



HAL
open science

MeshAndCollect: an automated multi-crystal data-collection workflow for synchrotron macromolecular crystallography beamlines

Ulrich Zander, Gleb Bourenkov, Alexander N. Popov, Daniele de Sanctis, Olof Svensson, Andrew A. Mccarthy, Ekaterina Round, Valentin I. Gordeliy, Christoph Mueller-Dieckmann, Gordon A. Leonard

► To cite this version:

Ulrich Zander, Gleb Bourenkov, Alexander N. Popov, Daniele de Sanctis, Olof Svensson, et al.. MeshAndCollect: an automated multi-crystal data-collection workflow for synchrotron macromolecular crystallography beamlines. *Acta crystallographica Section D: Structural biology* [1993-..], 2015, 71, pp.2328-2343. 10.1107/S1399004715017927 . hal-01572772

HAL Id: hal-01572772

<https://hal.science/hal-01572772>

Submitted on 8 Aug 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



MeshAndCollect: an automated multi-crystal data-collection workflow for synchrotron macromolecular crystallography beamlines

Ulrich Zander,^a Gleb Bourenkov,^b Alexander N. Popov,^a Daniele de Sanctis,^a Olof Svensson,^a Andrew A. McCarthy,^{c,d} Ekaterina Round,^{e,f,g,h,i} Valentin Gordeliy,^{e,f,g,h,i} Christoph Mueller-Dieckmann^a and Gordon A. Leonard^{a*}

Received 19 June 2015

Accepted 24 September 2015

Edited by M. Schiltz, Fonds National de la Recherche, Luxembourg

Keywords: macromolecular crystallography; synchrotron serial crystallography; multi-crystal data collection; hierarchical cluster analysis.

PDB references: thermolysin, 5a3y; lysozyme, 5a3z; bacteriorhodopsin, 5a44; 5a45; thaumatin, 5a47

Supporting information: this article has supporting information at journals.iucr.org/d

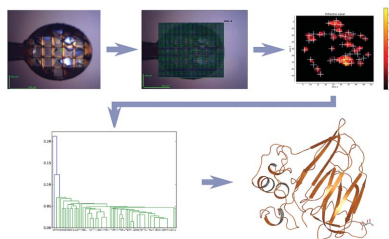
^aStructural Biology Group, European Synchrotron Radiation Facility, CS 40220, 38043 Grenoble, France, ^bEuropean Molecular Biology Laboratory, Hamburg Outstation, Notkestrasse 85, 22607 Hamburg, Germany, ^cEuropean Molecular Biology Laboratory, Grenoble Outstation, 71 Avenue des Martyrs, CS 90181, 38042 Grenoble, France, ^dUnit of Virus Host-Cell Interactions, Université Grenoble Alpes–EMBL–CNRS, 38042 Grenoble, France, ^eUniversité Grenoble Alpes, IBS, 38044 Grenoble, France, ^fCNRS, IBS, 38044 Grenoble, France, ^gCEA, IBS, 38044 Grenoble, France, ^hICS-6: Molecular Biophysics, Institute of Complex Systems (ICS), Research Centre Juelich, 52425 Juelich, Germany, and ⁱLaboratory for Advanced Studies of Membrane Proteins, Moscow Institute of Physics and Technology, Dolgoprudny 141700, Russian Federation. *Correspondence e-mail: leonard@esrf.fr

Here, an automated procedure is described to identify the positions of many cryocooled crystals mounted on the same sample holder, to rapidly predict and rank their relative diffraction strengths and to collect partial X-ray diffraction data sets from as many of the crystals as desired. Subsequent hierarchical cluster analysis then allows the best combination of partial data sets, optimizing the quality of the final data set obtained. The results of applying the method developed to various systems and scenarios including the compilation of a complete data set from tiny crystals of the membrane protein bacteriorhodopsin and the collection of data sets for successful structure determination using the single-wavelength anomalous dispersion technique are also presented.

1. Introduction

Merging partial X-ray diffraction data sets from many crystals to produce a complete data set can be traced back to the very beginnings of macromolecular crystallography (MX). Indeed, in order to cope with the radiation damage observed at room temperature, the crystal structure solution of myoglobin required the merging of partial data sets, each comprising a single precession photograph, from 22 crystals per heavy-atom derivative (Kendrew *et al.*, 1960). However, with the introduction of cryogenic data-collection techniques (Hope, 1988) the effects of radiation damage could be limited dramatically. This generally allowed the collection of complete data sets from single crystals of biological macromolecules, even at beamlines at high-intensity third-generation synchrotron sources, and this soon became the norm.

The emergence of X-ray free-electron lasers (XFELs) saw multi-crystal data collection in MX revived and taken to its logical limits. Owing to the exceptionally bright nature of XFEL beams, experimenters adopted a ‘diffraction before destruction’ approach, dubbed serial femtosecond crystallography (SFX), streaming microcrystals through the X-ray beam and collecting still diffraction images where the crystal and X-ray laser pulse coincide (Chapman *et al.*, 2011). Complete data sets are then compiled by combining data from many thousands of still diffraction images. While SFX is likely to prove a watershed in MX, chiefly because the crystal structures determined using the technique should be largely



OPEN  ACCESS

free of radiation damage (Neutze *et al.*, 2000; Boutet *et al.*, 2012), the technique is not without its disadvantages. In particular, the amount of protein required in SFX experiments is currently rather high even for structure solution based on

molecular-replacement techniques (*i.e.* only native data sets are collected). Moreover, although software for SFX data analysis is rapidly developing (Sauter *et al.*, 2013; White *et al.*, 2012, 2013; Barty *et al.*, 2014; Sawaya *et al.*, 2014), as the

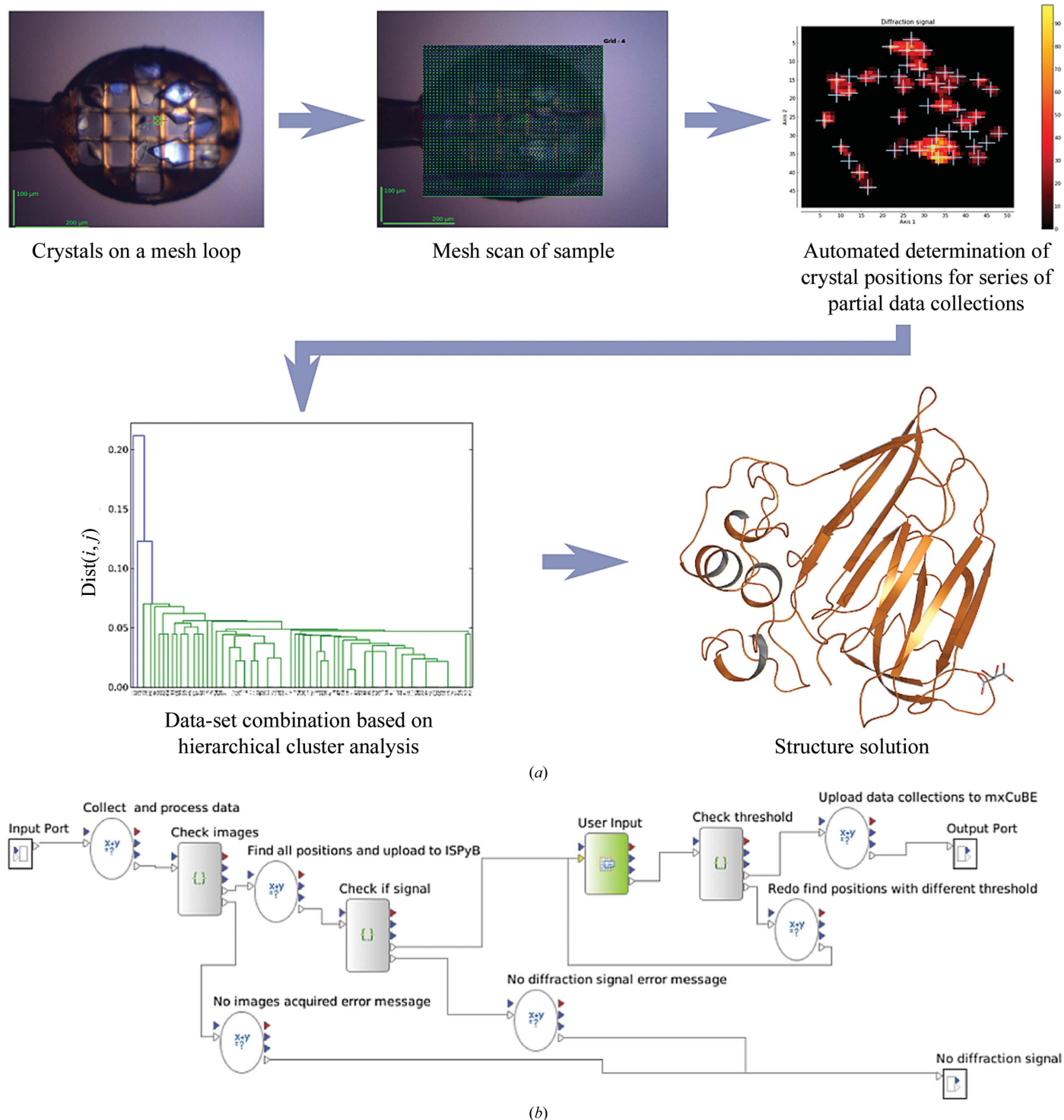


Figure 1

The *MeshAndCollect* workflow for a multicrystal data-collection method. (a) A mesh scan is performed on the sample. The resulting images are automatically inspected for protein diffraction and scored according to diffraction strength. A heat map is generated that represents the diffraction intensity, where the positions for partial data collections are marked. After the user has selected the settings for the partial data collections, the *MxCuBE2* data-collection queue is automatically filled and all partial data sets are collected. Once the partial data sets have been automatically processed, HCA can then be used to choose which data sets to merge to produce a final data set for structure solution. (b) Flow diagram of the *MeshAndCollect* workflow used.

diffraction images collected in such experiments contain predominantly partially recorded reflections measured from crystals of different sizes with laser pulses of different spectral content, estimation of the intensity (and its standard deviation) of any given reflection is problematical and data-processing methods will have to evolve significantly if the quality of SFX-collected data is to approach that currently available in 'traditional' MX experiments.

Inspired by the success of SFX, experimenters at synchrotron MX beamlines have used similar paradigms (bright X-ray beams, fast read-out detectors, small crystals, single-exposure experiments) to develop synchrotron serial crystallography (SSX), showing that it is possible to compile useful data sets from hundreds or thousands of crystals introduced into the synchrotron beam either *via* jets (Nogly *et al.*, 2015), liquid streams in glass capillaries (Stellato *et al.*, 2014), free-standing high-viscosity micro-streams (Botha *et al.*, 2015), sandwiched between two silicon nitride (Si_3N_4 ; Coquelle *et al.*, 2015) or cyclic olefin copolymer (COC; Huang *et al.*, 2015) wafers that are translated through the X-ray beam, or contained on a cryocooled sample holder (Gati *et al.*, 2014). In the latter case the whole sample mount is continuously rastered through the X-ray beam, being rotated at the same time (de Sanctis *et al.*, 2012), and diffraction images are recorded on the fly at set time intervals. As for crystals introduced into the X-ray beam in liquid streams or on silicon nitride wafers, the large majority of diffraction images collected contain no useful information. However, that fact that the sample is also rotated while being rastered means that where the crystal and the X-ray beam coincide some diffraction images could contain fully recorded reflections, thus rendering the processing and scaling of diffraction images using standard software packages relatively straightforward and improving the overall data quality. Moreover, for crystals larger than the X-ray beam, diffraction images can be grouped into those originating from the same crystal, thus also facilitating data processing and improving the resulting data quality (Gati *et al.*, 2014).

While for the same crystal volume and X-ray beam size the resolution obtainable in SSX experiments is likely to always be lower than that in SFX, SSX will become an important technique in MX. In particular, initial crystals of many systems are often small and SSX provides a means to study them without the need for the often time-consuming and cumbersome optimization of crystal size and/or quality. Indeed, when combined with the extremely bright X-rays beams that will be available at future low-emittance fourth-generation storage rings (see, for example, http://www.esrf.fr/Apache_files/Upgrade/ESRF-orange-book.pdf), such experiments may well become the norm. However, even when rastering samples contained on a cryocooled sample holder through the X-ray beam, SSX often suffers, as does SFX, from the fact that no attempt is made to synchronize the intersection of the X-ray beam and crystal during the experiment. Moreover, as the SFX 'diffraction before destruction' principle currently does not apply in SSX experiments on cryocooled samples, the amount of diffraction data collected from any given crystal is far from optimized.

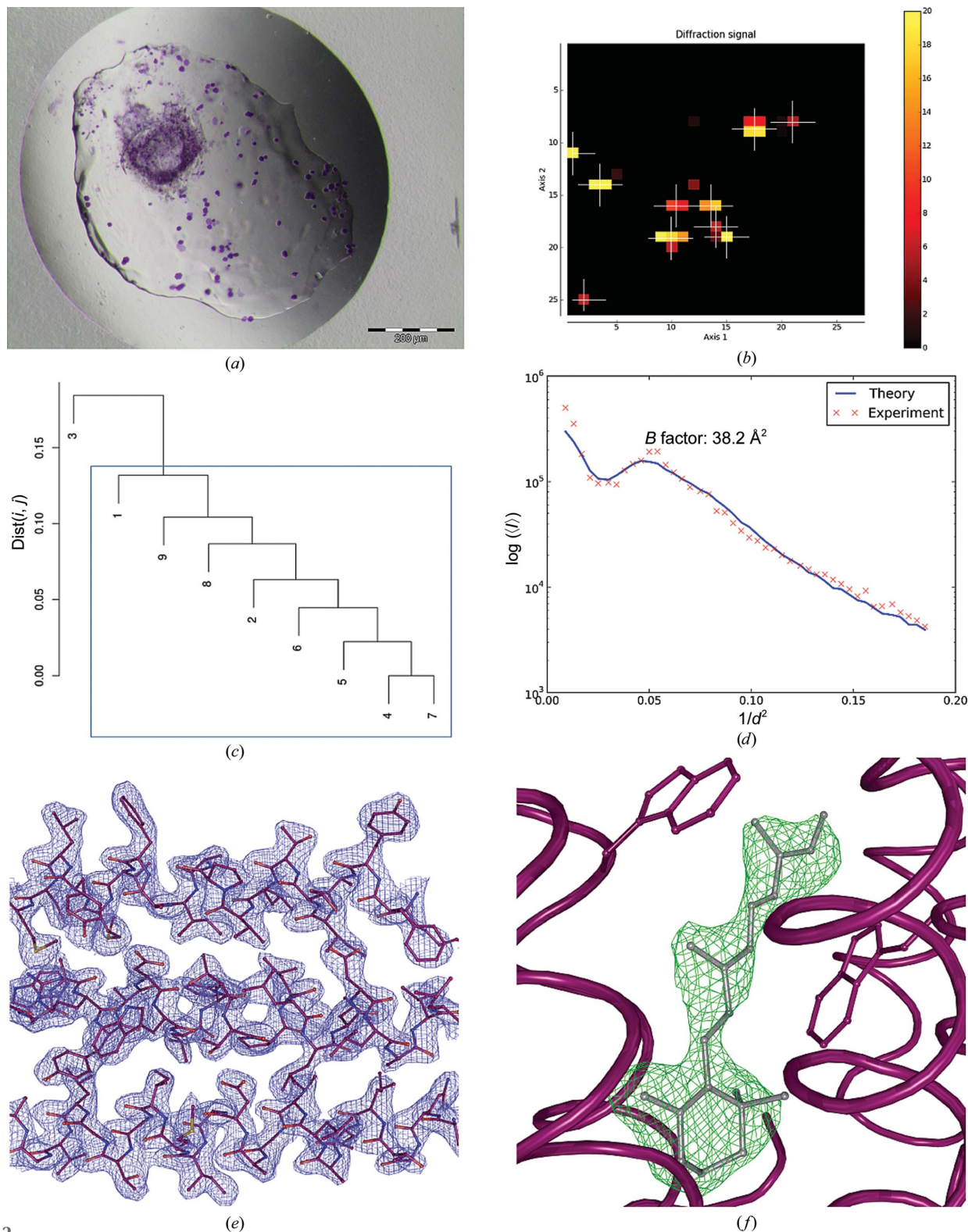
Recent developments based on either the optical (Huang *et al.*, 2015) or diffraction-based (Soares *et al.*, 2014) pre-interrogation of multi-crystal sample holders have ensured the synchronization of X-ray beam and crystals in SSX protocols and have enormously reduced the amount of sample required for a successful experiment. In a further step towards the optimal collection of diffraction data in SSX experiments from samples which can sustain the collection of many X-ray diffraction images before significant radiation damage occurs, we have developed an automatic procedure (Fig. 1). Here, the positions of many randomly oriented (micro)crystals contained in a single cryocooled sample holder are determined using an X-ray-based two-dimensional scan, the diffraction strength of each crystal found is automatically ranked and partial data sets from each crystal are collected and processed online. Subsequent manual hierarchical cluster analysis (HCA; Giordano *et al.*, 2012) is then used to decide the most correlated partial data sets to merge to produce the best quality data set for use in downstream analysis and structure solution. The protocol developed can in principle be applied to crystals mounted in almost any type of currently available mounting platform (*i.e.* nylon loops, micro-meshes, Si_3N_4 or COC wafers *etc.*) and is applicable not only to multi-crystal data collection but additionally automates multi-position data collection from large crystals when exploiting mini-focus or micro-focus X-ray beams.

As proof of the general usefulness of the protocol developed, we present the results of applying this method to various systems and scenarios. These include the compilation of a complete data set from microcrystals of the membrane protein bacteriorhodopsin, the collection and merging of partial data sets collected from different positions of larger crystals and the collection of data sets for use in structure determination using single-wavelength anomalous dispersion techniques.

2. Methods

In the experiments described here, the best results were obtained from crystals mounted in a flat sample holder (*i.e.* MiTeGen MicroMeshes; MiTeGen, USA; Fig. 1*a*), avoiding stacking of crystals and an excess of surrounding mother liquor, before either flash-cooling in liquid nitrogen or directly on the beamline. When mounted on a goniometer, the plane of the sample holder should be perpendicular to the direction of the X-ray beam. This ensures that any crystal brought into the X-ray beam will remain illuminated over a relatively small rotation range ($\pm 5^\circ$ in the experiments described here¹). To make this adjustment, we usually exploit the mini-kappa

¹ The setup of the goniometers on the ESRF beamlines on which our experiments were performed means that, once its position has been identified, each crystal is moved into the X-ray beam. Here, the movement of the crystal is *via* two motors (sampx and sampy) and the rotation axis of the goniometer is not displaced. This movement ensures the correct vertical position of the crystal but leaves open the possibility that the crystal will be misaligned in the direction of the X-ray beam. A misalignment of 10 μm in this direction will result in a misplacement of less than 1 μm over a $\pm 5^\circ$ rotation. Thus, provided that the beam is larger than 1 μm in size the crystal will remain in the X-ray beam during the data collection.


Figure 2

Multi-crystal data collection and structure solution from larger crystals of bacteriorhodopsin. (a) Crystals of bacteriorhodopsin obtained from crystallization in lipidic mesophase (Borshchevskiy *et al.*, 2011); the average crystal size is $\sim 20 \times 20 \times 5 \mu\text{m}$. (b) Heat map after initial mesh scan of the sample holder. The colours from dark red to yellow represent the intensity of the detected diffraction signal at the respective position; the white crosses mark the positions that have been used for collection of partial data sets. In all heat plots shown the x axis represents the grid points along the horizontal translation of the sample holder and the y axis the vertical grid points. For both, the unit is the beam size. (c) Dendrogram based on HCA of $CC_I(i, j)$ values produced by *XSCALE*. The blue rectangle shows the partial data sets merged to produce the final data set. (d) Wilson plot derived from the final data set using *BEST* (Bourenkov & Popov, 2006). (e) Detail of the final $2mF_{\text{obs}} - DF_{\text{calc}}, \alpha_{\text{calc}}$ electron-density map (contoured at $1.5 \times \text{r.m.s.}$) obtained, with the refined structure shown in ball-and-stick representation. (f) OMIT difference density ($mF_{\text{obs}} - DF_{\text{calc}}, \alpha_{\text{calc}}$) map at the end of the refinement procedure (contoured at $2.5 \times \text{r.m.s.}$) for a retinal molecule (ball-and-stick representation).

Table 1
Data collection and processing.

Values in parentheses are for the outer shell.

	BR1	BR2	Thaumatin	Lysozyme	Thermolysin	Maelstrom
Diffraction source	ID29, ESRF	ID29, ESRF	ID29, ESRF	ID23-1, ESRF	ID23-1, ESRF	ID23-1, ESRF
Wavelength (Å)	0.976	0.976	0.969	0.979	1.282	0.979
Temperature (K)	100	100	100	100	100	100
Detector	PILATUS3 6M	PILATUS3 6M	PILATUS3 6M	PILATUS 6M	PILATUS 6M	PILATUS 6M
Beam diameter (µm)	20	10	10	10	10	10
Flux (photons s ⁻¹)	3 × 10 ¹¹	1.5 × 10 ¹¹	8.7 × 10 ¹¹	3.5 × 10 ¹⁰	4.0 × 10 ¹⁰	9.5 × 10 ¹⁰
Absolute dose per point, two-dimensional grid (MGy)	0.013	0.16	0.016	0.085	0.17	0.34
Absolute dose per partial data set (MGy)	3.4	6.8	2.1	11.9	20.4	4.5
Partial data sets collected	10	59	100	54	96	137
Partial data sets processed	10	38	78	40	77	122
Partial data sets merged	9	10	74	21	49	45
Space group	<i>P</i> 6 ₃	<i>P</i> 6 ₃	<i>P</i> 4 ₁ 2 ₁ 2	<i>P</i> 2 ₁	<i>P</i> 6 ₂ 2	<i>H</i> 32
Unit-cell parameters (Å, °)	<i>a</i> = <i>b</i> = 61.13, <i>c</i> = 110.31	<i>a</i> = <i>b</i> = 61.25, <i>c</i> = 110.89	<i>a</i> = <i>b</i> = 57.93, <i>c</i> = 150.64	<i>a</i> = 27.58, <i>b</i> = 62.64, <i>c</i> = 59.55, β = 91.06	<i>a</i> = <i>b</i> = 92.87, <i>c</i> = 92.87	<i>a</i> = <i>b</i> = 109.95, <i>c</i> = 623.53
Resolution range (Å)	20.05–2.29 (2.41–2.29)	19.73–2.54 (2.68–2.54)	19.76–1.19 (1.26–1.19)	19.73–1.59 (1.68–1.59)	19.88–1.27 (1.33–1.27)	20.10–3.46 (3.65–3.46)
Total No. of reflections	47395	37209	4188764	96305	3330113	562987
No. of unique reflections	9802	7306	81704	23004	80728	19167
Completeness (%)	92.7 (90.6)	96.7 (81.4)	99.4 (96.1)	85.0 (82.1)	92.0 (53.4)	98.7 (94.6)
Multiplicity	4.8 (4.6)	5.1 (4.3)	51.3 (48.4)	4.2 (4.0)	41.3 (8.6)	29.4 (27.7)
Half-set correlation <i>CC</i> _{1/2}	0.998 (0.409)	0.938 (0.263)	0.999 (0.692)	0.989 (0.541)	0.998 (0.321)	0.997 (0.39)
<i>I</i> / <i>σ</i> (<i>I</i>)	13.0 (2.7)	4.8 (1.2)	48.7 (2.8)	8.0 (2.2)	25.4 (2.7)	9.5 (1.2)
<i>R</i> _{p.i.m.}	0.050 (0.548)	0.181 (0.797)	0.034 (0.503)	0.080 (0.486)	0.017 (0.344)	0.078 (0.638)
<i>B</i> factor, Wilson plot (Å ²)	37.0	26.5	9.6	12.5	14.9	88.1

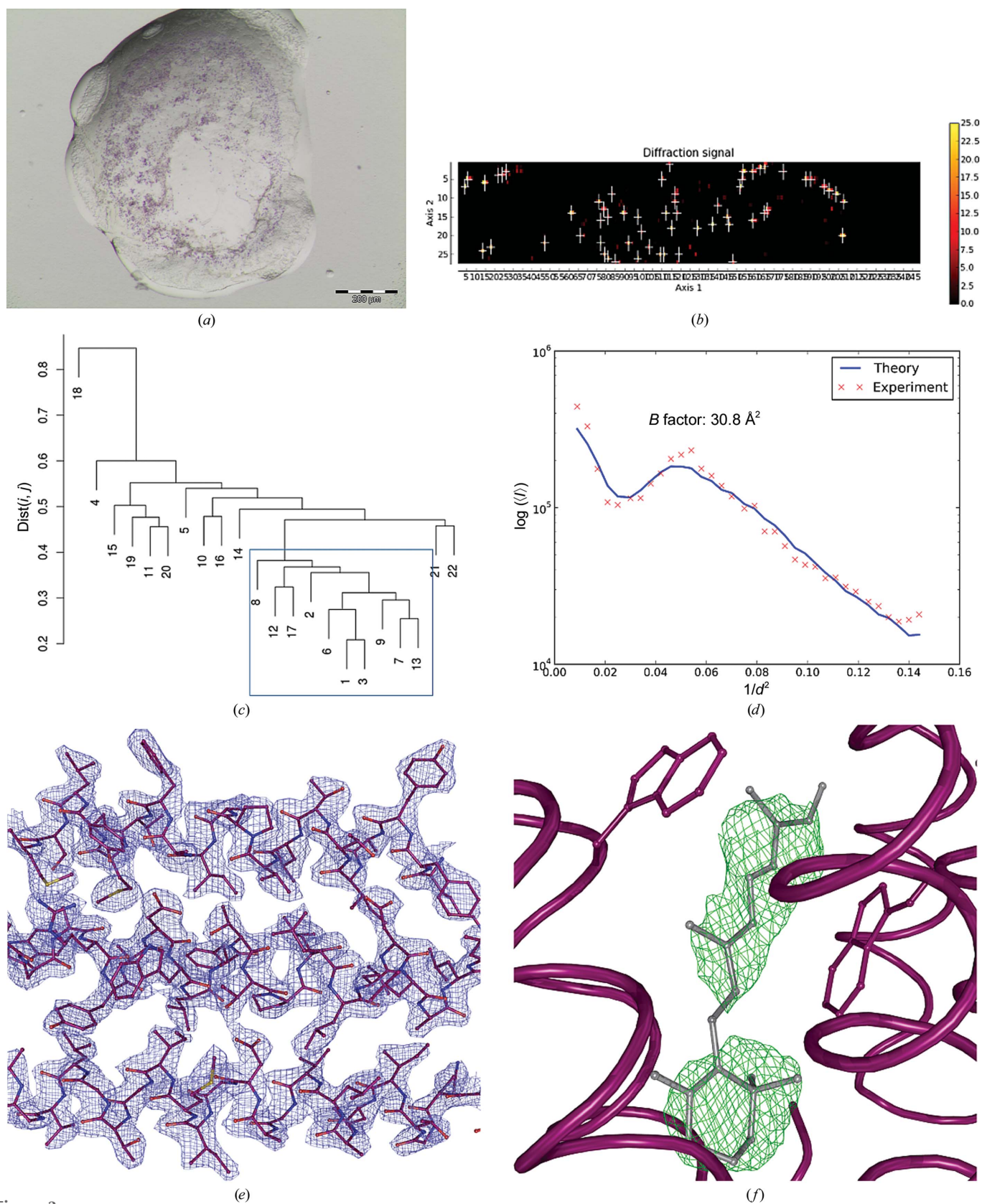
goniometers (Brockhauser *et al.*, 2013) installed on most of the MX beamlines at the ESRF. The *MeshAndCollect* protocol (Fig. 1*b*) is implemented in a customized Passerelle-EDM workflow engine (<http://isencia.be/passerelle-edm-en>) called the *Beamline Expert System* that is based on previous developments (Brockhauser *et al.*, 2012) and is embedded in the *MXCuBE2* beamline-control graphical user interface (Gabadinho *et al.*, 2010; de Sanctis & Leonard, 2014). Once the workflow has been launched the user defines the size of the X-ray beam to be used. Ideally, this should correspond to, or be smaller than, the minimum dimension of the crystals contained in the sample holder. The area over which the initial two-dimensional mesh scan is performed (Fig. 1*a*) is drawn by the user, with the dimensions of the grid and the X-ray beam size defining the number of points in the mesh scan. Diffraction images collected at each of these points are analysed on the fly for protein diffraction using the software *DOZOR* (§2.1). The user receives a heat map (Fig. 1), also stored in the ISPyB database (Delagenière *et al.*, 2011), showing the grid points at which diffraction has been observed. The user then has the possibility of adjusting the contrast level to include or exclude points for subsequent data collection. In the last experimental step partial data sets ($\pm 5^\circ$ total rotation range, 100 images per partial data set) are collected sequentially at each grid point with a *DOZOR* score above the threshold. Each partial data set is automatically processed using the *GrenADes* pipeline (Monaco *et al.*, 2013) based on *XDS* (Kabsch, 2010) running in parallel with the data collection. Partial data sets that have been successfully processed are then scaled together using *XSCALE* (Kabsch, 2010). The resulting *CC*_{*f*}(*i*, *j*) values calculated for the common unique intensities of each pair of data sets are used in a HCA protocol

(Giordano *et al.*, 2012) to produce a dendrogram (Fig. 1). This is then used to decide which partial data sets to combine to produce, using the *CCP4* programs *POINTLESS* and *AIMLESS* (Evans & Murshudov, 2013), the final data set for structure solution and refinement (Fig. 1). A feature of *POINTLESS* is that it uses the first partial data set provided as input as a reference data set. This avoids, where it might have been possible during automatic data processing, indexing ambiguities between partial data sets, with the result that the merged data set obtained is not artifactually merohedrally twinned (for a discussion of this, see Brehm & Diederichs, 2014). Any twinning then detected (*i.e.* using the ‘*H*-test’; Yeates, 1997) in the final merged data set, although an average over all crystals included, is likely to be real, facilitating determination of the true space group for use with the correct twinning fraction (if appropriate) in subsequent structure solution and refinement.

2.1. *DOZOR*

One of the core features of the protocol described here is the ability to automatically recognize and rank the series of single diffraction patterns collected during the low-dose mesh scan of the sample holder. This is carried out using the program *DOZOR*. As the algorithm used will be illustrated in more detail elsewhere, it will be only briefly described here.

In a first step, *DOZOR* determines the distribution of background intensity on a diffraction image as a function of the diffraction vector length *h*. This is accomplished by the iterative summation of pixel intensities and the sequential rejection of outliers. After azimuthal averaging this produces the one-dimensional background function $\hat{I}_{\text{background}}(h)$. This


Figure 3

Multi-crystal data collection and structure solution from microcrystals of bacteriorhodopsin. (a) Microcrystals of bacteriorhodopsin obtained from crystallization in lipidic mesophase; average crystal size $\sim 5 \times 5 \times 2 \mu\text{m}$. (b) Heat map after initial mesh scan of the sample holder. (c) Dendrogram based on HCA of $CC_{i,j}$ values produced by *XSCALE*. (d) Wilson plot from the final data set derived using *BEST*. (e) Detail of the final $2mF_{\text{obs}} - DF_{\text{calc}}$, α_{calc} electron-density map (contoured at $1.5 \times \text{r.m.s.}$), with the refined structure shown in ball-and-stick representation. (f) OMIT difference density ($mF_{\text{obs}} - DF_{\text{calc}}$, α_{calc}) map at the end of the refinement procedure (contoured at $2.0 \times \text{r.m.s.}$) for a retinal molecule (ball-and-stick representation).

Table 2
Structure solution and refinement.

Values in parentheses are for the outer shell.

	BR1	BR2	Thaumatin	Lysozyme	Thermolysin	Maelstrom
Resolution range (Å)	20.01–2.29	19.73–2.57	20.00–1.20	20.00–1.59	20.00–1.27	n.d.
No. of reflections, working set	9318	6924	77618	21804	76712	n.d.
No. of reflections, test set	480	365	3972	1182	3958	n.d.
Final R_{crist}	0.232	0.193	0.133	0.213	0.143	n.d.
Final R_{free}	0.239	0.218	0.151	0.265	0.166	n.d.
Cruickshank DPI	0.0988	0.3826	0.0311	0.1509	0.0467	n.d.
No. of non-H atoms						
Protein	1612	1677	1640	2014	2448	n.d.
Ion	—	—	—	—	5	n.d.
Ligand	95	20	10	14	31	n.d.
Water	10	—	290	181	326	n.d.
Total	1717	1697	1940	2209	2810	n.d.
R.m.s. deviations						
Bonds (Å)	0.012	0.012	0.028	0.008	0.007	n.d.
Angles (°)	1.59	1.46	2.18	1.27	1.32	n.d.
Average B factors (Å ²)						
Protein	39.2	29.5	13.7	16.9	16.5	n.d.
Ion	—	—	—	—	16.0	n.d.
Ligand	41.3	22.6	10.9	24.3	35.4	n.d.
Water	42.6	—	29.4	26.0	28.7	n.d.
Total	39.4	29.1	15.2	17.3	17.5	n.d.
Ramachandran plot						
Most favoured (%)	97.1	98.4	99.0	98.0	96.8	n.d.
Allowed (%)	2.8	1.6	1.0	2.0	3.2	n.d.
PDB code	5a44	5a45	5a47	5a3z	5a3y	

function should be smooth: any sharp peaks are an indication of ice rings or salt diffraction, and such areas are not used in further calculations.

In the case of diffraction from a crystal of a biological macromolecule, the function

$$\hat{I}_{\text{powder}}(h) = \frac{1}{N(h)} \sum_{i,j}^h I_{i,j} - \hat{I}_{\text{background}}(h), \quad (1)$$

where $N(h)$ is the number of detector pixels and $I_{i,j}$ is the intensity in any pixel which belongs to the resolution shell, h , will give the estimate of the mean intensity of Bragg spots as a function of resolution and will represent the well known Wilson plot, which for any protein crystal can be modelled using $\hat{J}_u(h)$, the unique pattern of average squared structure-factor magnitudes (Bourenkov & Popov, 2006). *DOZOR* approximates the experimental data by applying an isotropic Debye–Waller factor to the standard protein Wilson plot model,

$$\hat{I}_{\text{powder}}(h) = \text{scale} \hat{J}_u(h) \exp(-Bh^2/2). \quad (2)$$

The quality of the resulting fit is evaluated *via* the correlation coefficient between the left and right parts of (2), CC_{powder} . The program also identifies individual Bragg spots and makes a few simple geometrical checks which additionally validate the presence of diffraction from macromolecular crystals and allow the rejection of ice or salt contamination. Finally, a score of diffraction strength is estimated as the total averaged diffraction intensity multiplied by CC_{powder} , where $V(h)$ is the reciprocal volume of the resolution shell,

$$\text{score} = CC_{\text{powder}} \sum_h V(h) \text{scale} \hat{J}_u(h) \exp(-Bh/2) / \sum_h V(h). \quad (3)$$

In the case where *DOZOR* cannot find any Bragg spots, the score is determined as zero.

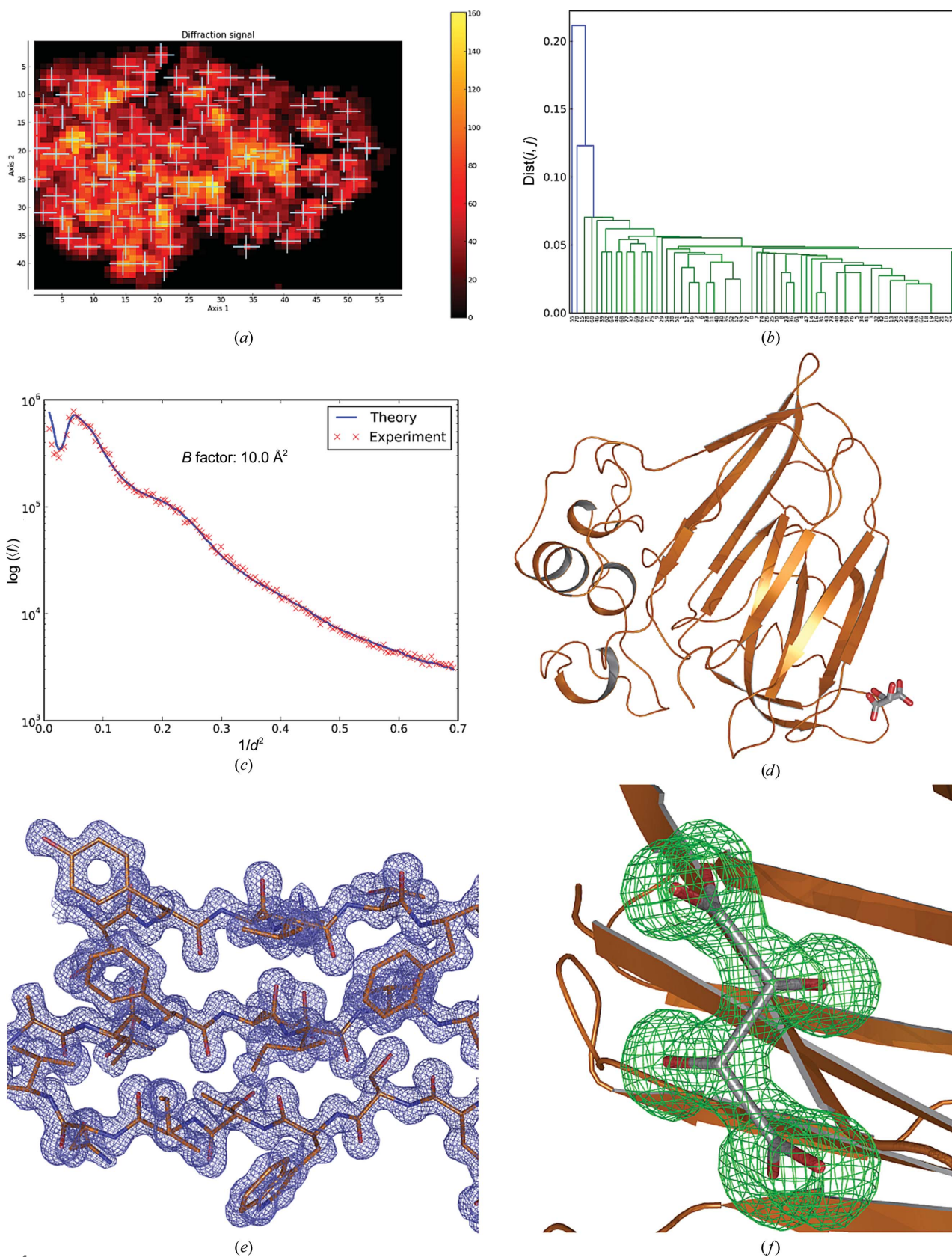
3. Results

3.1. Bacteriorhodopsin

Crystals of bacteriorhodopsin (BR) were prepared as described previously (Gordeliy *et al.*, 2003). In this study, two batches of bacteriorhodopsin crystals were used: BR1 (Fig. 2*a*), with dimensions of $\sim 20 \times 20 \times 5 \mu\text{m}$, and BR2 (Fig. 3*a*), with dimensions of $\sim 5 \times 5 \times 2 \mu\text{m}$. Diffraction data (Table 1) were collected on ESRF beamline ID29 (de Sanctis *et al.*, 2012) using a PILATUS3 6M pixel detector (Dectris, Baden, Switzerland).

For BR1 the initial mesh scan was carried out using a Gaussian X-ray beam of $20 \mu\text{m}$ in diameter with a flux of 3×10^{11} photons s^{-1} . The resulting heat map (Fig. 2*b*) revealed ten well diffracting positions from which partial data sets were collected. All partial data sets could be automatically processed and, after HCA (Fig. 2*c*), nine were chosen for scaling and merging to produce a final data set to $d_{\text{min}} = 2.3 \text{ \AA}$ (Table 1; Wilson plot shown in Fig. 2*d*).

For BR2, the initial mesh scan (X-ray beam of $10 \mu\text{m}$ in diameter with a flux of 1.5×10^{11} photons s^{-1}) produced a heat map (Fig. 3*b*) showing 59 diffracting positions in the sample holder from which partial data sets were collected. 38 partial data sets could be automatically processed and, after


Figure 4

Multi-crystal data collection and structure solution from crystals of thaumatin. (a) Heat map after initial mesh scan of the sample holder. (b) Dendrogram based on HCA of $CC_1(i, j)$ values produced by *XSCALE*. (c) Wilson plot from the final data set derived using *BEST*. (d) A ribbon diagram of the refined crystal structure of thaumatin produced (tartrate molecule in stick representation). (e) Detail of the final $2mF_{\text{obs}} - DF_{\text{calc}}$ electron-density map (contoured at $1.5 \times \text{r.m.s.}$), with the refined structure shown in ball-and-stick representation. (f) Difference density ($mF_{\text{obs}} - DF_{\text{calc}}$) for a tartrate molecule after structure refinement (OMIT map). The difference density is shown at a contour level of $3 \times \text{r.m.s.}$

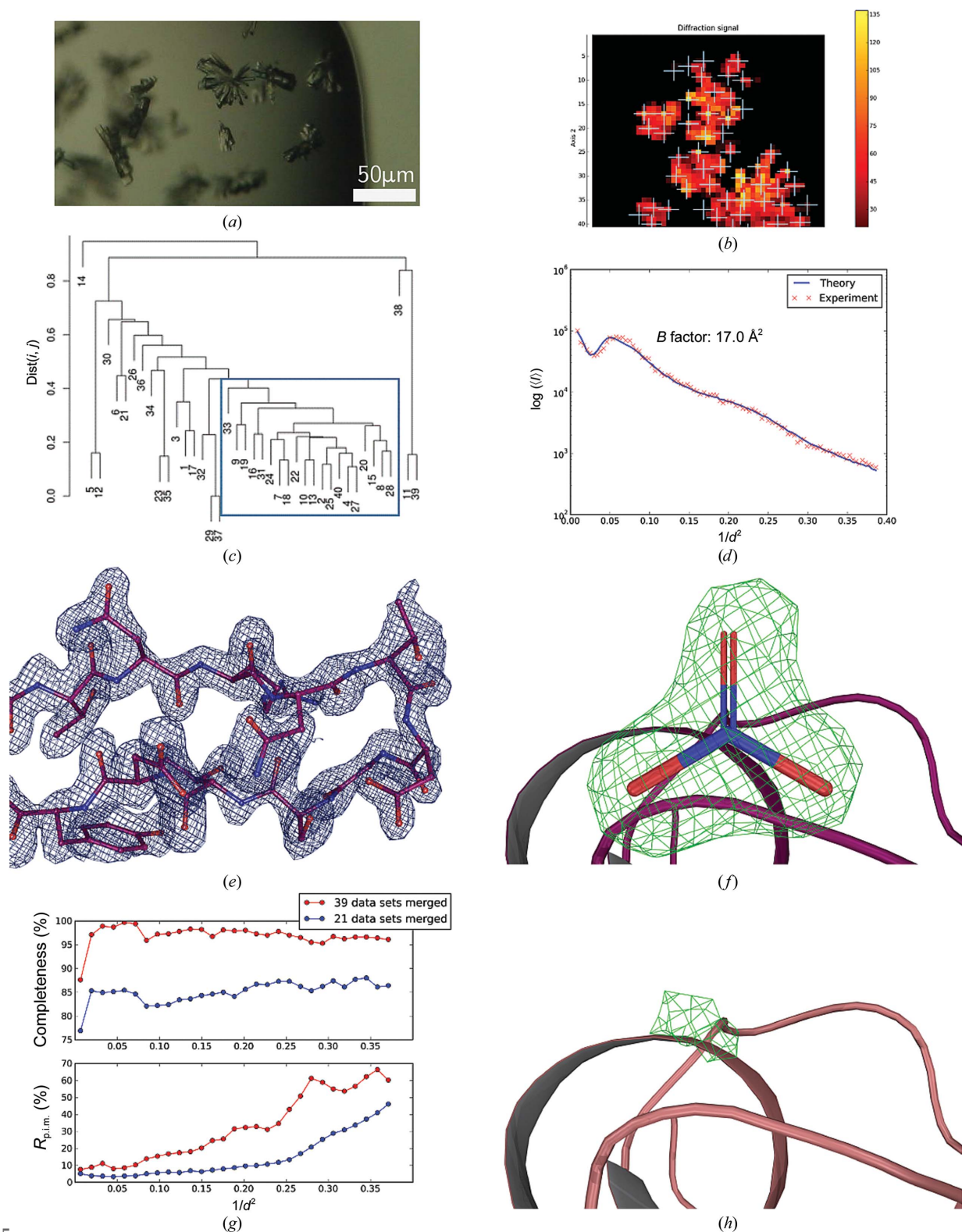


Figure 5

Multi-crystal data collection and structure solution from crystals of monoclinic lysozyme. (a) The ‘flowers’ of monoclinic lysozyme crystals produced by the crystallization procedure. (b) The heat map after an initial mesh scan of the sample used in the workflow described here. (c) Dendrogram based on HCA of $CC_2(i, j)$ values produced by *XSCALE*. (d) Wilson plot from the final data set derived using *BEST*. (e) Detail of the $2mF_{obs} - DF_{calc}$, α_{calc} electron-density map at the end of the refinement procedure (contoured at $1 \times$ r.m.s.; amino-acid residues shown in ball-and-stick representation). (f) Difference density ($mF_{obs} - DF_{calc}$, α_{calc}) for a nitrate molecule at the end of the structure-refinement procedure (OMIT map). The difference density is shown at a contour level of $3 \times$ r.m.s. (g) Plots showing comparisons of the completeness (top panel) and quality of data sets obtained following either the HCA-directed merging of data sets (21 data sets merged, blue) or the ‘blind’ merging of 39 of the 40 data sets collected. (h) Difference density ($mF_{obs} - DF_{calc}$, α_{calc}) for a nitrate molecule at the end of the structure-refinement procedure based on the data set obtained by merging 39 of the 40 data sets collected. The difference density is shown at a contour level of $2.5 \times$ r.m.s.

HCA (Fig. 3c), ten were merged to produce a final data set to $d_{\min} = 2.6 \text{ \AA}$; Table 1; Wilson plot shown in Fig. 3d).

For both BR1 (twinning fraction 0.06) and BR2 (twinning fraction 0.39) structure solution was carried out by molecular replacement using *MOLREP* (Vagin & Teplyakov, 2010) with PDB entry 3ns0 (Borshchevskiy *et al.*, 2011) stripped of water molecules and ligands as a search model. Structure refinement (Table 2) was carried out using the twinning refinement option in *REFMAC5* (Murshudov *et al.*, 2011) interspersed with rounds of manual rebuilding in *Coot* (Emsley *et al.*, 2010). In both crystal structures assignment of the retinal cofactor was possible from the interpretation of both electron-density and difference density maps and is well defined both in the final $2mF_{\text{obs}} - DF_{\text{calc}}$ electron density and in OMIT difference density maps (Figs. 2e, 2f, 3e and 3f).

3.2. Thaumatin

Thaumatin (Sigma–Aldrich catalogue No. T7638) was dissolved in double-distilled water to a concentration of 20 mg ml^{-1} . Crystals of approximate dimensions $40 \times 40 \times 60 \text{ \mu m}$ were obtained in 2 \mu l (1:1 ratio) hanging drops using 0.1 M HEPES pH 7.5, 0.7 M potassium/sodium tartrate, 20% glycerol as a reservoir. Crystals were mounted as described in §2 without further cryoprotection. Data were collected on ESRF beamline ID29. The initial mesh scan was performed with an X-ray beam of 10 \mu m in diameter with a flux of $8.7 \times 10^{11} \text{ photons s}^{-1}$. From the resulting heat map (Fig. 4a), 100 well diffracting points were chosen for the collection of partial data sets, of which 78 could be automatically integrated. After HCA (Fig. 4b) 74 were merged to produce a final data set to $d_{\min} = 1.2 \text{ \AA}$ (Table 1; Wilson plot shown in Fig. 4c).

Structure solution was carried out by molecular replacement using *MOLREP* with PDB entry 4axu (Cipriani *et al.*, 2012) stripped of water molecules and ligands as a search model. Structure refinement (Table 2, Fig. 4d), during which analysis of difference electron-density maps clearly allowed the assignment of tartrate (one molecule; Figs. 4e and 4f) and glycerol (one molecule) moieties bound to the protein, was carried out in *REFMAC5* alternated with manual rebuilding in *Coot*.

3.3. Monoclinic lysozyme

Lysozyme (Roche Applied Science, catalogue No. 10837059001) was dissolved in double-distilled water to a concentration of 40 mg ml^{-1} . ‘Flowers’ of monoclinic (space group $P2_1$) lysozyme crystals (Fig. 5a), with each petal $\sim 80 \text{ \mu m}$ in the largest dimension, were then obtained from 2 \mu l (1:1 ratio) hanging drops using 0.6 M NaNO_3 as the precipitant/reservoir. Prior to mounting, 1 \mu m 75% glycerol was added to the crystallization drop for cryoprotection. Diffraction data were collected on ESRF beamline ID23-1 (Nurizzo *et al.*, 2006) using an X-ray beam of 10 \mu m in diameter with a flux of $3.5 \times 10^{10} \text{ photons s}^{-1}$. The initial mesh scan produced a heat map (Fig. 5b) which was used as the basis for the collection of 54 partial data sets, of which 40 could be automatically processed. After HCA (Fig. 5c) 21 partial data sets were

merged to produce a final data set to $d_{\min} = 1.6 \text{ \AA}$ (Table 1; Wilson plot shown in Fig. 5d). Structure solution and refinement (Table 2, Fig. 5d) were then carried out as described above for thaumatin (using PDB entry 4axt stripped of water molecules and ligands as the search model for molecular replacement; Cipriani *et al.*, 2012), during which analysis of electron-density and difference electron density maps allowed the assignment of a nitrate (NO_3^-) ion bound to one of the lysozyme molecules in the asymmetric unit (Fig. 5f).

3.4. Thermolysin

Bacillus thermoproteolyticus thermolysin (Sigma–Aldrich catalogue No. T0331) was dissolved to 100 mg ml^{-1} in 45% DMSO, 0.05 M MES pH 6.0. The reservoir contained 35% saturated ammonium sulfate, whereas the drops were composed of the protein solution and a solution consisting of 0.05 M MES pH 6.0, 1 M NaCl, 45% DMSO in a 1:1 ratio. Rod-shaped crystals of between $40 \times 40 \times 150$ and $40 \times 40 \times 300 \text{ \mu m}$ in size were quick-soaked in 6 M trimethylamine *N*-oxide (TMAO; Mueller-Dieckmann *et al.*, 2011) for cryoprotection before mounting on a sample support (Fig. 6). Diffraction data were collected using an X-ray beam of 10 \mu m in diameter with a flux of $4.0 \times 10^{10} \text{ photons s}^{-1}$ at the peak of the Zn *K* absorption edge ($\lambda = 1.256 \text{ \AA}$) on beamline ID23-1 of the ESRF. The initial mesh scan produced a heat map (Fig. 6a) which was used as a basis for the collection of 96 partial data sets, 77 of which were automatically processed and 49 were manually merged after HCA analysis to produce a final data set to $d_{\min} = 1.37 \text{ \AA}$ (Table 1, Figs. 6b and 6c). Structure solution (Fig. 6d) was carried out using the SAD method (Dauter *et al.*, 2002) using the *SHELXC/D/E* pipeline (Sheldrick, 2008) as implemented in *HKL2MAP* (Pape & Schneider, 2004), with the initial *de novo*-obtained model of the crystal structure refined (Table 2, Fig. 6e) using iterative rounds of *REFMAC5* and manual rebuilding in *Coot*.

Our experiments with crystals of thermolysin reveal other features of the developed pipeline. In particular, when, as was the case here, the sample holder contains a series of crystals much larger than the X-ray beam (Fig. 6a) multi-crystal/multi-position data collection is also automated. Indeed, for crystals that are larger than the X-ray beam the rapid online analysis and ranking of diffraction characteristics using *DOZOR* (§2.1) provides diffraction cartographs (Bowler *et al.*, 2010) of the crystals contained on the sample mount (Fig. 6b). The workflow thus ensures that partial data sets are collected from only well diffracting areas of any given crystal.

3.5. MAEL domain of *Bombyx mori* Maelstrom

Diffraction data from crystals of the selenomethionyl derivative of the MAEL domain of *B. mori* Maelstrom (for crystallization conditions, see Chen *et al.*, 2015) were collected using an X-ray beam of 10 \mu m in diameter with a flux of $\sim 9.5 \times 10^{10} \text{ photons s}^{-1}$ at the peak of the Se *K* absorption edge ($\lambda = 0.979 \text{ \AA}$) on beamline ID23-1 at the ESRF. Crystals of this system ($20\text{--}50 \text{ \mu m}$ in the largest dimension) diffract rather poorly; therefore, in order to increase the data multiplicity to

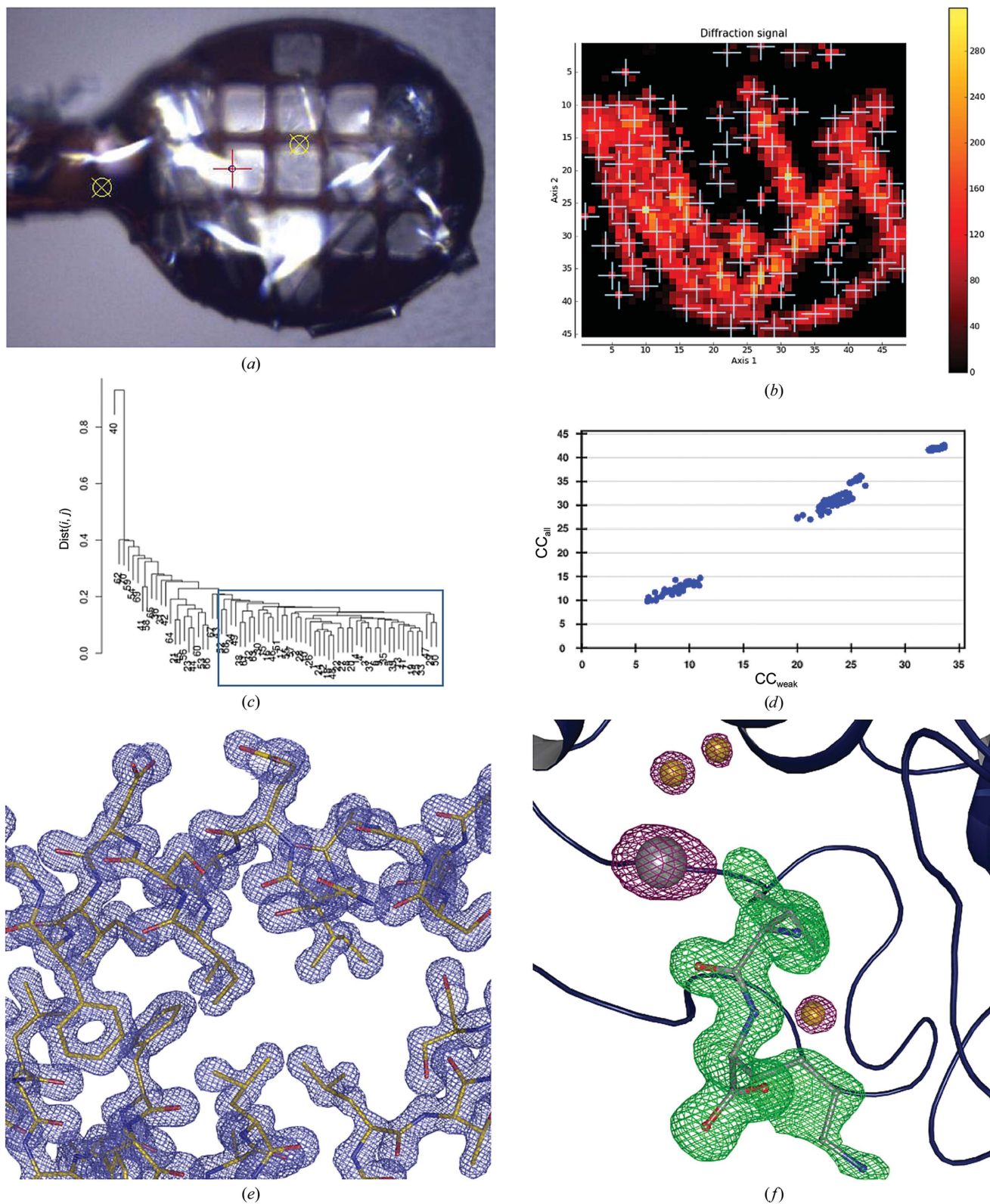


Figure 6 Multi-crystal data collection and SAD structure solution from crystals of thermolysin. (a) The sample holder mounted on ID23-1 immediately before launching the *MeshAndCollect* workflow. (b) The heat map after an initial mesh scan of the sample clearly shows the size and disposition of the crystals contained on the sample holder. (c) Dendrogram based on HCA of $CC_{i,j}$ values produced by *XSCALE*. (d) A plot of CC_{all} versus CC_{weak} from *SHELXD/HKL2MAP* for trial substructures, clearly indicating successful substructure solution. (e) Detail of the final $2mF_{obs} - DF_{calc}$, α_{calc} electron-density map at the end of the refinement procedure (contoured at $1.5 \times r.m.s.$; amino-acid residues in ball-and-stick representation). (f) Detail showing both anomalous difference map (ΔF_{ano} , $\alpha_{calc} + 90^\circ$) peaks (purple chicken wire) around the catalytic Zn^{2+} ion (grey sphere) and three Ca^{2+} ions (yellow spheres) and OMIT difference density ($mF_{obs} - DF_{calc}$, α_{calc} ; green chicken wire) in the region of a Val-Lys dipeptide found bound in the active site. The OMIT difference density is contoured at $3 \times r.m.s.$ and the anomalous difference density at $4.5 \times r.m.s.$

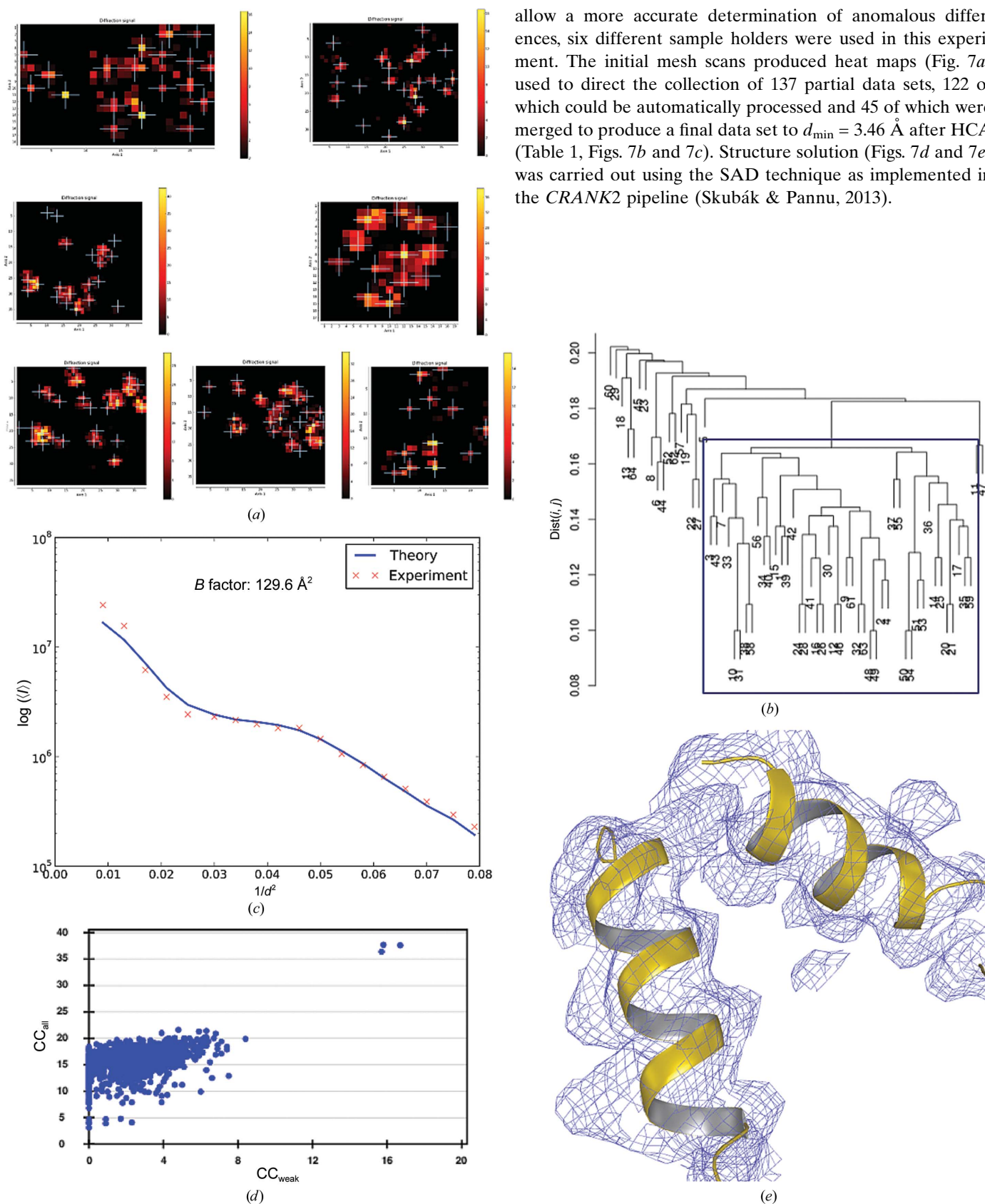


Figure 7 Multi-crystal data collection and SAD structure solution of Maelstrom. (a) Heat maps from initial mesh scans of the six sample holders analysed. (b) Dendrogram based on HCA of $CC_i(i, j)$ values produced by *XSCALE*. (c) Wilson plot from the final data set derived using *BEST*. (d) A plot of CC_{all} versus CC_{weak} from *SHELXD/HKL2MAP* for trial substructures, clearly indicating successful substructure solution. (e) Representative part of the $2mF_{obs} - DF_{calc}$, α_{calc} electron-density map after initial model building and refinement, with two α -helices shown in ribbon representation.

allow a more accurate determination of anomalous differences, six different sample holders were used in this experiment. The initial mesh scans produced heat maps (Fig. 7a) used to direct the collection of 137 partial data sets, 122 of which could be automatically processed and 45 of which were merged to produce a final data set to $d_{min} = 3.46 \text{ \AA}$ after HCA (Table 1, Figs. 7b and 7c). Structure solution (Figs. 7d and 7e) was carried out using the SAD technique as implemented in the *CRANK2* pipeline (Skubák & Pannu, 2013).

4. Discussion

4.1. General comments

The method that we describe here, while similar to the multi-crystal data-collection methods for samples mounted in micro-meshes described previously (Soares *et al.*, 2014), presents fundamental differences. Notably, a very low X-ray dose pre-screening of a sample mount is used to both identify the positions of crystals contained on the sample mount and to rank the diffraction characteristics of the crystals in order to create a priority for the subsequent automatic collection of partial data sets, and a HCA protocol is used to choose which partial data sets to merge to produce the best final data set. Moreover, when the sample holder contains a series of crystals much larger than the X-ray beam the method also automates the type of multi-crystal/multi-position data collection (Riekell *et al.*, 2005) that has become essential in the structural study of G protein-coupled receptors (GPCRs; Rasmussen *et al.*, 2011; Hollenstein *et al.*, 2013; Lebon *et al.*, 2011). Furthermore, for crystals larger than the X-ray beam the rapid online analysis and ranking of diffraction characteristics using *DOZOR* (§2.1) also provides diffraction cartographs (Bowler *et al.*, 2010) of the crystals contained on the sample mount, ensuring that partial data sets are only collected from well diffracting areas of any given crystal.

To demonstrate the general applicability of the workflow described here, we have applied it to various systems and scenarios in which many crystals of the same type are mounted on the same cryocooled sample holder. In all of the cases presented our workflow has yielded data sets that are fit for purpose (Table 1, §4.2). As might be expected (Fry *et al.*, 1996), the protocol described here is particularly amenable to systems (*i.e.* thaumatin, bacteriorhodopsin, Maelstrom) that crystallize in high-symmetry space groups. However, our experiments using monoclinic crystals of lysozyme show that the method can also be applied to low-symmetry systems. Furthermore, as the monoclinic form of lysozyme crystallized as clumps of intergrown crystals (Fig. 5*a*), the success of this latter experiment demonstrates that the protocol developed also automates the collection of diffraction data using mini-focus or micro-focus X-ray beams under conditions where mounting single crystals of a particular sample may prove to be difficult or impossible.

It is worth noting that the completeness of the data set obtained for monoclinic lysozyme following the HCA-directed merging of the partial data set collected is rather incomplete (21 of 40 automatically processed partial data sets merged, 85% completeness; Table 1). However, this is not the result of a combination of low-symmetry crystals lying in preferred orientations in the sample holder. Indeed, merging 39 of the 40 automatically processed partial data sets greatly improves the completeness (Fig. 5*g*). However, the quality of the resulting data set is seriously degraded compared with that obtained by merging only partial data sets in the main HCA cluster (Fig. 5*g*). Moreover, in contrast to what is observed following HCA-directed merging, the resulting difference electron density does not allow the proper identification of nitrate ions

bound to the protein (Figs. 5*f* and 5*h*). It is thus clear that HCA is an indispensable tool for the proper merging of partial data sets. Nevertheless, that the merged data set for monoclinic lysozyme obtained following HCA is somewhat incomplete suggests, for some low-symmetry systems at least, that data collection from samples in two loops with different orientations in the X-ray beam may be required to ensure a fully complete, high-quality data set.

The examples that we present include the compilation of complete diffraction data from partial data sets collected from a series of microcrystals ($\sim 5\ \mu\text{m}$ in the largest dimension), contained on the same sample holder, of a membrane protein (bacteriorhodopsin) grown in lipidic mesophase. Such mesophases are very important media for the growth of membrane-protein crystals (Gordeliy *et al.*, 2003), but are often opaque in nature, particularly when cooled. It can thus be challenging to identify, mount and centre in the X-ray beam small crystals produced in such media. That the workflow described here uses diffraction-based methods to identify the positions of crystals in a sample holder is clearly a major advantage in such cases as it obviates such problems, particularly when entire crystallization drops are harvested, by automating the collection of partial data sets from multiple crystals.

4.2. Structure solution and refinement

4.2.1. Diffraction data for structure solution by molecular replacement. The examples of bacteriorhodopsin (BR1 and BR2), thaumatin and monoclinic lysozyme described above clearly show that the protocol that we have developed yields, even for very small crystals, complete diffraction data sets that allow structure solution by MR. Moreover, despite the fact that all data sets were obtained by the merging of multiple partial data sets, electron-density ($2mF_{\text{obs}} - DF_{\text{calc}}$, α_{calc}) and difference density ($mF_{\text{obs}} - DF_{\text{calc}}$, α_{calc}) maps calculated during structure refinement clearly allow the identification of moieties not included in the MR search models: retinal (BR1 and BR2; Figs. 2*f* and 3*f*), tartrate (thaumatin; Fig. 4*f*) and NO_3^- (monoclinic lysozyme; Fig. 5*f*). This suggests that the method developed may, in the future, have a significant role to play in projects aimed at fragment screening (Murray & Blundell, 2010) as an aid in drug design. Traditionally, such projects are based around the production of relatively large, robust crystals for use in soaking experiments (Oster *et al.*, 2015). However, the results presented here show that this clearly does not need to be the case and that complete, high-quality data sets could straightforwardly be compiled from a series of smaller crystals mounted on the same sample holder. Moreover, as evidence suggests that smaller crystals require reduced fragment/ligand-soaking times to obtain the same occupancy of the fragment/ligand in crystal structures (Cole *et al.*, 2014), microcrystal-based fragment screening experiments may well become the norm, with soaking times based on the largest crystal contained in the crystallization drop ensuring the maximum occupancy of ligands/fragments in all of the crystals mounted on a single sample loop.

4.2.2. Diffraction data for structure solution exploiting anomalous scattering. In order to demonstrate the possibilities of the workflow presented here to produce data suitable for experimental phasing techniques that exploit anomalous scattering, two different systems were investigated. The first of these, thermolysin, contains one catalytic Zn^{2+} ion and three Ca^{2+} ions per protein chain (316 residues), producing a theoretical anomalous diffraction ratio ($\langle \Delta F/F \rangle$) of $\sim 2\%$ for data collected at the peak of the Zn *K* absorption edge. The second, the selenomethionyl derivative of the MAEL domain of *B. mori* Maelstrom (Chen *et al.*, 2015), produces a theoretical anomalous diffraction ratio of 4.0% for data collected at the peak of the Se *K* absorption edge. However, the crystals of this system diffract rather poorly (see Table 1). The collection of data of sufficiently high quality for the structure solution of both systems is thus clearly challenging, even from single crystals. Nevertheless, as can be seen in Figs. 6 and 7, for both systems our multi-crystal workflow clearly yields diffraction data of sufficient quality for structure solution. As might be expected, a high data multiplicity was important in both cases (Table 1) and to achieve this for Maelstrom required combining partial data sets from crystals mounted on six different sample holders (Fig. 7a).

4.3. Perspectives

We have developed an automatic procedure to locate, rank the diffraction characteristics of and collect partial data sets from large numbers of crystals contained on the same sample holder. Subsequent HCA of the partial data sets collected then allows the choice of which partial data sets to merge to produce a final data set for downstream structure solution and refinement. Compared with previously presented SSX protocols (Gati *et al.*, 2014; Nogly *et al.*, 2015; Stellato *et al.*, 2014), *MeshAndCollect* has several advantages, notably that small but contiguous data sets can, if desired, be collected from all crystals contained on the sample holder. Crystal wastage is thus not an issue, data reduction from raw diffraction images to structure factors and standard deviations is comparably straightforward and the quality of the final data set is improved. Moreover, the experiments described in §3 clearly demonstrate the capability of *DOZOR* to detect diffraction signal in low-dose two-dimensional mesh scans even for the smallest crystals (BR2; §3.1) studied in this work, which had an average volume of $\sim 50 \mu\text{m}^3$.

When starting this work, we presumed that cryocooled crystals contained on the same loop would be relatively isomorphous as all crystals are from the same crystallization drop and subject to similar handling during mounting and cryocooling (Giordano *et al.*, 2012). The dendrograms shown in Figs. 2, 3, 4, 5 and 6 suggest that this is the case, although in several of our examples many of the partial data sets collected are not used to construct the final result. Most of the above histograms contain one main cluster with high mutual correlation coefficients and a continuum of data sets with decreasingly low correlation to the main cluster. Such a pattern is indicative of strongly varying data quality between

partial data sets rather than crystal non-isomorphism and suggests that some partial data sets were collected from positions with overlapping crystal lattices or other issues such as crystal damage. Clearly, the evaluation of initial two-dimensional mesh scans with *DOZOR* did not filter such positions out. Furthermore, with only a 10° rotation range measured at each position it is difficult to detect such problematic data sets on the basis of their internal processing statistics, and HCA is required to filter out the worst partial data sets. In the case of Maelstrom, where partial data sets were measured from crystals on several different sample mounts, the dendrogram (Fig. 7) shows well populated clusters above a cutoff of $\text{dist}(i, j) = 0.15$ and a continuum of poorly correlated data sets below this cutoff. This suggests that both non-isomorphism and variation in data quality between partial data sets is present. However, as can be seen, both poor-quality and non-isomorphous partial data sets are successfully filtered by the HCA procedure.

Despite the success of the experiments described above, the procedure developed will eventually be improved in many areas. Here, all samples were mounted and cryocooled manually, and it may be that better results can be achieved by taking advantage of robotic crystal-handling methods both for the removal of mother liquor from the crystallization drop and the mounting and cryocooling of crystals in a suitable sample holder (Cipriani *et al.*, 2012). Moreover, for the different experiments described here the total absorbed doses per crystal (Table 1; calculated post-experiment using *RADDOSE*; Paithankar & Garman, 2010) are rather low compared with the Henderson/Garman limits (Henderson, 1990; Owen *et al.*, 2006) generally used in diffraction data collection from cryocooled single crystals of macromolecules. In future versions of the pipeline presented here, following the low-dose two-dimensional mesh scan the optimum total exposure time per crystal (partial data set) will be calculated before the data-collection step using the *EDNA* characterization software (Bourenkov & Popov, 2010; Incardona *et al.*, 2009), the result being better quality and/or higher resolution data collected per crystal. For crystals that are highly radiation-sensitive one might even imagine the use of a 'Burn Strategy' workflow (Leal *et al.*, 2011) to provide a precise estimation of the maximum allowable total absorbed dose per crystal.

As the *EDNA* procedure implies the indexing of diffraction patterns (Incardona *et al.*, 2009), comparison, for crystals larger than the X-ray beam, of orientation matrices will allow either the pre-clustering of partial data sets collected from different points on the same crystal or the measurement of crystal size and alignment in the sample holder. In the latter case this information could be used to automatically guide helical data collections (Flot *et al.*, 2010; de Sanctis *et al.*, 2012) that, provided that diffraction is homogenous, may allow the collection of complete data sets from each of the different crystals contained in the sample holder. For crystals of a similar or smaller size than the X-ray beam prior knowledge of the crystal orientation in the X-ray beam will allow a broader range of experiments than is currently the case. In particular, the order of the collection of partial data sets could be

constructed to ensure the compilation of a complete data set when only a few crystals are available or to ensure the collection of as highly redundant data as possible. Finally, for sample mounts containing many small robust, well diffracting crystals one can also imagine a modification to the pipeline in which complete diffraction data sets for structure solution and subsequent refinement are collected from all crystals contained in the sample holder. Separating such data sets into different clusters would result in ensembles of crystal structures for each target.

Once data collection and processing have been completed, a final improvement to the pipeline is in the choice of partial data sets to merge to produce a final data set. This choice clearly depends on the aim of the experiment in hand (*i.e.* structure solution by molecular replacement, *de novo* structure solution using SAD *etc.*), and in principle is best made using HCA based on $CC_i(i, j)$ (§2; Giordano *et al.*, 2012). However, for partial data sets from low-symmetry crystals the number of common unique reflections for each pair of data sets may be low, thus leading to artefacts, and a better approach may be to combine HCA with the type of ‘scale-and-merge’ algorithms currently implemented in the *PHENIX* package (Adams *et al.*, 2010; https://www.phenix-online.org/version_docs/dev-1977/reference/scale_and_merge.html) or recently described for other SSX protocols (Huang *et al.*, 2015).

5. Conclusions

We have presented here a pipeline for the routine collection of partial diffraction data sets from many randomly oriented crystals of the same biological macromolecule contained in a single cryocooled sample holder. The major advantages of the pipeline developed are (i) that it can be applied to crystals mounted in almost any available sample holder suitable for cryocooling, thus rendering the methodology available to the widest possible range of potential users, (ii) that the positions of all well diffracting crystals are determined and that their diffraction strength is ranked prior to data collection, (iii) that small, but contiguous, partial data sets are collected from as many crystals contained in the sample holder as is desired and (iv) that HCA is used to choose partial data sets for merging to produce the best possible data set for downstream analysis and structure solution. As described above, the protocol developed can be applied to both SSX-type experiments involving microcrystals and to multi-position data collection from crystals larger than the X-ray beam size. The results presented here suggest that the method developed will be useful in all areas of macromolecular crystallography, including the compilation of a complete data set from many very small crystals ($\sim 5\ \mu\text{m}$ in the largest dimension), in structure determination exploiting anomalous scattering and in projects aimed at rational drug design.

While we have confined our experiments to crystals mounted on cryocooled sample holders, there is no reason, providing that the increased radiation damage is taken into account, that the automated screening and data-collection

procedure developed cannot also be applied at room temperature, particularly in experiments that involve *in situ* screening and data collection (Axford *et al.*, 2012; Jacquamet *et al.*, 2004; le Maire *et al.*, 2011; Huang *et al.*, 2015). Moreover, *MeshAndCollect* should also be extendable to structure solution based on radiation damage-induced phasing (RIP; Ravelli *et al.*, 2003; de Sanctis & Nanao, 2012) or SAD experiments based on inverse-beam protocols (González, 2003).

Acknowledgements

The authors would like to gratefully acknowledge the allocation of beamtime (‘SSX BAG’) for this and other ongoing SSX projects at ESRF by the ESRF MX Beamtime Allocation Panel.

References

- Adams, P. D. *et al.* (2010). *Acta Cryst.* **D66**, 213–221.
 Axford, D. *et al.* (2012). *Acta Cryst.* **D68**, 592–600.
 Barty, A., Kirian, R. A., Maia, F. R. N. C., Hantke, M., Yoon, C. H., White, T. A. & Chapman, H. (2014). *J. Appl. Cryst.* **47**, 1118–1131.
 Borshchevskiy, V. I., Round, E. S., Popov, A. N., Büldt, G. & Gordeliy, V. I. (2011). *J. Mol. Biol.* **409**, 813–825.
 Botha, S., Nass, K., Barends, T. R. M., Kabsch, W., Latz, B., Dworkowski, F., Foucar, L., Panepucci, E., Wang, M., Shoeman, R. L., Schlichting, I. & Doak, R. B. (2015). *Acta Cryst.* **D71**, 387–397.
 Bourenkov, G. P. & Popov, A. N. (2006). *Acta Cryst.* **D62**, 58–64.
 Bourenkov, G. P. & Popov, A. N. (2010). *Acta Cryst.* **D66**, 409–419.
 Boutet, S. *et al.* (2012). *Science*, **337**, 362–364.
 Bowler, M. W., Guijarro, M., Petitdemange, S., Baker, I., Svensson, O., Burghammer, M., Mueller-Dieckmann, C., Gordon, E. J., Flot, D., McSweeney, S. M. & Leonard, G. A. (2010). *Acta Cryst.* **D66**, 855–864.
 Brehm, W. & Diederichs, K. (2014). *Acta Cryst.* **D70**, 101–109.
 Brockhauser, S., Ravelli, R. B. G. & McCarthy, A. A. (2013). *Acta Cryst.* **D69**, 1241–1251.
 Brockhauser, S., Svensson, O., Bowler, M. W., Nanao, M., Gordon, E., Leal, R. M. F., Popov, A., Gerring, M., McCarthy, A. A. & Gotz, A. (2012). *Acta Cryst.* **D68**, 975–984.
 Chapman, H. N. *et al.* (2011). *Nature (London)*, **470**, 73–77.
 Chen, K. M., Campbell, E., Pandey, R. R., Yang, Z., McCarthy, A. A. & Pillai, R. S. (2015). *RNA*, **21**, 833–839.
 Cipriani, F., Röwer, M., Landret, C., Zander, U., Felisaz, F. & Márquez, J. A. (2012). *Acta Cryst.* **D68**, 1393–1399.
 Cole, K. *et al.* (2014). *PLoS One*, **9**, e101036.
 Coquelle, N., Brewster, A. S., Kapp, U., Shilova, A., Weinhausen, B., Burghammer, M. & Colletier, J.-P. (2015). *Acta Cryst.* **D71**, 1184–1196.
 Dauter, Z., Dauter, M. & Dodson, E. J. (2002). *Acta Cryst.* **D58**, 494–506.
 Delagenière, S. *et al.* (2011). *Bioinformatics*, **27**, 3186–3192.
 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
 Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* **D69**, 1204–1214.
 Flot, D., Mairs, T., Giraud, T., Guijarro, M., Lesourd, M., Rey, V., van Brussel, D., Morawe, C., Borel, C., Hignette, O., Chavanne, J., Nurizzo, D., McSweeney, S. & Mitchell, E. (2010). *J. Synchrotron Rad.* **17**, 107–118.
 Fry, E., Logan, D. & Stuart, D. (1996). *Crystallographic Methods and Protocols*, edited by C. Jones, B. Mulloy & M. Sanderson, pp. 319–363. Totowa: Humana Press.
 Gabadinho, J. *et al.* (2010). *J. Synchrotron Rad.* **17**, 700–707.
 Gati, C., Bourenkov, G., Klinge, M., Rehders, D., Stellato, F., Oberthür, D., Yefanov, O., Sommer, B. P., Mogk, S., Duszhenko,

- M., Betzel, C., Schneider, T. R., Chapman, H. N. & Redecke, L. (2014). *IUCrJ*, **1**, 87–94.
- Giordano, R., Leal, R. M. F., Bourenkov, G. P., McSweeney, S. & Popov, A. N. (2012). *Acta Cryst. D***68**, 649–658.
- González, A. (2003). *Acta Cryst. D***59**, 1935–1942.
- Gordeliy, V. I., Schlesinger, R., Efremov, R., Büldt, G. & Heberle, J. (2003). *Methods Mol. Biol.* **228**, 305–316.
- Henderson, R. (1990). *Proc. R. Soc. B Biol. Sci.* **241**, 6–8.
- Hollenstein, K., Kean, J., Bortolato, A., Cheng, R. K., Doré, A. S., Jazayeri, A., Cooke, R. M., Weir, M. & Marshall, F. H. (2013). *Nature (London)*, **499**, 438–443.
- Hope, H. (1988). *Acta Cryst. B***44**, 22–26.
- Huang, C.-Y., Olieric, V., Ma, P., Panepucci, E., Diederichs, K., Wang, M. & Caffrey, M. (2015). *Acta Cryst. D***71**, 1238–1256.
- Incardona, M.-F., Bourenkov, G. P., Levik, K., Pieritz, R. A., Popov, A. N. & Svensson, O. (2009). *J. Synchrotron Rad.* **16**, 872–879.
- Jacquamet, L., Ohana, J., Joly, J., Borel, F., Pirocchi, M., Charrault, P., Bertoni, A., Israel-Gouy, P., Carpentier, P., Kozielski, F., Blot, D. & Ferrer, J.-L. (2004). *Structure*, **12**, 1219–1225.
- Kabsch, W. (2010). *Acta Cryst. D***66**, 125–132.
- Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. & Shore, V. C. (1960). *Nature (London)*, **185**, 422–427.
- Leal, R. M. F., Bourenkov, G. P., Svensson, O., Spruce, D., Guijarro, M. & Popov, A. N. (2011). *J. Synchrotron Rad.* **18**, 381–386.
- Lebon, G., Warne, T., Edwards, P. C., Bennett, K., Langmead, C. J., Leslie, A. G. W. & Tate, C. G. (2011). *Nature (London)*, **474**, 521–525.
- Maire, A. le, Gelin, M., Pochet, S., Hoh, F., Pirocchi, M., Guichou, J.-F., Ferrer, J.-L. & Labesse, G. (2011). *Acta Cryst. D***67**, 747–755.
- Monaco, S., Gordon, E., Bowler, M. W., Delagenière, S., Guijarro, M., Spruce, D., Svensson, O., McSweeney, S. M., McCarthy, A. A., Leonard, G. & Nanao, M. H. (2013). *J. Appl. Cryst.* **46**, 804–810.
- Mueller-Dieckmann, C., Kauffmann, B. & Weiss, M. S. (2011). *J. Appl. Cryst.* **44**, 433–436.
- Murray, C. W. & Blundell, T. L. (2010). *Curr. Opin. Struct. Biol.* **20**, 497–507.
- 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.
- Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. (2000). *Nature (London)*, **406**, 752–757.
- Nogly, P. *et al.* (2015). *IUCrJ*, **2**, 168–176.
- Nurizzo, D., Mairs, T., Guijarro, M., Rey, V., Meyer, J., Fajardo, P., Chavanne, J., Biasci, J.-C., McSweeney, S. & Mitchell, E. (2006). *J. Synchrotron Rad.* **13**, 227–238.
- Oster, L., Tapani, S., Xue, Y. & Kack, H. (2015). *Drug Discov. Today*, **20**, 1104–1411.
- Owen, R. L., Rudiño-Piñera, E. & Garman, E. F. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 4912–4917.
- Paithankar, K. S. & Garman, E. F. (2010). *Acta Cryst. D***66**, 381–388.
- Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
- Rasmussen, S. G. F. *et al.* (2011). *Nature (London)*, **477**, 549–555.
- Ravelli, R. B. G., Leiros, H.-K. S., Pan, B., Caffrey, M. & McSweeney, S. (2003). *Structure*, **11**, 217–224.
- Riek, C., Burghammer, M. & Schertler, G. (2005). *Curr. Opin. Struct. Biol.* **15**, 556–562.
- Sanctis, D. de *et al.* (2012). *J. Synchrotron Rad.* **19**, 455–461.
- Sanctis, D. de & Leonard, G. (2014). *Not. Neutroni Luce Sincrotrone*, **19**, 24–26.
- Sanctis, D. de & Nanao, M. H. (2012). *Acta Cryst. D***68**, 1152–1162.
- Sauter, N. K., Hattne, J., Grosse-Kunstleve, R. W. & Echols, N. (2013). *Acta Cryst. D***69**, 1274–1282.
- Sawaya, M. R. *et al.* (2014). *Proc. Natl Acad. Sci. USA*, **111**, 12769–12774.
- Sheldrick, G. M. (2008). *Acta Cryst. A***64**, 112–122.
- Skubák, P. & Pannu, N. S. (2013). *Nature Commun.* **4**, 2777.
- Soares, A. S., Mullen, J. D., Parekh, R. M., McCarthy, G. S., Roessler, C. G., Jackimowicz, R., Skinner, J. M., Orville, A. M., Allaire, M. & Sweet, R. M. (2014). *J. Synchrotron Rad.* **21**, 1231–1239.
- Stellato, F. *et al.* (2014). *IUCrJ*, **1**, 204–212.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst. D***66**, 22–25.
- White, T. A., Barty, A., Stellato, F., Holton, J. M., Kirian, R. A., Zatsepin, N. A. & Chapman, H. N. (2013). *Acta Cryst. D***69**, 1231–1240.
- White, T. A., Kirian, R. A., Martin, A. V., Aquila, A., Nass, K., Barty, A. & Chapman, H. N. (2012). *J. Appl. Cryst.* **45**, 335–341.
- Yeates, T. O. (1997). *Method Enzymol.* **276**, 344–358.