BIOLOGICAL CRYSTALLOGRAPHY

ISSN 1399-0047

On the accuracy of unit-cell parameters in protein crystallography

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

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

Received 18 July 2015
Accepted 19 August 2015

Edited by Z. S. Derewenda, University of Virginia, USA

Keywords: protein crystallography; unit-cell parameter accuracy; symmetry.

The availability in the Protein Data Bank (PDB) of a number of structures that are presented in space group *P*1 but in reality possess higher symmetry allowed the accuracy and precision of the unit-cell parameters of the crystals of macromolecules to be evaluated. In addition, diffraction images from crystals of several proteins, previously collected as part of in-house projects, were processed independently with three popular software packages. An analysis of the results, augmented by published serial crystallography data, suggests that the apparent precision of the presentation of unit-cell parameters in the PDB to three decimal points is not justified, since these parameters are subject to errors of not less than 0.2%. It was also noticed that processing data including full crystallographic symmetry does not lead to deterioration of the refinement parameters; thus, it is not beneficial to treat the crystals as belonging to space group *P*1 when higher symmetry can be seen.

1. Introduction

Unit-cell parameters are estimated in crystallographic diffraction experiments from diffraction angles according to Bragg's law: $n\lambda = 2d\sin\theta$. In four-circle diffractometry the angles θ are measured directly and if the X-ray wavelength λ is known accurately, as is the case for sealed-tube or rotating-anode sources, with a well calibrated goniostat, the measurement of diffraction angles for several reflections spread over reciprocal space permits the estimation of the resolution of each reflection (or the corresponding interplanar spacing d_{hkl}) and the crystal unit-cell parameters with high accuracy from the following geometric relations (Giacovazzo, 2011).

The more elaborate forms of Bragg's equation are as follows. In reciprocal space,

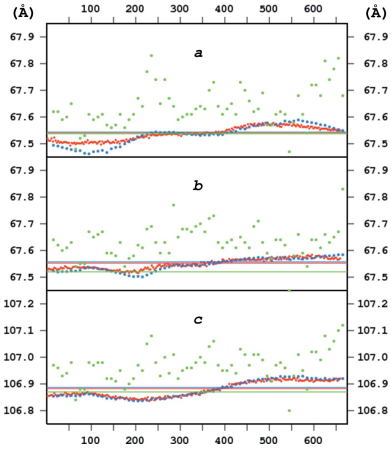
$$\begin{aligned} d_{hkl}^2 &= \lambda/4\sin^2\theta \\ &= 1/(h^2a^{*2} + k^2b^{*2} + l^2c^{*2} + 2hka^*b^*\cos\gamma^* \\ &\quad + 2hla^*c^*\cos\beta^* + 2klb^*c^*\cos\alpha^*), \end{aligned} \tag{1}$$

or in direct space,

$$\begin{aligned} d_{hkl}^2 &= \lambda/4\sin^2\theta \\ &= [h^2\sin^2\alpha/a^2 + k^2\sin^2\beta/b^2 + l^2\sin^2\gamma/c^2 \\ &\quad + 2kl(\cos\beta\cos\gamma - \cos\alpha)/bc + 2lh(\cos\alpha\cos\gamma \\ &\quad - \cos\beta)/ac + 2hk(\cos\alpha\cos\beta - \cos\gamma)/ab]/ \\ &\quad (1 - \cos^2\alpha - \cos^2\beta - \cos^2\gamma + 2\cos\alpha\cos\beta\cos\gamma). \end{aligned} \tag{2}$$

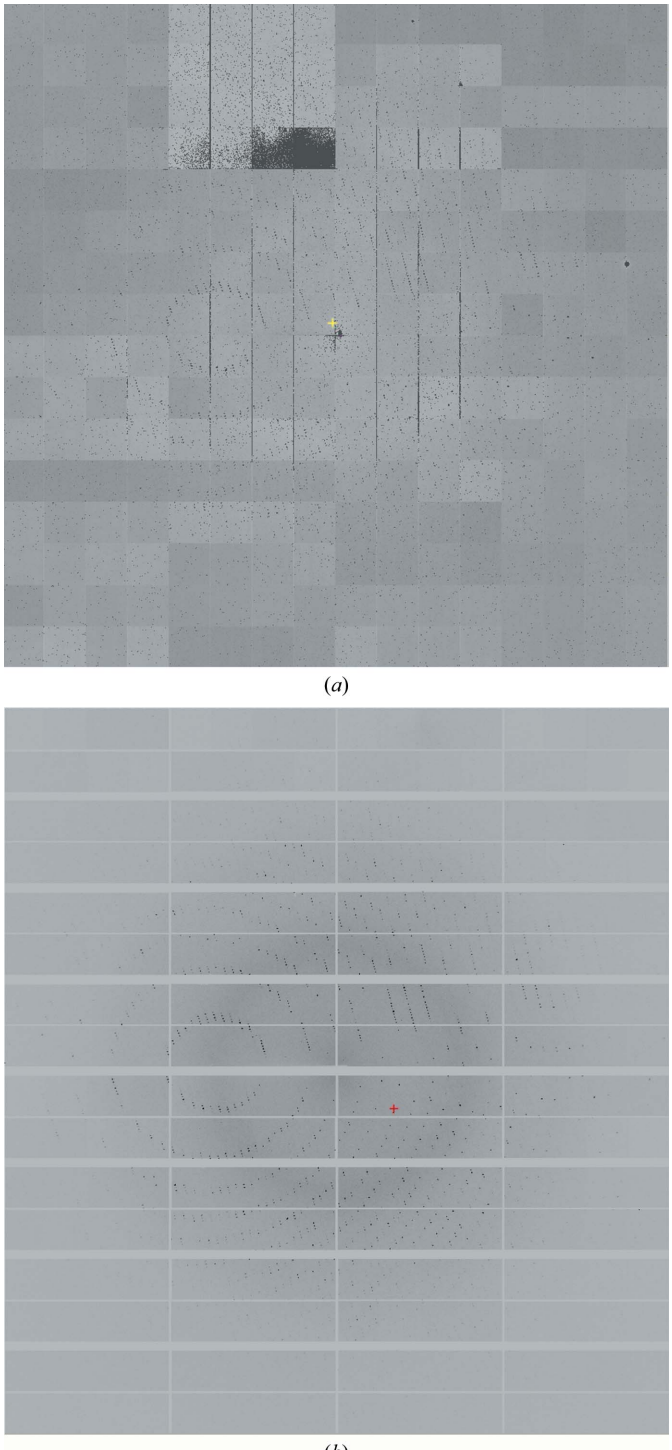
For higher symmetry crystals these relationships simplify according to the constraints on some unit-cell parameters.

If diffraction data are measured on two-dimensional detectors, as is currently routine not only in macromolecular



© 2015 International Union of Crystallography

crystallography but also in small-structure work, the estimation of unit-cell parameters is somewhat complicated and is potentially influenced by various systematic errors. Some unit-cell parameters may be poorly determined in a single diffraction image, since such an image represents only a narrow cross-section of reciprocal space. The crystal-to-detector distance is usually not known with high accuracy and the value of the X-ray wavelength may also not be set very accurately. Precise calibration of these parameters requires special attention, which is rarely followed in routine experiments at synchrotron stations, where the detector distance, X-ray wavelength and other beam parameters are frequently changed from one experiment to another. An additional factor contributing to the uncertainty is owing to the geometric calibration of a particular detector, as represented in its correction tables. The necessity of such geometric correction is evident from a visual comparison of a 'raw' diffraction image and a corrected diffraction image (Fig. 1). Many geometric parameters of the experimental system are refined during data processing, such as, for example, the detector 'tilt' and 'twist' that describe its deviation from perpendicularity to the beam. It is, however, always assumed that the spindle axis is precisely perpendicular to the beam and this feature is not a refinable parameter. If this condition is not strictly preserved, the unit-cell parameters of the crystal and its orientation angles estimated during data processing change somewhat from image to image while the crystal is rotated.



(*a*)

(*b*)

Figure 1
A raw diffraction image collected on a pixel detector (*a*) and the same image after application of the calibration and geometric corrections (*b*). These figures were obtained courtesy of Dr K. Rajashankar.

In addition, certain systematic errors and uncertainties may be introduced by the crystals themselves. Some macromolecular crystals are not perfectly uniform in terms of their lattices and diffraction properties within their whole volume. If the X-ray beam cross-section is smaller than the crystal size, the diffraction data may come from different non-isomorphous parts of the specimen while the crystal rotates during data collection. Protein crystals irradiated by strong synchrotron X-radiation are influenced by radiation damage, causing not only local structural and chemical changes in the investigated samples but also degradation of diffracted intensities, especially at high resolution. These effects lead to changes in the crystal mosaicity and unit-cell parameters (Ravelli & McSweeney, 2000). If diffraction data are measured and merged from several crystals, as is practiced in the 'serial crystallography' approach (Gati *et al.*, 2014), the individual specimens may be non-isomorphous to some extent, having somewhat different unit-cell parameters. All of these effects diminish the accuracy of the finally obtained unit-cell parameters, which then represent a set of 'averaged' values.

The optimal values of the unit-cell and certain other parameters (for example, crystal mosaicity *etc.*) in data collection by the rotation method with two-dimensional detectors are obtained during data merging and scaling by the so-called post-refinement procedure. This involves the global optimization of many parameters, some refined as constant for all diffraction images (mainly unit-cell parameters) and others estimated individually for each image or batch of a few images (crystal orientation, mosaicity, crystal-to-detector distance). Details of post-refinement algorithms and their execution differ in different data-processing programs, and thus the resulting unit-cell parameters also differ to some extent. In addition, processing the same diffraction images in different Laue groups (but the same lattice), applying the appropriate symmetry constraints, results in somewhat different unit-cell parameters.

Table 1
Selected structures presented in the PDB in *P*1 symmetry that in reality possess higher symmetry.

The unit-cell parameters are given in the original *P*1 space group, in the equivalent, nonstandard centered lattice and in the correct, higher symmetry space group. The data resolution and the R_{merge} values as quoted in the PDB and resulting from merging the deposited data in higher symmetry are also given, as well as R and R_{free} quoted in *P*1 and resulting from re-refinement in higher symmetry. Maximum deviations of the cell lengths (as an absolute number and as a percentage) and angles between the *P*1 values and those calculated utilizing the higher symmetry lattice restrictions are given in the last three columns.

Space group	*a* (Å)	*b* (Å)	*c* (Å)	*α* (°)	*β* (°)	*γ* (°)	Resolution (Å)	R_{merge} (%)	R/R_{free} (%)	*δ*(*a*) (Å)	*δ*(*a*)/*a* (%)	*δ*(*α*) (°)
1f1g (Hart *et al.*, 1999)												
*P*1	72.500	72.480	72.470	109.20	109.55	109.21	1.35	8.8	17.1/n/a			
*R*1	118.153	118.185	73.023	89.82	90.20	119.86				0.030	0.04	0.35
*R*32	118.169	118.169	73.023	90.0	90.00	120.00		10.1	17.4/18.9			
2h6l (Northeast Structural Genomics Consortium, unpublished work)												
*P*1	47.673	47.714	47.665	70.75	70.81	70.83	2.00	n/a	18.7/23.7			
*R*1	55.217	55.234	106.342	89.98	89.95	119.94				0.049	0.10	0.08
*R*3	55.225	55.225	106.342	90.00	90.00	120.00		5.3	19.2/28.1			
2zdc (RIKEN Structural Genomics/Proteomics Initiative, unpublished work)												
*P*1	53.434	53.408	53.419	108.56	108.57	108.48	2.00	3.7	21.8/5.6			
*R*1	86.699	86.731	55.837	89.96	89.98	119.97				0.026	0.05	0.09
*R*3	86.715	86.715	55.837	90.00	90.00	120.00		7.7	20.1/27.0			
3kse (M. Renko & D. Turk, unpublished work)												
*P*1	35.233	83.948	83.906	118.07	98.04	98.04	1.71	4.0	15.3/20.4			
*R*1	143.911	143.932	35.233	90.00	89.99	119.97				0.042	0.05	0.00
*R*3	143.921	143.921	35.233	90.00	90.00	120.00		4.0	16.6/21.8			
1sed (Midwest Center for Structural Genomics, unpublished work)												
*P*1	56.784	64.550	64.503	111.37	107.19	107.13	2.10	6.1	18.9/20.9			
*R*1	106.591	107.125	56.784	89.81	90.03	119.77				0.047	0.07	0.06
*R*3	106.858	106.858	56.784	90.00	90.00	120.00		6.7	15.6/19.8			
1op8 (Hink-Schauer *et al.*, 2003)												
*P*1	49.480	94.550	94.870	117.12	100.25	100.12	2.50	4.5	21.8/28.4			
*R*1	160.912	161.394	49.480	90.32	89.78	119.81				0.320	0.34	0.13
*R*3	161.153	161.153	49.480	90.00	90.00	120.00		5.4	17.2/28.0			
3ds5 (Bartonova *et al.*, 2008)												
*P*1	51.359	51.321	51.358	109.20	109.63	109.55	2.40	2.5	22.3/27.0			
*I*1	59.188	59.228	59.478	90.02	89.95	89.95				0.037	0.07	0.08
*I*41	59.208	59.208	59.478	90.00	90.00	90.00		2.6	22.1/n/a			
3u58 (Zeng *et al.*, 2011)												
*P*1	83.046	83.110	82.885	108.45	111.59	108.42	2.61	10.3	21.9/25.9			
*I*1	97.041	97.171	93.279	89.98	90.15	89.92				0.161	0.19	0.03
*I*41	97.106	97.106	93.279	90.00	90.00	90.00		9.2	19.2/26.0			
4mjm (Center for Structural Genomics of Infectious Diseases, unpublished work)												
*P*1	84.328	84.249	84.313	110.01	109.22	109.19	2.25	5.2	20.9/23.2			
*I*1	97.665	97.667	96.671	89.99	90.05	89.97				0.064	0.08	0.03
*I*4	97.666	97.666	96.671	90.00	90.00	90.00		6.2	14.7/21.2			
2oqy (Rakus *et al.* (2009)												
*P*1	104.635	104.640	104.462	109.50	109.45	109.47	2.00	6.7	21.5/22.9			
*I*1	120.754	120.827	120.682	89.95	89.90	90.04				0.005	0.01	0.05
*I*4	120.790	120.790	120.682	90.00	90.00	90.00		4.9	18.2/24.8			
3es8 (Rakus *et al.*, 2009)												
*P*1	105.584	105.695	105.718	109.31	109.47	109.66	2.20	7.9	22.4/24.1			
*I*1	121.698	121.997	122.302	90.02	89.99	89.92				0.109	0.10	0.16
*I*4	121.848	121.848	122.302	90.00	90.00	90.00		8.9	18.9/27.1			
2r8e (Biswas *et al.*, 2009)												
*P*1	82.877	83.005	85.864	118.84	118.77	90.06	1.40	3.5	16.0/18.8			
*I*1	82.877	83.005	125.616	89.97	89.94	90.06				0.128	0.15	0.07
*I*4	82.941	83.941	125.616	90.00	90.00	90.00		4.0	15.1/18.9			
3hz2 (Aravind *et al.*, 2009)												
*P*1	29.300	54.202	54.179	85.81	74.31	74.32	1.86	3.7	17.2/21.2			
*I*1	73.782	73.784	29.300	90.00	90.00	89.97				0.023	0.04	0.01
*I*4	73.783	73.783	29.300	90.00	90.00	90.00		3.0	13.7/19.2			
1gc0 (Motoshima *et al.*, 2000)												
*P*1	72.861	81.030	81.282	70.56	63.17	63.38	1.70	3.7	21.0/23.6			
*I*1	72.861	93.747	110.583	89.93	90.07	90.23				0.252	0.31	0.21
*I*222	72.804	93.804	110.583	90.00	90.00	90.00		4.5	18.9/21.4			
1c03 (Song *et al.*, 1999)												
*P*1	66.343	66.480	66.491	106.37	106.66	115.33	2.30	4.4	20.6/24.8			
*I*1	71.043	79.329	79.681	89.79	90.22	90.04				0.137	0.21	0.29
*I*222	71.186	79.186	79.681	90.00	90.00	90.00		5.8	16.6/23.5			

Comparison of the values of unit-cell parameters obtained by processing the same images in various symmetries and by different programs gives a chance to estimate the lower limit of the uncertainties of these parameters. We have selected a group of structures from the Protein Data Bank (PDB; Berman *et al.*, 2000) presented in space group *P*1 but in reality possessing higher symmetry, and compared the obtained unit-cell parameters with the results of merging the deposited data in higher symmetry. In addition, we have reprocessed several of our own sets of diffraction images with three popular data-reduction programs. Although the true symmetry of these crystals was higher, we assumed space group *P*1. The results obtained with *HKL*-2000 (Otwinowski & Minor, 1997), *XDS* (Kabsch, 2010) and *MOSFLM* (Leslie, 2006) were compared in order to evaluate the true precision (*i.e.* how many decimal digits are meaningful) and accuracy (*i.e.* the deviation of the estimated values from the true values) that is practically possible to obtain. Finally, we analyzed the variations of unit-cell parameters reported in a typical serial data-collection experiment (Axford *et al.*, 2015).

2. Materials and methods

The PDB was searched for structures with *P*1 symmetry having two unit-cell lengths differing by less than 1.0 Å and the two corresponding angles similar to within 1.0°. Among 126 such structures identified in April 2015, 111 were accompanied by diffraction data. The true symmetry for the latter data sets was evaluated with *POINTLESS* (Evans, 2011) and *XPREP* (Sheldrick, 2008). Those data sets in which the possible presence of higher symmetry was indicated were merged in the appropriate space group and submitted to molecular-replacement structure solution by *MOLREP* (Vagin & Teplyakov, 2010), followed by ten cycles of refinement by *REFMAC* (Murshudov *et al.*, 2011), accompanied by automatic incorporation of water molecules with *ARP*/*wARP* (Perrakis *et*

Table 1 (continued)

Space group	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	Resolution (Å)	R_{merge} (%)	R/R_{free} (%)	$\delta(a)$ (Å)	$\delta(a)/a$ (%)	$\delta(\alpha)$ (°)
3ebn (Zhong *et al.*, 2009)												
*P*1	51.395	51.350	51.390	112.22	112.00	104.36	2.40	5.1	20.8/24.3			
*I*1	57.288	57.477	63.001	90.01	90.08	90.02				0.045	0.09	0.22
*I*222	57.383	57.383	63.001	90.00	90.00	90.00		4.6	18.6/26.0			
1u8t (Dyer *et al.*, 2004)												
*P*1	54.280	53.480	54.100	60.36	60.75	60.57	1.50	10.0	20.0/27.3			
*C*1	94.042	53.480	54.803	89.91	125.32	90.02				0.180	0.33	0.21
*C*2	94.042	53.480	54.803	90.00	125.32	90.00		3.2	22.9/27.6			
4bm1 (Fernández-Fueyo *et al.*, 2014)												
*P*1	39.990	75.370	75.610	69.75	75.69	75.82	1.10	5.0	13.1/14.7			
*C*1	123.868	86.323	39.990	90.16	107.46	89.81				0.240	0.32	0.13
*C*2	123.868	86.323	39.990	90.00	107.46	90.00		5.6	14.3/15.8			
2q5c (New York SGX Research Center for Structural Genomics, unpublished work)												
*P*1	45.950	45.948	58.284	72.94	72.98	82.07	1.49	9.8	18.8/21.0			
*C*1	69.319	60.333	58.284	89.97	112.86	90.00				0.002	0.01	0.04
*C*2	69.319	60.333	58.284	90.00	112.86	90.00		5.7	18.8/22.1			
1f5v (Kobori *et al.*, 2001)												
*P*1	51.560	52.860	52.830	75.79	60.71	61.17	1.70	3.3	18.9/20.6			
*C*1	92.152	51.560	64.917	90.31	134.50	89.92				0.030	0.06	0.54
*C*2	92.152	51.560	64.917	90.00	134.50	90.00		5.4	17.2/18.7			
1vg8 (Rak *et al.*, 2004)												
*P*1	57.058	57.071	74.310	71.73	71.69	77.65	1.70	4.1	19.8/22.4			
*C*1	88.914	71.552	74.310	89.97	113.75	89.99				0.013	0.02	0.04
*C*2	88.914	71.552	74.310	90.00	113.75	90.00		3.5	18.8/21.7			
2ywv (RIKEN Structural Genomics/Proteomics Initiative, unpublished work)												
*P*1	49.890	56.397	56.410	76.39	64.36	64.35	1.75	3.8	18.8/19.1			
*C*1	88.656	69.753	49.890	90.00	123.41	89.99				0.013	0.02	0.01
*C*2	88.656	69.753	49.890	90.00	123.41	90.00		2.7	15.6/19.3			
3e17 (Chen *et al.*, 2009)												
*P*1	30.177	41.325	41.292	80.05	68.63	68.52	1.75	4.9	20.8/23.7			
*C*1	61.062	55.603	30.177	90.00	119.52	90.00				0.033	0.08	0.11
*C*2	61.062	55.603	30.177	90.00	119.52	90.00		2.0	20.3/24.0			
2ddt (Ago *et al.*, 2006)												
*P*1	50.844	50.893	59.506	81.87	81.84	79.68	1.80	4.0	19.5/23.0			
*C*1	78.117	65.177	59.506	89.98	100.63	89.94				0.049	0.10	0.03
*C*2	78.117	65.177	59.506	90.00	100.63	90.00		5.3	19.3/24.4			
2ag5 (Guo *et al.*, 2006)												
*P*1	62.092	62.055	74.042	106.05	105.95	100.97	1.84	6.2	16.8/22.4			
*C*1	78.992	95.774	74.042	89.94	115.67	89.97				0.037	0.06	0.10
*C*2	78.992	95.774	74.042	90.00	115.67	90.00		5.4	19.9/24.6			
2rh0 (Joint Center for Structural Genomics, unpublished work)												
*P*1	44.568	64.504	64.423	74.51	81.06	81.22	1.95	9.6	19.2/23.5			
*C*1	102.619	78.048	44.568	89.88	101.16	89.93				0.081	0.13	0.16
*C*2	102.619	78.048	44.568	90.00	101.16	90.00		6.9	188/23.2			
2a1f (New York SGX Research Center for Structural Genomics, unpublished work)												
*P*1	77.369	79.889	79.899	94.85	96.68	96.88	2.10	6.5	21.3/26.3			
*C*1	108.105	117.667	77.369	89.87	100.05	89.99				0.010	0.01	0.20
*C*2	108.105	117.667	77.369	90.00	100.05	90.00		8.6	19.8/25.4			
4ifc (Gao, Mechin *et al.*, 2013)												
*P*1	52.320	52.260	78.980	104.47	104.63	93.01	2.13	5.2	24.3/25.6			
*C*1	71.982	75.866	78.980	90.12	111.41	89.93				0.060	0.12	0.16
*C*2	71.982	75.866	78.980	90.00	111.41	90.00		5.3	18.1/25.7			
3ox8 (Liu *et al.*, 2011)												
*P*1	60.279	68.269	68.318	70.20	84.40	84.45	2.16	6.6	18.5/23.4			
*C*1	111.749	78.538	60.279	90.05	96.82	89.96				0.049	0.07	0.05
*C*2	111.749	78.538	60.279	90.00	96.82	90.00		5.1	17.7/23.0			
4loh (Gao, Ascano *et al.*, 2013)												
*P*1	36.500	59.209	59.205	83.98	85.83	85.87	2.25	2.6	18.8/21.5			
*C*1	88.013	79.219	36.500	89.97	95.59	90.00				0.004	0.01	0.04
*C*2	88.013	79.219	36.500	90.00	95.59	90.00		2.4	16.2/22.9			
3d6e (Addington *et al.*, 2011)												
*P*1	39.551	54.813	54.910	61.38	85.72	86.10	2.40	4.3	21.3/25.5			
*C*1	94.355	56.002	39.551	90.38	94.76	89.88				0.097	0.18	0.38
*C*2	94.355	56.002	39.551	90.00	94.76	90.00		4.8	15.4/23.5			
1p7h (Giffin *et al.*, 2003)												
*P*1	74.107	80.321	80.308	71.20	78.97	78.94	2.60	6.0	23.1/26.5			
*C*1	130.608	93.506	74.107	90.03	103.63	89.99				0.013	0.02	0.03
*C*2	130.608	93.506	74.107	90.00	103.63	90.00		2.6	20.1/28.7			

al., 1997). The models were not revised manually. The 32 structures successfully solved and refined in symmetry higher than *P*1 are presented in Table 1. The table includes the unit-cell parameters from the original PDB deposition in *P*1 and after transformation by *XPREP* to the lattice corresponding to higher symmetry, with and without applying the appropriate constraints. Other parameters shown in Table 1 include the data resolution, the R_{merge} values as deposited in the PDB from processing in *P*1 and those from merging by *XPREP* in higher symmetry, as well as the R and R_{free} values quoted in the PDB after refinement in *P*1 and those obtained from refinement in higher symmetry. The maximum differences between those unit-cell lengths and angles that should be equal in higher symmetry space groups are also indicated.

Several large sets of images from protein crystals previously investigated in our laboratories (corresponding to at least 200° of total crystal rotation) were processed in the proper space group and in *P*1 symmetry with three data-processing and reduction programs: *HKL*-2000 (Otwinowski & Minor, 1997), *XDS* (Kabsch, 2010) and *MOSFLM* (Leslie, 2006). The resulting unit-cell parameters are presented in Table 2. The selected examples included crystals of two plant proteins, the transcription regulator MFT, several metal and ligand complexes of L-histidinol phosphate phosphatase (HPP; M. Ruszkowski, personal communication) and proteinase K (Wang *et al.*, 2006), as well as the fluorescent proteins LSS (Pletnev *et al.*, 2014) and RFP (S. Pletnev, personal communication).

By default, *HKL*-2000 and *XDS* estimate unit-cell parameters common to batches of three (*HKL*-2000) or ten (*XDS*) consecutive images and proceed to integrate reflection intensities with these values within each batch. However, the number of images per batch can also be changed by the user. The unit-cell parameters are therefore allowed to vary among different batches and the final unit-cell parameters are estimated from the subsequent post-

Table 2
Examples of unit-cell parameters obtained from data processing with *HKL*-2000, *XDS* and *MOSFLM* in the proper symmetry, as well as symmetry reduced to *P*1.

For each data set the structure name, data resolution, total rotation range and detector type is given on the first line, followed by the unit-cell parameters refined in *P*1 and in higher symmetry by the three data-processing programs. Maximum deviations of the unit-cell lengths $\delta(a)$ (absolute and relative) and angles $\delta(\alpha)$ between data processing in the two symmetries are given in the last three columns.

Program	Space group	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	$\delta(a)$ (Å)	$\delta(a)/a$ (%)	$\delta(\alpha)$ (°)
MFT, 1.70 Å, 360°, MAR300HS										
HKL-2000	$P4_12_12$	61.553	61.553	106.990	90	90	90			
HKL-2000	*P*1	61.577	61.565	106.990	90.007	89.991	89.932	0.024	0.04	0.068
XDS	$P4_12_12$	61.541	61.541	106.983	90	90	90			
XDS	*P*1	61.514	61.525	106.947	90.005	90.017	90.065	0.036	0.06	0.065
MOSFLM	$P4_12_12$	61.51	61.51	106.99	90	90	90			
MOSFLM	*P*1	61.44	61.58	106.97	90.02	89.93	90.00	0.07	0.11	0.07
HPP_INO, 1.50 Å, 200°, MAR300HS										
HKL-2000	$P2_1$	61.933	89.337	92.326	90	96.985	90			
HKL-2000	*P*1	61.905	89.318	92.326	89.924	97.008	90.101	0.028	0.04	0.101
XDS	$P2_1$	61.931	89.364	92.365	90	97.005	90			
XDS	*P*1	61.908	89.331	92.334	90.090	97.002	89.920	0.033	0.04	0.090
MOSFLM	$P2_1$	61.85	89.11	92.20	90	97.00	90			
MOSFLM	*P*1	61.86	89.13	92.22	89.94	97.00	90.06	0.02	0.02	0.06
HPP_Ca_INO, 1.40 Å, 250°, MAR300HS										
HKL-2000	$P2_1$	61.972	89.450	92.224	90	96.857	90			
HKL-2000	*P*1	61.968	89.430	92.208	90.156	96.863	90.041	0.020	0.02	0.156
XDS	$P2_1$	62.024	89.458	92.230	90	96.914	90			
XDS	*P*1	62.027	89.467	92.228	90.118	96.909	90.009	0.009	0.01	0.118
MOSFLM	$P2_1$	61.98	89.43	92.28	90	96.86	90			
MOSFLM	*P*1	61.95	89.38	92.12	89.83	96.85	89.98	0.16	0.02	0.17
HPP_Mg_HOL, 1.44 Å, 360°, MAR300HS										
HKL-2000	$P2_1$	61.951	89.461	92.203	90	96.928	90			
HKL-2000	*P*1	61.950	89.441	92.213	90.062	96.904	89.905	0.020	0.02	0.095
XDS	$P2_1$	62.083	89.513	92.153	90	96.752	90			
XDS	*P*1	62.092	89.528	92.163	90.025	96.750	90.009	0.015	0.02	0.025
MOSFLM	$P2_1$	61.94	89.26	92.08	90	96.90	90			
MOSFLM	*P*1	61.94	89.44	92.21	90.11	96.93	89.85	0.18	0.02	0.15
HPP_Mg_HOLP, 1.44 Å, 250°, MAR300HS										
HKL-2000	$P2_1$	61.883	88.858	92.175	90	96.910	90			
HKL-2000	*P*1	61.876	88.841	92.169	90.048	96.916	90.992	0.017	0.02	0.048
XDS	$P2_1$	61.864	88.796	92.140	90	96.918	90			
XDS	*P*1	61.859	88.784	92.128	90.045	96.917	90.004	0.012	0.01	0.045
MOSFLM	$P2_1$	61.88	88.83	92.20	90	96.87	90			
MOSFLM	*P*1	61.88	88.82	92.18	89.99	96.88	90.00	0.02	0.02	0.01
HPP_Mg_INO, 1.50 Å, 250°, MAR300HS										
HKL-2000	$P2_1$	62.158	89.459	92.361	90	96.900	90			
HKL-2000	*P*1	62.145	89.427	92.302	89.989	96.891	90.104	0.059	0.06	0.104
XDS	$P2_1$	62.090	89.519	92.161	90	96.752	90			
XDS	*P*1	62.099	89.535	92.171	90.024	96.750	90.010	0.016	0.02	0.024
MOSFLM	$P2_1$	62.02	89.16	92.13	90	96.90	90			
MOSFLM	*P*1	62.05	89.21	92.17	90.02	96.91	90.05	0.05	0.06	0.05
HPP_Mg_soakHOLP, 1.43 Å, 250°, MAR300HS										
HKL-2000	$P2_1$	61.890	89.826	92.602	90	97.071	90			
HKL-2000	*P*1	61.881	89.813	92.590	89.982	97.069	89.997	0.013	0.01	0.018
XDS	$P2_1$	61.865	89.805	92.569	90	97.057	90			
XDS	*P*1	61.841	89.769	92.532	90.033	97.056	90.003	0.037	0.04	0.033
MOSFLM	$P2_1$	61.91	89.90	92.64	90	97.07	90			
MOSFLM	*P*1	61.94	89.94	92.66	90.03	97.07	89.97	0.04	0.04	0.03
HPP_Mg_paratone, 1.50 Å, 360°, MAR300HS										
HKL-2000	$P42_12$	86.947	86.947	61.836	90	90	90			
HKL-2000	*P*1	86.903	86.872	61.790	90.023	90.120	90.095	0.075	0.09	0.095
XDS	$P42_12$	86.993	86.993	61.862	90	90	90			
XDS	*P*1	86.821	86.868	61.765	90.123	90.018	90.085	0.172	0.20	0.123
MOSFLM	$P42_12$	87.00	87.00	61.82	90	90	90			
MOSFLM	*P*1	86.91	86.92	61.78	89.87	89.94	90.02	0.09	0.10	0.13
PROTK, 1.30 Å, 330°, MAR300										
HKL-2000	$P4_32_12$	67.547	67.547	106.883	90	90	90			
HKL-2000	*P*1	67.541	67.554	106.884	90.006	89.998	89.993	0.007	0.01	0.007
XDS	$P4_32_12$	67.550	67.550	106.886	90	90	90			
XDS	*P*1	67.543	67.557	106.886	90.003	90.006	90.005	0.007	0.01	0.006
MOSFLM	$P4_32_12$	67.54	67.54	106.87	90	90	90			
MOSFLM	*P*1	67.55	67.53	106.88	90.01	90.00	90.01	0.01	0.02	0.01

refinement of all integrated data in the scaling and merging step. *MOSFLM* initially integrates reflection intensities from a few selected images with crystal spindle orientations differing by 45° and/or 90°, refines the unit-cell parameters by a post-refinement procedure and keeps these final values fixed during the subsequent integration of reflections on all images.

To compare the behavior of unit-cell parameters estimated in small batches by *HKL*-2000 and *XDS*, the 660 images obtained from a crystal of proteinase K were divided into groups of ten and each batch was separately submitted to *MOSFLM* for evaluation of its unit-cell parameters. The results of unit-cell estimation for individual batches by the three programs for proteinase K are illustrated in Fig. 2.

3. Results and discussion

During data processing, the values of the crystal unit-cell parameters are highly correlated with some other parameters, such as, for example, the crystal-to-detector distance. Their variation may also result from imperfect crystal centering on the spindle axis or nonperpendicularity of the spindle and beam directions. According to the variational principle, the use of more refinable parameters leads to better agreement between the observed and calculated reflection spot positions, resulting in their more accurate intensity evaluation, which is of primary importance at the stage of integration of intensities. Therefore, it may be beneficial to always integrate data in *P*1 symmetry, and to impose the proper crystal symmetry at the later stages of scaling and merging. However, the unit-cell parameters should, of course, eventually be estimated in the proper Laue symmetry as accurately as possible.

3.1. Variation of unit-cell parameters during data collection

To illustrate the problem of variation of unit-cell parameters during data collection, if they are estimated from small batches of consecutive diffraction

Table 2 (continued)

Program	Space group	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)	$\delta(a)$ (Å)	$\delta(a)/a$ (%)	$\delta(\alpha)$ (°)
LSS, 1.40 Å, 200°, MAR225										
HKL-2000	$P2_1$	37.436	107.366	56.602	90	102.147	90			
HKL-2000	$P1$	37.437	107.368	56.604	90.003	102.147	89.990	0.002	0.01	0.010
XDS	$P2_1$	37.433	107.341	56.602	90	102.155	90			
XDS	$P1$	37.435	107.346	56.604	90.006	102.155	90.008	0.005	0.01	0.008
MOSFLM	$P2_1$	37.42	107.31	56.61	90	102.15	90			
MOSFLM	$P1$	37.41	107.29	56.59	89.97	102.14	89.97	0.02	0.04	0.03
RFP, 1.46 Å, 200°, MAR225										
HKL-2000	$P2_12_12_1$	52.336	53.156	106.466	90	90	90			
HKL-2000	$P1$	52.337	53.155	106.466	90.006	89.997	89.985	0.001	0.01	0.015
XDS	$P2_12_12_1$	52.485	53.200	106.533	90	90	90			
XDS	$P1$	52.480	53.125	106.498	90.038	90.030	90.155	0.075	0.01	0.155
MOSFLM	$P2_12_12_1$	52.35	53.20	106.48	90	90	90			
MOSFLM	$P1$	52.36	53.20	106.47	89.99	90.02	89.98	0.01	0.02	0.02

images, the set of images measured from a crystal of proteinase K was integrated with three popular data-processing programs: *HKL*-2000 (Otwinowski & Minor, 1997), *XDS* (Kabsch, 2010) and *MOSFLM* (Leslie, 2006). These very highly accurate diffraction data led to the successful solution of the structure of proteinase K from the anomalous signal of sulfur using the short X-ray wavelength of 0.98 Å (Wang *et al.*, 2006). The set consisted of 660 images of 0.5° rotation width and the data resolution extended to 1.3 Å. The integration was performed in *P*1 without imposing the constraints of the proper space group, which is $P4_32_12$. In *DENZO* (the integrating program within the *HKL*-2000 system) the unit-cell parameters were refined together with a number of other parameters, utilizing three consecutive images, without invoking the post-refinement procedure. In *XDS* the corresponding parameters were refined and post-refined in batches consisting of ten images. These procedures are standard and are routinely used in *DENZO* and *XDS*. However, processing of data with *MOSFLM* did not use the standard procedure, since we elected to integrate and post-refine batches consisting of ten images each and then evaluated the resulting unit-cell parameters. In this program the unit-cell parameters are normally evaluated from several images recorded at a spindle-axis interval of 90°, post-refined and subsequently kept constant during integration of the whole set. We need to stress that the procedure used by us here in *MOSFLM* was nonstandard and was only applied for testing purposes.

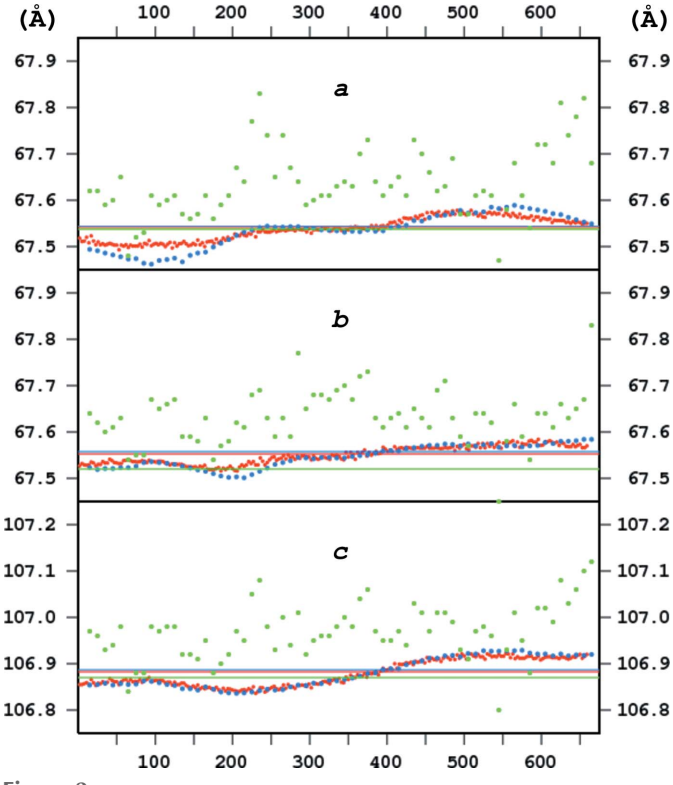


Figure 2
The unit-cell lengths a, b, c obtained from diffraction images from a $P4_32_12$ crystal of proteinase K (Wang *et al.*, 2006) processed in *P*1 symmetry by three data-processing programs. The set consisted of 660 images of 0.5° recorded on a MAR300 CCD detector at the SER-CAT beamline of the APS synchrotron with a wavelength of 0.98 Å, a crystal-to-detector distance of 150 mm and data extending to 1.3 Å resolution. The colored dots correspond to parameter values obtained by processing separate batches of three consecutive images by *HKL*-2000 (red), ten images by *XDS* (blue) and ten images by *MOSFLM* (green). The horizontal lines correspond to the values obtained from processing the whole set of images.

Inspection of Fig. 2 shows that the unit-cell parameters estimated by the three programs varied during the integration of the whole set of 660 images. In *DENZO* the maximum variation of the unit-cell lengths and angles was 0.097 Å (0.13%) and 0.05° and in *XDS* they were 0.127 Å (0.19%) and 0.84°, whereas in *MOSFLM* the variations were 0.38 Å (0.56%) and 0.19°, respectively. The variations of the unit-cell parameters were accompanied by differences in the crystal-to-detector distance of 0.15 mm (0.10%) in *DENZO*, of 0.17 mm (0.11%) in *XDS* and of 0.60 mm (0.40%) in *MOSFLM*. By default, in *HKL*-2000 and *XDS* the unit-cell parameters are output with a precision of three decimal digits, whereas in *MOSFLM* only two digits are retained, although, if the estimations of the systematic errors are provided on input, the programs give realistic uncertainties of the unit-cell parameters.

The overall, final values of the unit-cell parameters resulting from processing all of the proteinase K data in *P*1 symmetry, this time utilizing the standard *MOSFLM* procedure, are as follows: *HKL*-2000, a = 67.541, b = 67.554, c = 106.884 Å, α = 90.006, β = 89.998, γ = 89.993°; *XDS*, a = 67.543, b = 67.557, c = 106.886 Å, α = 90.003, β = 90.006, γ = 90.005°; *MOSFLM*, a = 67.55, b = 67.53, c = 106.88 Å, α = 90.01, β = 90.00, γ = 90.01°.

The comparison of the results obtained for a high-quality crystal of proteinase K which diffracted very well to almost atomic resolution and based on highly redundant data (Table 2) shows that the accuracy of the final unit-cell parameters does not exceed 0.02 Å in lengths and 0.01° in angles. This is a fairly ideal case and, if the data multiplicity is lower, as in the case of processing sets consisting of a more limited

number of images, the accuracy of the estimated unit-cell parameters may be substantially lower.

3.2. Higher symmetry structures presented in the PDB in space group *P*1

Table 1 contains information about a selected group of 32 structures presented in the PDB in space group *P*1 for which the real, higher symmetry was originally overlooked, but was eventually validated by us by merging deposited data in the appropriate symmetry, allowing successful refinement of the atomic models. The structures were not refined exhaustively here (no manual rebuilding was attempted), but the obtained statistics R and $R_{\rm free}$, as well as visual inspection of the electron-density maps, unambiguously confirmed the correctness of the reassigned space groups. None of these structures suggested the presence of pseudomerohedral twinning, which would support the treatment of these structures in lower than apparent symmetry.

If two of the unit-cell lengths and the two related unit-cell angles are approximately equal, the metric of the cell corresponds to the monoclinic *C*-centered lattice, and indeed 16 of the 32 cases summarized in Table 1 possess *C*2 symmetry. If the unit-cell parameters fulfill certain other relations the lattice may correspond to higher symmetry metrics, for example tetragonal *I*-centered or rhombohedral, as shown in some other examples in Table 1.

Fig. 3 illustrates, for the 32 investigated structures presented in the PDB in space group *P*1, the maximum differences of the unit-cell lengths between these pairs of values, which have to be equal in the true lattice metrics of these crystals. The relevant numerical values are presented in Table 1. These deviations range from a negligible amount to about 0.35%, corresponding to 0.35 Å for a crystal with a 100 Å unit-cell length. The maximum deviation of the unit-cell angles from the expected constrained value is 0.38°. The median value of discrepancies of the unit-cell lengths is 0.07% and that in the unit-cell angles is 0.09°. There is no apparent correlation between the discrepancies in the unit-cell parameters and the quality indicators for data processing or for the resulting structural model, such as resolution, $R_{\rm merge}$ or R factor.

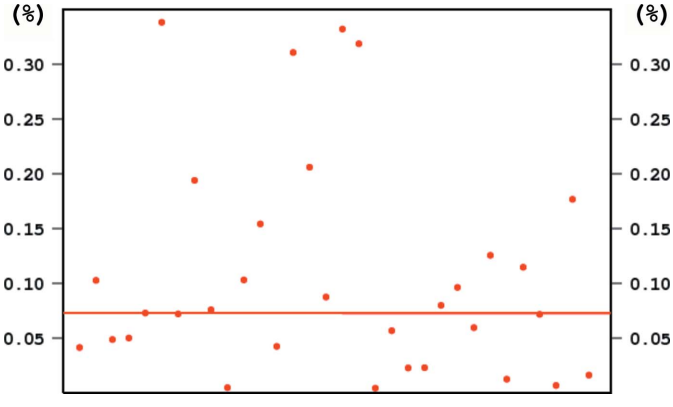


Figure 3
The maximum relative differences (normalized by the unit-cell length, as a percentage) between pairs of unit-cell lengths presented in the PDB in *P*1 symmetry, which in the true, higher symmetry, space group have to be equal. The red dots correspond to the 32 structures from the PDB in the same order of presentation as in Table 1. The horizontal line at 0.07% represents the median value.

3.3. Representation of higher symmetry structures in space group *P*1

Several sets of images from in-house projects were processed with three programs, *HKL*-2000, *XDS* and *MOSFLM*, in the proper symmetry and in space group *P*1, and the results are presented in Table 2. In each case the standard protocol for data processing was used, relying mostly on the default values of certain parameters and procedures.

The maximum deviations between the unit-cell parameters obtained in *P*1 and in the higher symmetry for the same structures are 0.08 Å (0.09%) for unit-cell lengths and 0.16° for unit-cell angles processed with *HKL*-2000, 0.17 Å (0.20%) and 0.15° for data processed with *XDS* and 0.18 Å (0.11%) and 0.17° for data processed with *MOSFLM*.

The unit-cell parameters resulting from processing diffraction data by the three programs in the proper symmetry, which was higher than *P*1, also differ somewhat. This is shown in Fig. 4, which illustrates the relative deviations of each of the three estimations of each unit-cell length from the average value (normalized by the unit-cell length) for the 11 structures listed in Table 2. The differences between the estimations of

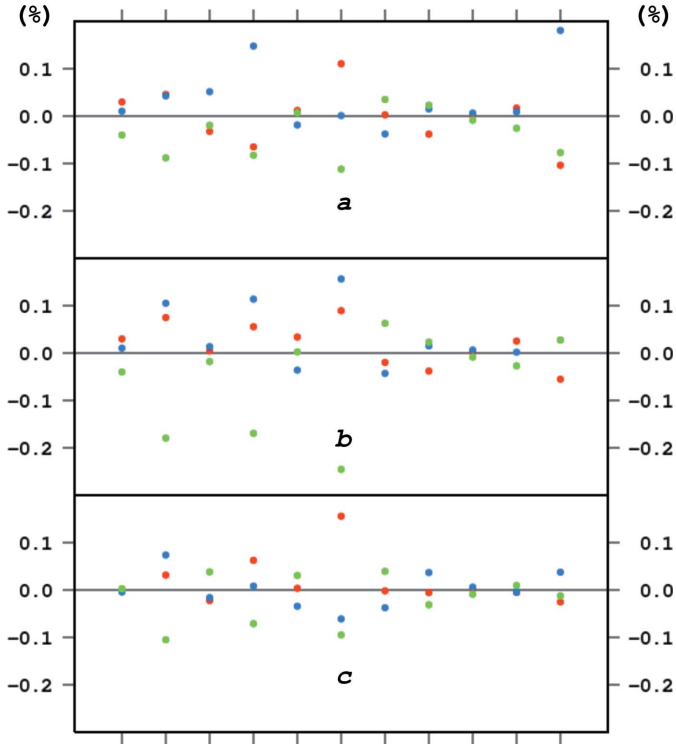


Figure 4
The relative differences between unit-cell lengths (*a*, *b*, *c*) obtained from processing 11 in-house data sets in the correct symmetry by three data-processing programs: *HKL*-2000 (red), *XDS* (blue) and *MOSFLM* (green). The horizontal line corresponds to the average value of the three estimations of each unit-cell length and the position of each dot represents the relative (normalized by the unit-cell length, as a percentage) difference from the average value. The sequence of the structures is the same as in Table 2.

the three programs are higher than 0.2% in a few cases, which is equivalent to 0.2 Å for a cell length of 100 Å.

The observed variations in unit-cell parameters, estimated from the same diffraction data by different programs and in different circumstances, obviously result from some differences in the algorithms used by the programs and from details of the protocols used by them. It has already been mentioned above that *DENZO* and *XDS* integrate reflections in batches of a few images after refining all parameters, including unit-cell parameters, whereas *MOSFLM* estimates unit-cell parameters from several images spread widely over the whole set and keeps them constant during the integration of all images, while refining other parameters in small batches. There are many more subtle differences in the detailed procedures of parameter refinement, building of standard reflection profiles, post-refinement algorithm *etc*.

3.4. Effect of radiation damage and of the utilization of multi-crystal data sets

Radiation damage had already been identified as a problem in the early days of protein crystallography. The effects resulting from the exposure of crystals of macromolecules to X-rays during diffraction data collection at ambient temperatures rapidly cause a degree of non-isomorphism, manifested by certain specific chemical and structural changes of the biological material and changes of various crystal properties, culminating in the loss of diffraction power. The introduction of cryo-techniqes alleviated this problem to some extent, but the very strong contemporary synchrotron X-ray sources introduce significant amounts of radiation damage even at cryogenic temperatures (Garman, 2003). Among various other effects, radiation damage causes changes in unit-cell parameters, usually resulting in an increase of the unit-cell volume, with an increase of up to 2% reported for some protein crystals (Ravelli & McSweeney, 2000). However, change of unit-cell parameters is not in itself a reliable metric of radiation damage, since it differs significantly between various types of crystals (Murray & Garman, 2002).

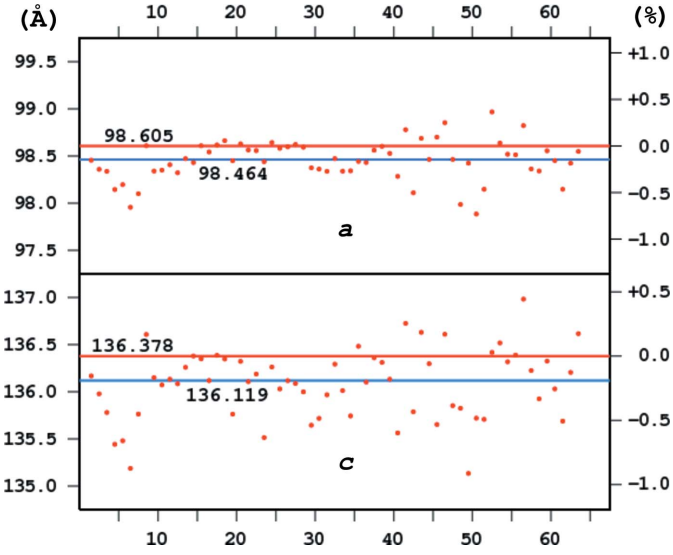


Figure 5
The *a* and *c* unit-cell lengths estimated for 63 individual crystals of an integral membrane protein with data collected in serial mode (Axford *et al.*, 2015; PDB entry 4ycr). The unit-cell lengths are marked in Å on the left side of the graph and as the percentage difference from the finally accepted overall value published by the authors (shown as a red horizontal line) on the right side. The blue line, which is significantly different from the red line, represents the median value of the parameters; the difference between these two estimates of the unit-cell length, which we cannot explain, vastly exceeds the three-digit precision of their presentation.

Radiation damage is therefore an additional cause of changes observed in the unit-cell parameters of crystals during diffraction data collection, even at moderate levels of X-ray exposure. This effect may have different magnitudes for different crystals and experimental conditions, but it contributes to the problem of the uncertainty in the evaluation of the 'final' unit-cell parameters resulting from data collection.

To prevent severe radiation damage, in the past diffraction data have frequently been measured and merged from more than a single crystal. Recently, the idea of merging data recorded in small wedges from many crystals has been revived in the form of 'serial crystallography', especially at the most intense synchrotron beamlines (Gati *et al.*, 2014) and when radiation damage has to be minimized to increase the accuracy of the weak anomalous phasing signal (Liu *et al.*, 2013). Obviously, such an approach is necessary at X-ray free-electron laser facilities, where one very small crystal can deliver only a single diffraction image before its total destruction (Chapman *et al.*, 2011).

In serial data collection, with diffraction data merged from many crystals, the evaluation of crystal unit-cell parameters constitutes a problem, since each crystal may have slightly different unit-cell parameters. This is illustrated in Fig. 5, where the unit-cell parameters *a* and *c* of an integral membrane protein crystallized with *R*3 symmetry are shown for 63 separate small batches of images recorded from 57 individual crystals (see the Supporting Information in Axford *et al.*, 2015). These unit-cell parameters vary by more than 1% from one crystal to another. Obviously the estimation of the overall, averaged values cannot be performed with high accuracy, and the values presented in the PDB deposition of this structure (PDB entry 4ycr) with three decimal digits clearly have an excessive precision.

4. Conclusions

The main aim of this paper was to draw the attention of the community to the problem of an unbiased evaluation of the expected level of accuracy and reproducibility of the unit-cell parameters for macromolecular crystals. The limitation of the accuracy of the unit-cell parameters obtained in routine macromolecular data-collection experiments may be caused by several factors, which are dependent on both the facility and the crystals. The uncertainties in the accurate values of the X-ray wavelength, crystal-to-detector distance or imperfections of detector calibration are difficult to estimate. On the other hand, macromolecular crystals are sometimes not perfectly isomorphous within their bulk volume and undergo

radiation damage. Additional problems arise if data are merged from several specimens that may differ somewhat in their lattice properties.

The discrepancies in unit-cell parameter values obtained by different programs and in different circumstances from identical diffraction images suggest that the accuracy of the unit-cell parameters estimated during routine macromolecular diffraction data processing is rarely better than ±0.1% for unit-cell lengths (approximately 0.05–0.2 Å for a typical protein crystal) and ±0.1° for unit-cell angles, although it is not easily possible to estimate the reliable values of their uncertainties in particular cases. These difficulties are indirectly acknowledged by the macromolecular crystallography community, as there is no requirement for the published values of unit-cell parameters to be accompanied by their standard uncertainties. This is in contrast to small-structure crystallography, where all quoted unit-cell parameters must be accompanied by their estimated standard deviations.

Taking into account the limitations of the practically achievable accuracy of unit-cell parameters, it is clear that unit-cell lengths and angles should be presented with the precision limited to two decimal digits at most. The excessive precision of these parameters usually quoted in publications and in the PDB is a result of an arbitrary choice of the numerical output format of numbers obtained from various mathematical procedures.

However, it may be acknowledged that inaccuracies of 0.2 Å in unit-cell lengths of 100 Å would not significantly distort the refined atomic models of macromolecules. Such a difference of 0.2% corresponds to the change of a 1.5 Å long bond by only 0.003 Å, a value comparable to the accuracy of the stereochemical restraint targets and smaller than the accuracy of atomic positions achievable in macromolecular crystallography, especially at resolutions lower than fully atomic.

Our analysis of the accuracy of unit-cell parameters led us to the investigation of the numerous structures that have been deposited in the PDB after being analyzed in space group *P*1, whereas their true symmetry was higher. In principle, any crystal structure can be expressed in *P*1 symmetry, correctly representing the positions of all atoms and the spatial interactions between them. However, the refinement of a multiplied number of parameters with the same number of observables (reflection intensities and, possibly, restraints) necessarily leads to less accurate results. Processing data in lower than correct symmetry also produces a larger number of reflections, but they are not independent even if their intensities differ somewhat owing to inaccuracies in their measurement. Merging of these genuine symmetry-related reflections will lead to more accurately estimated intensities, eventually producing a more accurate structure in the correct space group. Working in too low symmetry may also lead to computational problems with numerical singularities *etc.* It is also clear that it is easier to solve and refine a structure with one symmetrically independent molecule in, say, space group *R*32 than a structure consisting of six (or 18, depending on the choice of a rhombohedral *versus* hexagonal setting) molecules expressed in *P*1 symmetry.

It is appropriate to quote the opinion of Richard Marsh, the highly regarded expert on issues related to crystal symmetry (Marsh & Bernal, 1995):

> ...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'. [...] 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.

Acknowledgements

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

References

Addington, T., Calisto, B., Alfonso-Prieto, M., Rovira, C., Fita, I. & Planas, A. (2011). *Proteins*, **79**, 365–375.
Ago, H., Oda, M., Takahashi, M., Tsuge, H., Ochi, S., Katunuma, N., Miyano, M. & Sakurai, J. (2006). *J. Biol. Chem.* **281**, 16157–16167.
Aravind, P., Mishra, A., Suman, S. K., Jobby, M. K., Sankaranarayanan, R. & Sharma, Y. (2009). *Biochemistry*, **48**, 12180–12190.
Axford, D., Foadi, J., Hu, N.-J., Choudhury, H. G., Iwata, S., Beis, K., Evans, G. & Alguel, Y. (2015). *Acta Cryst.* D**71**, 1228–1237.
Bartonova, V., Igonet, S., Sticht, J., Glass, B., Habermann, A., Vaney, M.-C., Sehr, P., Lewis, J., Rey, F. & Kräusslich, H.-G. (2008). *J. Biol. Chem.* **283**, 32024–32033.
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.
Biswas, T., Yi, L., Aggarwal, P., Wu, J., Rubin, J. R., Stuckey, J. A., Woodard, R. W. & Tsodikov, O. V. (2009). *J. Biol. Chem.* **284**, 30594–30603.
Chapman, H. N. *et al.* (2011). *Nature (London)*, **470**, 73–77.
Chen, H., Tong, S., Li, X., Wu, J., Zhu, Z., Niu, L. & Teng, M. (2009). *Acta Cryst.* F**65**, 327–330.
Dyer, C. M., Quillin, M. L., Campos, A., Lu, J., McEvoy, M. M., Hausrath, A. C., Westbrook, E. M., Matsumura, P., Matthews, B. W. & Dahlquist, F. W. (2004). *J. Mol. Biol.* **342**, 1325–1335.
Evans, P. R. (2011). *Acta Cryst.* D**67**, 282–292.
Fernández-Fueyo, E., Ruiz-Dueñas, F. J., Martínez, M. J., Romero, A., Hammel, K. E., Medrano, F. J. & Martínez, A. T. (2014). *Biotechnol. Biofuels*, **7**, 2.
Gao, P., Ascano, M., Zillinger, T., Wang, W., Dai, P., Serganov, A. A., Gaffney, B. L., Shuman, S., Jones, R. A., Deng, L., Hartmann, G., Barchet, W., Tuschl, T. & Patel, D. J. (2013). *Cell*, **154**, 748–762.
Gao, Q., Mechin, I. *et al.* (2013). *J. Biol. Chem.* **288**, 30125–30138.
Garman, E. F. (2003). *Curr. Opin. Struct. Biol.* **13**, 545–551.
Gati, C., Bourenkov, G., Klinge, M., Rehders, D., Stellato, F., Oberthür, D., Yefanov, O., Sommer, B. P., Mogk, S., Duszenko, M., Betzel, C., Schneider, T. R., Chapman, H. N. & Redecke, L. (2014). *IUCrJ*, **1**, 87–94.
Giacovazzo, C. (2011). Editor. *Fundamentals of Crystallography*, 3rd ed., ch. 2. Oxford University Press.
Giffin, M. J., Stroud, J. C., Bates, D. L., von Koenig, K. D., Hardin, J. & Chen, L. (2003). *Nature Struct. Biol.* **10**, 800–806.
Guo, K., Lukacik, P., Papagrigoriou, E., Meier, M., Lee, W. H., Adamski, J. & Oppermann, U. (2006). *J. Biol. Chem.* **281**, 10291–10297.
Hart, J. P., Balbirnie, M. M., Ogihara, N. L., Nersissian, A. M., Weiss, M. S., Valentine, J. S. & Eisenberg, D. (1999). *Biochemistry*, **38**, 2167–2178.

Hink-Schauer, C., Estébanez-Perpiñá, E., Kurschus, F., Bode, W. & Jenne, D. E. (2003). *Nature Struct. Biol.* **10**, 535–540.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kobori, T., Sasaki, H., Lee, W. C., Zenno, S., Saigo, K., Murphy, M. E. P. & Tanokura, M. (2001). *J. Biol. Chem.* **276**, 2816–2823.
Leslie, A. G. W. (2006). *Acta Cryst.* D**62**, 48–57.
Liu, J., Chen, Y. & Ren, E. C. (2011). *Eur. J. Immunol.* **41**, 2097–2106.
Liu, Q., Liu, Q. & Hendrickson, W. A. (2013). *Acta Cryst.* D**69**, 1314–1332.
Marsh, R. E. & Bernal, I. (1995). *Acta Cryst.* B**51**, 300–307.
Motoshima, H., Inagaki, K., Kumasaka, T., Furuichi, M., Inoue, H., Tamura, T., Esaki, N., Soda, K., Tanaka, N., Yamamoto, M. & Tanaka, H. (2000). *J. Biochem.* **128**, 349–354.
Murray, J. & Garman, E. (2002). *J. Synchrotron Rad.* **9**, 347–354.
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.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. & Lamzin, V. S. (1997). *Nature Struct. Biol.* **6**, 458–463.
Pletnev, S., Shcherbakova, D. M., Subach, O. M., Pletneva, N. V., Malashkevich, V. N., Almo, S. C., Dauter, Z. & Verkhusha, V. V. (2014). *PLoS One*, **9**, e99136.
Rak, A., Pylypenko, O., Niculae, A., Pyatkov, K., Goody, R. S. & Alexandrov, K. (2004). *Cell*, **117**, 749–760.
Rakus, J. F., Kalyanaraman, C., Fedorov, A. A., Fedorov, E. V., Mills-Groninger, F. P., Toro, R., Bonanno, J., Bain, K., Sauder, J. M., Burley, S. K., Almo, S. C., Jacobson, M. P. & Gerlt, J. A. (2009). *Biochemistry*, **48**, 11546–11558.
Ravelli, R. B. G. & McSweeney, S. M. (2000). *Structure*, **8**, 315–328.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Song, H. K., Lee, J. Y., Lee, M. G., Moon, J., Min, K., Yang, J. K. & Suh, S. W. (1999). *J. Mol. Biol.* **293**, 753–761.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Wang, J., Dauter, M. & Dauter, Z. (2006). *Acta Cryst.* D**62**, 1475–1483.
Zeng, Z., Min, B., Huang, J., Hong, K., Yang, Y., Collins, K. & Lei, M. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 20357–20361.
Zhong, N., Zhang, S., Xue, F., Kang, X., Zou, P., Chen, J., Liang, C., Rao, Z., Jin, C., Lou, Z. & Xia, B. (2009). *Protein Sci.* **18**, 839–844.