programs would treat this case properly. For example, *Coot* (Emsley & Cowtan, 2004) will properly display all symmetry-equivalent molecules around any position in the lattice, but the program *CONTACT* (Winn *et al.*, 2011) will miss all intermolecular contacts if the translation component is larger than two unit cells. Table 1 contains examples of structures consisting of only a single molecule, which is located further than three unit-cell lengths from the origin. Some of these crystals contain polar rotation axes with a floating cell origin and molecules shifted away from the cell along these axes during model refinement, but in other cases there is not even such an explanation for such an illogical placement of molecules.

An interesting case is a collection of 25 structures of different complexes of influenza virus endonuclease with various inhibitors that were published in a single paper (Beylkin *et al.*, 2017). These structures were crystallized in three space groups, with a single protein molecule found in the asymmetric unit, but were presented in various locations around the unit cell (Table 1*b*). The molecule resides within the cell only in the case of the hexagonal crystal form, whereas 15 different tetragonal coordinate sets are widely scattered (Fig. 1).

In our opinion, it is definitely beneficial to place molecules in a standardized location within one selected asymmetric unit of the cell, such as that defined in *International Tables for Crystallography* (2005). This can be achieved, for example, by using the *ACHESYM* server (Kowiel *et al.*, 2014). It is unfortunate that the various molecular-replacement programs that are used to solve the majority of structures to be deposited in the PDB tend to locate the molecules outside the cell, even if their algorithms are able to construct most compact multimolecular oligomers.

This problem has another aspect related to comparisons of multiple versions of identical or isomorphous crystal structures in the PDB. This was previously pointed out by us (Dauter, 2013) by highlighting an example showing that 59 isomorphous crystal structures of bovine trypsin were present in the PDB with molecules in 11 different locations. Many of them were evidently independently 'solved' by molecular

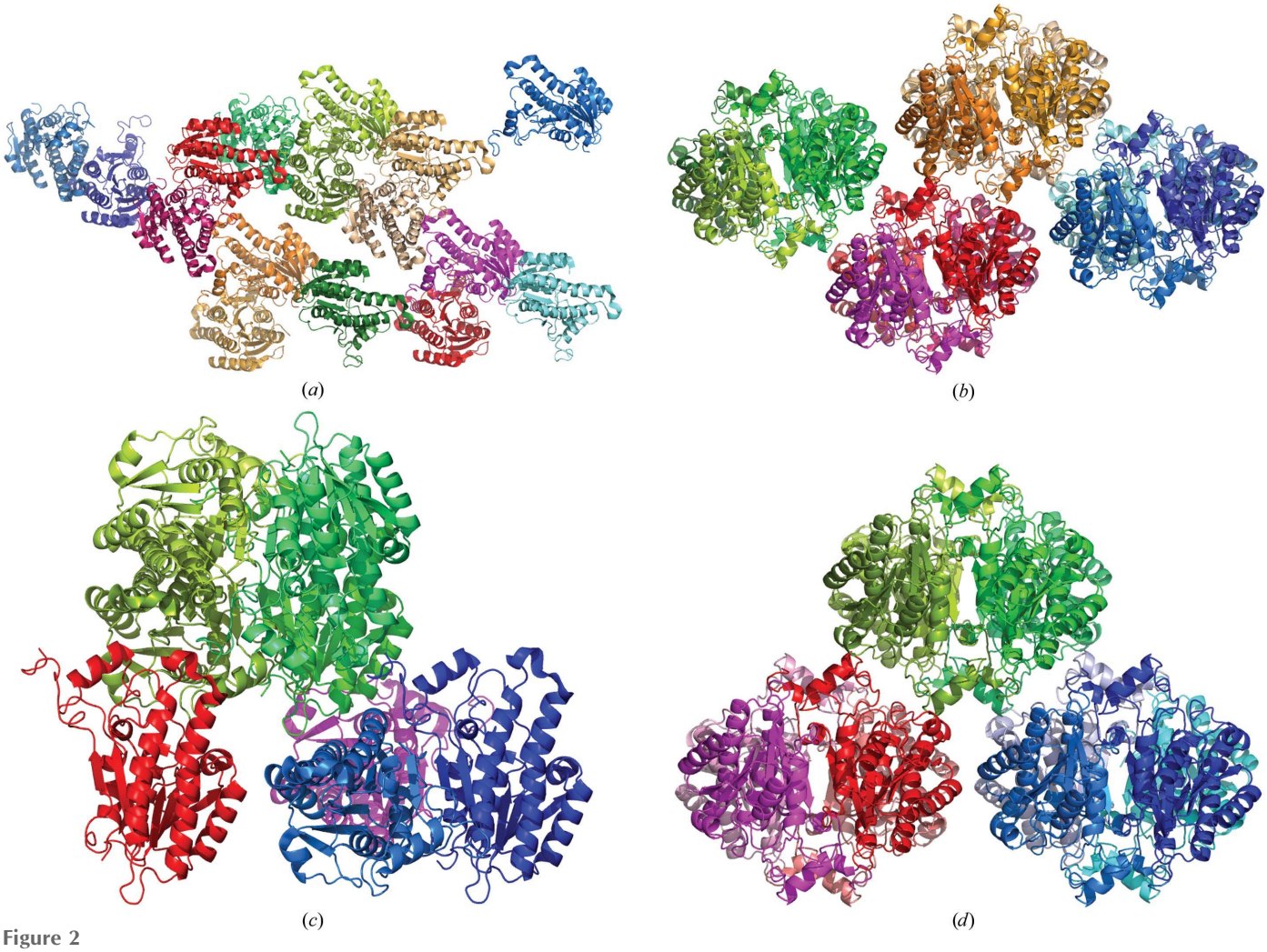


Figure 2
Arrangement of molecules in the structures with PDB codes 4cr7 and 4cr8, with molecules belonging to the same tetramers presented in different shades of the same color. (*a*) 16 molecules in the PDB presentation of the structure 4cr7. (*b*) The same 16 molecules presented as four compact tetramers. (*c*) 12 molecules in the structure 4cr8 as presented in the PDB. (*d*) The same molecules rearranged into compact assemblies with two tetramers (blue and red) completed by two additional molecules (in pale colors) related by crystallographic twofold axes.

electronic reprint

replacement, instead of 'borrowed' from the available previous structures. To illustrate this point again, we have identified 27 different locations of molecules of hen egg-white lysozyme in 500 isomorphous structures determined in the tetragonal form and deposited in the PDB (Table 2). This might not appear to be of crucial importance, yet it is our postulation that, for example, one of the reasons for the misinterpretation of the identity of ligands in three isomorphous structures of kynurenine aminotransferase (Wlodawer *et al.*, 2018) could have been a side effect of each of them locating the molecules elsewhere, thus making it impossible (or at least more difficult) to simultaneously look at all electron densities of the ligands.

3.2. Disintegration of oligomers

Even more confusing are cases in which the subunits of multimeric proteins are spread in different locations, instead of forming compact oligomers that represent their functional forms. Table 3 contains examples of wrongly presented oligomers, limited only to space group *P*1 with more than five molecules in the asymmetric unit.

An example is found in the paper by Sola-Carvajal *et al.* (2014) that discusses the structures of a tetrameric enzyme in three different states: the apoenzyme as well as complexes with a substrate and with a cofactor. These three structures were determined in different crystal forms. The apo structure (PDB entry 4cr6) is presented as a single tetramer with 222 symmetry. The substrate complex (PDB entry 4cr7) in fact contains four tetramers, but is presented as one trimer and a separated monomer, as well as three instances of a dimer and two separated monomers. The cofactor complex (PDB entry 4cr8) contains one tetramer in a general position in the *C*2 cell and two dimers located at the twofold axes forming tetramers with themselves. However, one of these dimers is presented as two separated monomers. The locations of the molecules in the structures with PDB codes 4cr7 and 4cr8 as presented in the PDB are shown in Figs. 2(*a*) and 2(*c*), respectively, and those after a more logical rearrangement are shown in Figs. 2(*b*) and 2(*d*), respectively. Sola-Carvajal *et al.* (2014) clearly state that this enzyme always forms tetrameric oligomers, but the presentation in the PDB may suggest that the formation of complexes changes the oligomeric state of this enzyme. This is just one example illustrating why the optimal presentation of the structures in the PDB matters.

3.3. Structures presented with incorrect symmetry

All crystal structures can, in principle, be expressed in space group *P*1, ignoring all crystallographic symmetry operations other than lattice translations. Such a mode of presentation of crystal structures will correctly preserve all intermolecular interactions but, nevertheless, will not be strictly correct. The arrangement of molecules in the crystal lattice always has a particular symmetry, which is preserved after averaging over space and time. The presentation of crystal structures in a wrong (too low) symmetry negatively affects the process of model refinement and distorts the statistical results obtained.

Table 3
The structures in *P*1 symmetry with more than five independent molecules spread in different locations, but in reality forming compact oligomers.

The entry '4 × 2 ABCD,EF+G+H' means that the structure contains two tetramers, one tetramer *ABCD* and a second tetramer formed by the dimer *EF* and molecules *G* and *H* transformed by the space-group symmetry operations.

PDB code	Oligomers	Presentation in the unit cell
1tzn	28 × 1	ABCDEFOabcdefo+GHIJKLMghijklm
4riq	12 × 2	ABCJKLVXZ+DEF,MNOPQRSTU+GHI
5aew	6 × 4	STUVWX,CDEFIJ,ABKLOP,GHMN+QR
2xso	6 × 4	CDEFGH,KLMNQR,STUVWX,IJOP+AB
5k02	2 × 12	AB,EJ,MN,QV,C+L,D+K,F+I,G+H,O+X,P+W,R+U,S+T
4fmj	10 × 2	ABCDE+KLMNO,FGHIJ+PQRST
3gnd	10 × 2	ABCDEFGHIJ,KLMNO+PQRST
3gkf	10 × 2	ABCDEFGHIJ,KLMNO+PQRST
1xls	16 × 1	ACEGIKMO+BFJN+DHLP
5exc	8 × 2	ACEHIKMP,BDJL+FGNO
4cr7	4 × 4	AK+H+N,BLM+G,CE+JP,DF+IO
2ov2	4 × 4	ABIJ,EHMP,CK+DL,FN+GO
3zds	6 × 2	ABCDEG,HIK+F+J+L
1l2w	6 × 2	CDGHJL,ABJ+EFK
3m6s	6 × 2	ABCDEF,IJKL+GH
4qql	3 × 3	ABC,EHI,DF+G
3ukf	8 × 1	BDEG,AH+CF
3uhj	8 × 1	ABCD+EH+FG
1s5u	8 × 1	ABCD+EFGH
1nr4	8 × 1	AB+CD+EF+GH
3n6q	8 × 1	ABCD+EFGH
4cr8	4 × 2	CDEF,AB+A′B′,G+H+G′+H′
5jyo	4 × 2	BCDE,A+F+G+H
4uxj	4 × 2	ABCD,EF+GH
4c7u	4 × 2	BCH+F,AE+D+G
3ukl	4 × 2	ABDF,CE+GH
3tpc	4 × 2	ABCD,EG+F+H
5lsm	2 × 4	BC,DF,A+H,E+G,
4u0z	2 × 4	BH,A+G,C+E,D+F
4obu	2 × 4	BC,EH,A+U,F+G
4bof	2 × 4	AH,BC,FG,D+E
4bjq	2 × 4	AC,BE,DH,F+G
3n9r	2 × 4	Ae,B+j,K+Z,P+U
3h06	2 × 4	EP,GJ,HL,B+N
2j4e	2 × 4	AB,FG,C+H,D+E
4wa1	3 × 2	BCD,EF+A
3dls	3 × 2	ABC,D+E+F
5mms	2 × 3	BF,DE,A+C
4tkz	2 × 3	AB,C+D,E+F

It multiplies the number of refined parameters, whereas the number of independent observables (reflections) is not increased. The intensities in groups of reflections related by symmetry that has not been accounted for are not independent, and all observed differences between them are only caused by measurement errors. Merging diffraction data in higher symmetry, if such is indeed present, increases their multiplicity and enhances the resulting accuracy of estimated intensities.

We have analyzed the low-symmetry (and cubic) crystal structures having a potentially higher lattice metric than the symmetry declared in the PDB depositions and identified many instances that could be successfully refined in a higher symmetry space group. Our analysis is by no means comprehensive, but the representative examples with different combinations of low/high-symmetry space groups are presented in Table 4. These crystal structures refine in higher symmetry without the assumption of twinning, a phenomenon

electronic reprint

Table 4
Examples of high-symmetry crystal structures deposited in the PDB in low-symmetry space groups.

The columns show the original unit-cell parameters, space group and R and R_{free} factors taken from the header of the PDB deposition, as well as the results of simple refinement in the higher symmetry space group (after appropriate conversion of the unit-cell parameters; not shown).

PDB code	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	Space group	R/R_{free} (%)	Space group	R/R_{free} (%)
3kqu	116.534	116.467	71.114	90.00	90.00	119.97	$P1$	19.7/21.5	$P6_1$	14.8/23.3
1sed	56.784	64.550	64.503	111.37	107.19	107.13	$P1$	18.9/20.9	$R3$	15.6/19.8
3l1w	68.565	68.558	104.245	89.88	89.79	60.07	$P1$	16.7/19.8	$P3_121$	17.5/20.8
4p9j	37.580	37.570	95.820	89.97	89.97	120.05	$P1$	16.6/22.0	$P3_1$	15.6/22.7
5jyo	126.404	126.629	126.269	112.88	102.81	112.74	$P1$	19.0/21.3	$I4_122$	16.9/19.6
3hz2	29.300	54.202	54.179	85.81	74.31	74.32	$P1$	17.2/21.2	$I4$	13.7/19.2
5ja3	59.612	60.444	60.472	90.07	90.06	89.93	$P1$	22.7/26.7	$P4$	17.7/22.7
1gc0	72.861	81.030	81.282	70.56	63.17	63.38	$P1$	21.0/23.6	$I222$	18.9/21.4
2qmh	69.059	106.368	106.486	119.50	90.02	89.96	$P1$	22.5/29.8	$C222_1$	20.3/27.3
5hd6	42.665	50.816	138.780	90.02	90.02	90.10	$P1$	13.8/17.0	$P2_12_12_1$	14.7/18.4
2z66	49.787	67.971	122.779	89.86	89.96	90.01	$P1$	21.0/21.9	$P2_12_12$	17.8/21.9
5god	38.240	38.260	48.000	98.15	98.15	109.67	$P1$	20.3/22.2	$C2$	18.7/22.5
3tqs	38.007	57.895	112.938	90.00	90.01	104.34	$P1$	18.7/24.6	$P2_1$	16.3/22.4
2o3f	58.898	48.060	58.854	90.00	119.92	90.00	$P2$	18.3/21.5	$P6_4$	16.4/21.2
1xtf	57.900	40.490	195.890	90.00	90.25	90.00	$P2$	22.0/27.3	$P2_12_12$	18.6/26.2
3m9b	176.787	176.652	176.633	90.00	90.04	90.00	$P2_1$	27.6/30.4	$P2_13$	21.6/25.7
3cdf	72.959	98.043	73.074	90.00	119.60	90.00	$P2_1$	19.1/23.1	$P6_1$	20.5/21.4
3cue	115.120	115.400	290.070	90.00	90.28	90.00	$P2_1$	26.5/29.9	$P4_12_12$	20.7/30.0
1aw2	89.660	137.820	89.540	90.00	90.94	90.00	$P2_1$	20.0/21.9	$C222_1$	16.9/21.7
1xfy	315.622	182.044	141.024	90.00	89.93	90.00	$C2$	26.9/28.9	$P6_522$	20.7/30.8
5apg	115.950	66.970	106.420	90.00	89.99	90.00	$C2$	19.3/20.3	$P3_221$	17.8/20.1
1xs0	81.339	46.962	88.044	90.00	89.95	90.00	$C2$	22.5/26.0	$P3_121$	20.9/25.7
2inu	159.138	91.907	93.022	90.00	124.75	90.00	$C2$	19.8/22.6	$R32$	16.2/19.2
3dg7	171.047	124.265	117.485	90.00	133.39	90.00	$C2$	18.8/20.3	$I422$	16.9/20.2
3tox	122.975	122.942	139.818	90.00	90.03	90.00	$C2$	20.3/25.5	$P4_12_12$	19.2/23.9
3kb4	160.106	167.692	113.220	90.00	135.16	90.00	$C2$	19.4/22.0	$I222$	18.3/21.4
3hx9	43.971	64.622	71.082	90.00	90.01	90.00	$C2$	18.9/23.1	$C222_1$	17.2/22.0
1zwk	73.314	73.331	78.672	90.00	90.00	90.00	$P222$	22.9/27.3	$P4_222$	18/4/28.6
5sww	90.148	90.199	167.257	90.00	90.00	90.00	$P222_1$	21.4/26.3	$P4_322$	19.5/27.4
3cn9	96.423	96.363	68.044	90.00	90.00	90.00	$P2_12_12$	19.6/24.0	$P42_12$	17.8/22.1
5t9d	122.220	122.380	122.240	90.00	90.00	90.00	$P2_12_12_1$	21.4/23.1	$P2_13$	17.7/22.4
3imd	42.239	42.243	110.384	90.00	90.00	90.00	$P2_12_12_1$	19.9/23.7	$P4_12_12$	16.7/23.6
4dd1	77.887	78.703	37.074	90.00	90.00	90.00	$P2_12_12_1$	19.1/22.7	$P4_32_12$	17.5/23.0
5xz0	174.565	174.630	48.769	90.00	90.00	90.00	$C222$	20.8/24.3	$P42_12$	17.8/25.2
4db9	131.920	228.880	116.530	90.00	90.00	90.00	$C222_1$	21.3/24.6	$P6_322$	19.5/24.8
3w51	94.276	95.038	116.497	90.00	90.00	90.00	$C222_1$	19.0/24.3	$P4_12_12$	17.7/23.3
3gzn	135.030	228.715	229.307	90.00	90.00	90.00	$I222$	23.3/28.7	$I422$	19.6/26.0
3wi6	180.061	179.684	254.101	90.00	90.00	90.00	$I2_12_12_1$	24.2/28.3	$F4_132$	20.4/25.8
4xc5	112.930	112.960	113.220	90.00	90.00	90.00	$I2_12_12_1$	19.3/22.3	$I2_13$	19.0/22.7
2nsd	91.263	91.327	184.259	90.00	90.00	90.00	$I2_12_12_1$	23.6/25.9	$I4_122$	20.8/24.6
3jxe	170.198	170.198	170.198	90.00	90.00	90.00	$P2_13$	23.9/25.8	$P4_132$	18.6/23.8
2xns	265.261	265.261	265.261	90.00	90.00	90.00	$I23$	22.3/24.4	$I432$	19.3/22.4
3wiz	152.661	152.661	152.661	90.00	90.00	90.00	$I2_13$	19.8/23.8	$I4_132$	19.0/23.6
4wwv	234.323	234.323	234.323	90.00	90.00	90.00	$F23$	20.7/26.3	$F4_132$	19.7/26.6

that potentially would explain the low-symmetry treatment of these crystals.

An instructive example is the structure of a DNA-binding protein (PDB entry 5t9d; Billon *et al.*, 2017) presented as a symmetric trimer positioned at the diagonal of the orthorhombic cell in space group $P2_12_12_1$, with the three cell dimensions almost identical and, of course, all right angles (Fig. 3). This data set can be rescaled by applying the cubic 23 symmetry with an R_{merge} of 3.4% and it refines in $P2_13$ to R and R_{free} values of 17.7 and 22.4%, respectively, which are lower than the values quoted for the orthorhombic symmetry.

The problem of presenting crystal structures in the wrong symmetry was addressed in the past by Richard Marsh, who was very active in correcting ('marshing') wrongly interpreted structures of small molecules. We completely agree with his opinion, as expressed in the following quotation (Marsh & Bernal, 1995):

> Finally, a referee suggests that it would be well to emphasize *why* it matters that the symmetry be correct, noting that noncrystallographers 'are prone to thinking papers like this one are hopelessly pedantic'. It is tempting to knock that complaint aside with the comment that almost all details, such as the precise conditions for a chemical reaction, may be considered 'hopelessly pedantic' by scientists in other fields. But that is beside the point: accepting incorrect results in order to avoid the label 'pedantic' is contrary to accepted standards of scientific behavior ... We can think of no valid excuse for considering the choice of space group as unimportant, or for condoning an incorrect choice.

Of course, not all structures presented in a low-symmetry space group with metrics indicating a possibility of higher symmetry result from carelessness of the depositors. An example may be provided by the structure of the ICAP1–integrin $\beta1$ complex (PDB entry 4dx9; Liu *et al.*, 2013), with 32

electronic reprint

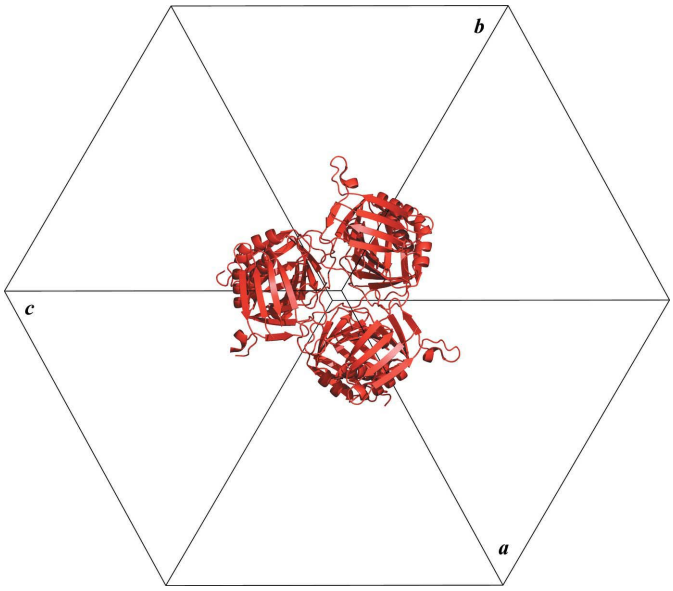


Figure 3
The structure with PDB code 5t9d, consisting of three molecules forming a symmetric trimer in the $P2_12_12_1$ unit cell with three equal edges and three right angles. The view is almost along the cell diagonal.

molecules placed in a cell in space group $P1$. The unit-cell parameters a = 75.62, b = 122.21, c = 135.27 Å, α = 89.97, β = 89.99, γ = 108.11° in space group $P1$ may be transformed to a = 75.41, b = 230.84, c = 134.84 Å, $\alpha = \beta = \gamma = 90°$ in space group $C222_1$. Since the original diffraction images are available on the SBGrid server (https://doi.org/10.15785/SBGRID/550), we independently reprocessed these data and re-solved the structure. Although the systematic absences clearly indicate the presence of a 2_1 axis and scaling in the orthorhombic space group is acceptable, it was not possible to refine the structure in this setting. We conclude that this example of a highly twinned crystal may indicate the presence of crystallographic artifacts that neither the original authors nor the authors of this paper could unambiguously identify, leaving refinement in $P1$ as the only feasible option.

4. Conclusion

We hope that this note will initiate a discussion that might ultimately result in the requirement that the deposited coordinates be presented in the most logical way, preferably by including checking of the symmetry and proper placement of the molecules as part of the validation process, as well as requiring that isomorphous structures are presented in a consistent manner.

Funding information

This work was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

References

Beylkin, D., Kumar, G., Zhou, W., Park, J., Jeevan, T., Lagisetti, C., Harfoot, R., Webby, R. J., White, S. W. & Webb, T. R. (2017). *Sci. Rep.* **7**, 17139.
Billon, P., Li, J., Lambert, J.-P., Chen, Y., Tremblay, V., Brunzelle, J. S., Gingras, A.-C., Verreault, A., Sugiyama, T., Couture, J.-F. & Côté, J. (2017). *Mol. Cell*, **65**, 78–90.
Burley, S. K., Berman, H. M., Christie, C., Duarte, J. M., Feng, Z., Westbrook, J., Young, J. & Zardecki, C. (2018). *Protein Sci.* **27**, 316–330.
Dauter, Z. (2013). *Acta Cryst.* D**69**, 2–4.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Evans, P. R. (2011). *Acta Cryst.* D**67**, 282–292.
International Tables for Crystallography (2005). Vol. A, *Space-group Symmetry*, 5th ed., edited by T. Hahn. Dordrecht: Kluwer Academic Publishers.
Kowiel, M., Jaskolski, M. & Dauter, Z. (2014). *Acta Cryst.* D**70**, 3290–3298.
Liu, W., Draheim, K. M., Zhang, R., Calderwood, D. A. & Boggon, T. J. (2013). *Mol. Cell*, **49**, 719–729.
Marsh, R. E. & Bernal, I. (1995). *Acta Cryst.* B**51**, 300–307.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Perrakis, A., Morris, R. M. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Sheldrick, G. M. (2015). *Acta Cryst.* C**71**, 3–8.
Sola-Carvajal, A., Gil-Ortiz, F., García-Carmona, F., Rubio, V. & Sánchez-Ferrer, Á. (2014). *Biochem. J.* **462**, 499–511.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Wlodawer, A., Dauter, Z., Porebski, P. J., Minor, W., Stanfield, R., Jaskolski, M., Pozharski, E., Weichenberger, C. X. & Rupp, B. (2018). *FEBS J.* **285**, 444–466.

electronic reprint