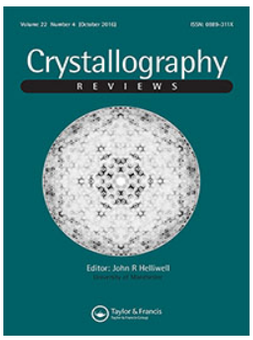


Crystallography Reviews

ISSN: 0889-311X (Print) 1476-3508 (Online) Journal homepage: https://www.tandfonline.com/loi/gcry20

Fully automatic macromolecular crystallography: the impact of MASSIF-1 on the optimum acquisition and quality of data

Matthew W. Bowler, Olof Svensson & Didier Nurizzo

To cite this article: Matthew W. Bowler, Olof Svensson & Didier Nurizzo (2016) Fully automatic macromolecular crystallography: the impact of MASSIF-1 on the optimum acquisition and quality of data, Crystallography Reviews, 22:4, 233-249, DOI: 10.1080/0889311X.2016.1155050

To link to this article: https://doi.org/10.1080/0889311X.2016.1155050

Published online: 22 Mar 2016.

Crystallography Reviews, 2016
Vol. 22, No. 4, 233–249, http://dx.doi.org/10.1080/0889311X.2016.1155050

Taylor & Francis
Taylor & Francis Group

REVIEW

Fully automatic macromolecular crystallography: the impact of MASSIF-1 on the optimum acquisition and quality of data

Matthew W. Bowler[a,b*], Olof Svensson[c] and Didier Nurizzo[c]

[a]*European Molecular Biology Laboratory, Grenoble Outstation, 71 avenue des Martyrs, CS 90181 F-38042 Grenoble, France;* [b]*Unit for Virus Host Cell Interactions, Univ. Grenoble Alpes-EMBL-CNRS, 71 avenue des Martyrs, CS 90181 F-38042 Grenoble, France;* [c]*European Synchrotron Radiation Facility, 71 avenue des Martyrs, CS 40220 F-38043 Grenoble, France*

(*Received 29 January 2016; accepted 13 February 2016*)

Automation is beginning to transform the way data are collected in almost all scientific disciplines. The combination of robotics and software now allows data to be collected consistently and reproducibly, eliminating human error and boredom. This approach has been applied to macromolecular crystallography at MASSIF-1, a fully automated beamline at the European Synchrotron Radiation Facility (ESRF). Considerable human effort is still dedicated to evaluating protein crystals in order to find the few crystals that diffract well or collecting hundreds of data sets to screen potential new drug candidates. The combination of ESRF-developed robotic sample handling and advanced software protocols now provides a new tool to structural biologists. Not only is the beamline used efficiently, running 24 h a day without getting tired, data collection is also performed consistently by an expert system, often better than with a human operator. In this review, we will focus on the impact this level of automation has had on the optimum acquisition of data from crystals of biological macromolecules.

Keywords: automation; X-ray crystal centring; synchrotron instrumentation; macromolecular crystallography

1. Introduction

Determining the atomic structure of proteins, nucleic acids and their complexes is of vital importance in understanding fundamental biology and medical research. Despite enormous technical advances since the determination of the first structures by X-ray crystallography [1,2] the process remains a challenge and demands considerable human effort. Automation has been a cornerstone of the advances made in the last two decades, making the process of protein production,[3,4] purification,[5] crystallization [6–9] and data collection [10–13] more streamlined. Coupling the extremely bright and small beams that can be produced at modern third-generation synchrotrons with automatic sample changers and on-line data analysis [10,12–20] has been one of the great advances, allowing not only large numbers of structures to be solved [21] but also incredibly challenging systems to be characterized structurally.[22–25] As the field of macromolecular crystallography (MX) reaches maturity, efforts are now focussing on automating the steps that still require considerable human involvement: crystal mounting,[26] data collection,[7,11,27] processing [16,28–31] and validation.[32]

*Corresponding author. Email: mbowler@embl.fr

The full automation of MX data collection, where human involvement is limited to sending samples and accessing data, has been discussed within the field for many years. Many synchrotrons offer 'mail-in' services, where a dedicated operator collects data for users [33–35] but this service has remained largely the proviso of the pharmaceutical industry. Lilly research laboratories collaborative access team at the advanced photon source and the stanford synchrotron radiation laboratory have led the way in truly automatic data collection [36,37] but as the services rely on optical centring, they tend to be restrictive and are again mostly used for proprietary research. This leaves a niche for a fully automatic system that is able to completely replace a human operator, in that it can work with the most challenging systems and make complex decisions based on the analysis of the data. The first beamline of the Massively Automated Sample Selection Integrated Facility (MASSIF-1) at the European Synchrotron Radiation Facility (ESRF) has been developed to fill this need.[38] The beamline relies on new developments in the automation of sample mounting [39] coupled to a state-of-the-art synchrotron beamline to provide a stable environment for user-free experiments. At the heart of the beamline is a system that will locate, centre to the optimal diffraction volume, characterize and collect data, if possible, from multiple cryo-cooled crystals.[40] The beamline has been operational since September 2014 and at the time of writing has processed over 15,000 samples without any user intervention. This new instrument not only increases the efficiency of data collection but also offers a new tool to structural biologists as the consistent treatment of samples can in many cases collect better data than a human operator. In this review, we give a brief overview of the technology developed for MASSIF-1 and concentrate on the impact this has had on how data are collected and the quality of the data obtained.

2. MASSIF-1

MASSIF-1 is the first beamline of the MASSIF complex to come online and is dedicated to the fully automatic evaluation and data collection from crystals of macromolecules.[38] The beamline is one of the facilities that were funded as part of the ESRF phase I upgrade and was initially intended to be a purely screening hub for the other beamlines.[41] The scientific case for the beamline was built during the age of charge coupled device detectors, when the collection of a full data set would take between 20 min and 2 h. This severely restricted the number of data sets that could be collected in a typical allocation of 24 h of beamtime. Therefore, sample screening was essential in order to maximize the efficiency and was a successful strategy for many high-profile projects.[23,25,42] MASSIF was intended to outsource the screening stage by automating the mounting and characterization of large numbers of crystals, allowing researchers to then take the best crystals and use dedicated data collection beamlines more efficiently. However, the advent of pixel array detectors (PADs) dramatically changed the use of synchrotron beamlines.[43] In addition to the enormous benefits of zero readout noise, the continuous readout ability of PADs allows the collection of data sets in the order of minutes, effectively eliminating the need for pre-screening. While this changed the focus of the beamlines, it has enabled them to become much more effective instruments than originally proposed. The increased throughput enabled by PADs does not remove the need to test the diffraction quality of crystals and also allows the collection of hundreds of data sets in the allocated beamtime. The focus of the MASSIF beamlines was therefore shifted to cover full automation of both screening and data collection using the capabilities of PADs. The full automation of highly complex experiments relies on three core technologies that will be briefly described in the following sections.

2.1. *The beam*

For a beamline to run autonomously a high degree of stability is required while retaining the beam properties demanded at a modern synchrotron. MASSIF-1 uses a synthetic diamond as

monochromator (using the same concept of beam splitting monochromators pioneered on ID14 [44]) and a compound refractive lens as the only focussing element.[45,46] This provides a highly intense beam (3×10^{12} ph s^{-1}) at the sample position of 100 µm × 65 µm ($H \times V$, full width half maximum). The use of only two optical elements leads to a high level of stability that simply requires an alignment of the experimental table once a day. The beam diameter can be further tailored by the insertion of apertures between 50 and 10 µm in diameter. The energy is fixed at 12.8 keV, allowing *de novo* structure solution using single wavelength anomalous diffraction (SAD) protocols.[38,47]

2.2. *Robotic sample mounting*

The automated mounting of cryo-cooled samples is now an essential component of any MX experiment. While the robotic sample changers implemented across the world have increased throughput, they often require small manual interventions to recover from errors. In order for a system to run completely autonomously, pitfalls must be avoided and routines for recovery need to be in place. This need stimulated the design of a new robot at the ESRF, named RoboDiff.[39] Based on an industrial standard 6-axis robot, the system acts as both sample changer and goniometer (Figure 1(a)). The system has been designed with a high level of diagnostics in order to allow recovery from problems encountered during the mounting and unmounting of samples. Samples are tracked at multiple points in the trajectory from dewar to sample position. The

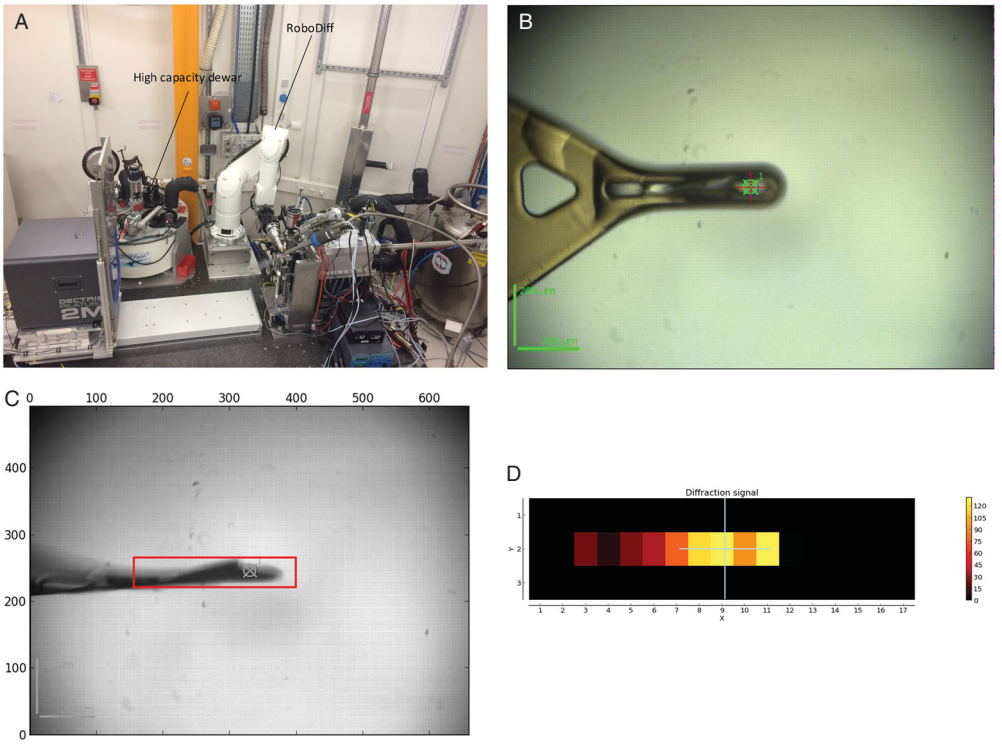


Figure 1. Overview of MASSIF-1 and the crystal location process. (a) View of the MASSIF-1 experimental hutch. The RoboDiff is shown in the goniometer position. (b) A sample support after optical centring. (c) The result of the mesh area determination is shown (red box), in this case the smallest orientation has been selected. (d) The output from the mesh scan. Diffraction images are scored by intensity distribution as a function of resolution. The centre of mass is calculated between positions within 50% of each other and used as the point for centring (white cross).

flexibility of an anthropomorphic arm, coupled to low-level software routines, allows the robot to escape from potentially blocking problems and continue to the next sample. Industrial robots are precise instruments; however, while their accuracy is excellent for sample manipulation, it is insufficient for the manipulation of a crystal in the 10–50 µm range in a beam of a similar size over a 360° rotation. In order to overcome this, the robot is only used to move the sample to the beam position; all subsequent movements are performed by a high-accuracy stage located below the robot for sample translations and an air bearing with a sphere of confusion < 2 µm diameter for rotations. In combination with a high-capacity dewar containing up to 240 Structural Proteomics IN Europe standard samples (Figure 1(a)) the system is able to run autonomously for between 24 and 48 h, depending on experiment requirements.

2.3. *Sample location, characterization and data collection algorithms*

The ability to autonomously place hundreds of crystals into a stable, highly intense beam is only the start. Until now, automatic systems have stopped at this point, relying on a large beam and matching crystal size to that of the loop to ensure that the crystal remains in the beam during rotation. In order to completely replace a human operator, an automatic system must be launched that can locate the crystal, often the most difficult step, and make multiple decisions on sample quality and data collection strategy.[40] On MASSIF-1, once a sample is mounted, an optical procedure moves the tip of the sample support to the centre of rotation of the goniometer that is aligned to the beam position (Figure 1(b)). Protein crystals come in a wide variety of shapes and sizes and are mounted in a dazzling array of supports. In order to efficiently locate samples without missing any part of the support, a fast optical tomography routine is launched. This determines the minimum or maximum areas that should be scanned through the X-ray beam (Figure 1(c)). The minimum area is selected when a single crystal is present; if multiple crystals are present the maximum orientation is selected, in order to distinguish crystals. The fast readout of the PAD allows continuous scans to be performed meaning that the crystal can be rapidly located by diffraction. Once located, the centre of mass of the best part of the crystal is centred to the beam (Figure 1(d)). A line scan is then launched 90° away in the ω axis in order to position the crystal at the beam over the full rotation range. Once centred, the crystal is characterized using the EDNA/BEST [14,17] pipeline and an optimized data collection strategy calculated. Data sets are then collected and automatically processed [28] before the next sample is mounted. In cases where indexing fails, a default data collection of 180° with a dose calculated to be the Garman limit (the dose at which diffracted intensity reaches 0.7 of initial intensity for an average crystal) given the measured flux is launched.

The system has been designed to make intelligent decisions at all stages; however, scientists usually know their samples best and despite the beamline being fully automatic, we wanted to introduce considerable flexibility in data collection parameters. This has been implemented by allowing user preferences (such as type of data collection (native/SAD/screening), resolution required, beam size, etc.) to be entered at the sample level in the diffraction plan (Table 1). An online booking system adds to the flexibility: rather than the rigid system in place on other beamlines, users can book time in an online calendar with very short notice and samples then enter a queuing system.

3. The impact of automation: a year in crystals

After just over a year of operation the beamline has processed over 15,000 crystals. Of these, 9872 were handled in 2015. Here, we have used the 2015 samples as the basis for an analysis of the impact MASSIF-1 has had on how data are collected and their quality. This cohort comes from a wide variety of projects from across the European structural biology community

Table 1. Diffraction plan entries.

Diffraction plan entry	Definition	Default value
Protein acronym	Defines the protein that is registered with the ESRF safety group	Required field
Sample name	User-defined unique identifier	Required field
Protein data bank file	Coordinates to be associated with an acronym	None
Pin barcode	Barcode identifier	None
Experiment type	Define MXPressE/O/SAD/Score	MXPressE
Space group	If present used for strategy calculation and autoprocessing	None
Pre-observed resolution	Resolution that will define the detector distance for mesh scans, characterization images and default data collection	2.0 Å
Required resolution	Threshold resolution, samples below cut-off will not be collected	None
Radiation sensitivity	BEST input in case of highly radiation sensitive crystals	1
Required completeness	–	99%
Required multiplicity	–	4
Number of positions	For multiple crystals	1
Preferred beamsize	Select appropriate beamsize for crystals	50 µm
SMILES	SMILES string for an associated ligand	None

Notes: The diffraction plan is filled for each sample in the Information System for Protein crYstallography Beamlines (ISPyB) and used by the system to tailor the experiment to the sample. In the absence of an entry, sensible default values are used. MXPressE is a workflow optimized for native data collection; MXPressO collects 180°; SAD optimizes a data set for structure solution via SAD and Score characterizes samples without data collection.

(Figure 2) and represents a reasonable snapshot of crystallographic projects. MASSIF-1 is used primarily for two types of project: those involving a large amount of screening to find the one crystal out of many that diffracts sufficiently and projects where large numbers of repetitive data collections are required, such as fragment screening campaigns. The full automation of data collection not only frees researcher's' time but also allows an expert system to make crucial decisions in the process using a large amount of information not usually used by human operators. The analysis presented here represents a general view of the impact of automation and also demonstrates some of the direct consequences for specific projects.

3.1. *The robot does not get tired*

The most obvious benefit of a completely autonomous system is that it can run day and night without getting tired. This is particularly useful at synchrotron sources where the facilities are scheduled for 24 h of continuous operation. As MASSIF-1 runs automatically, all beamtime can be used efficiently without the need for users or beamline scientists to work overnight. This is demonstrated by analysing the time that data collections are initiated on MASSIF-1 compared to the human-operated beamlines at the ESRF in 2015 (Figure 3). The most obvious trend on MASSIF-1 is that data collection rates remain steady regardless of time of day (Figure 3(a)). Fewer data collections are launched after 21:00 as this is the time when 100 crystals are completed, a common number received at the beamline. This contrasts with user-operated beamlines where a steady decrease in the number of data collections launched after 22:00 is observed (Figure 3(b)). The deleterious effects of sleep deprivation on cognitive ability are well studied [48,49] and map to the decrease in data collections (Figure 3(b)). By the time a user has been awake for 22 h, relative cognitive ability is equivalent to having consumed $\sim$5 glasses of wine.[49] The beamline acts efficiently through the night, performing exactly the same detailed

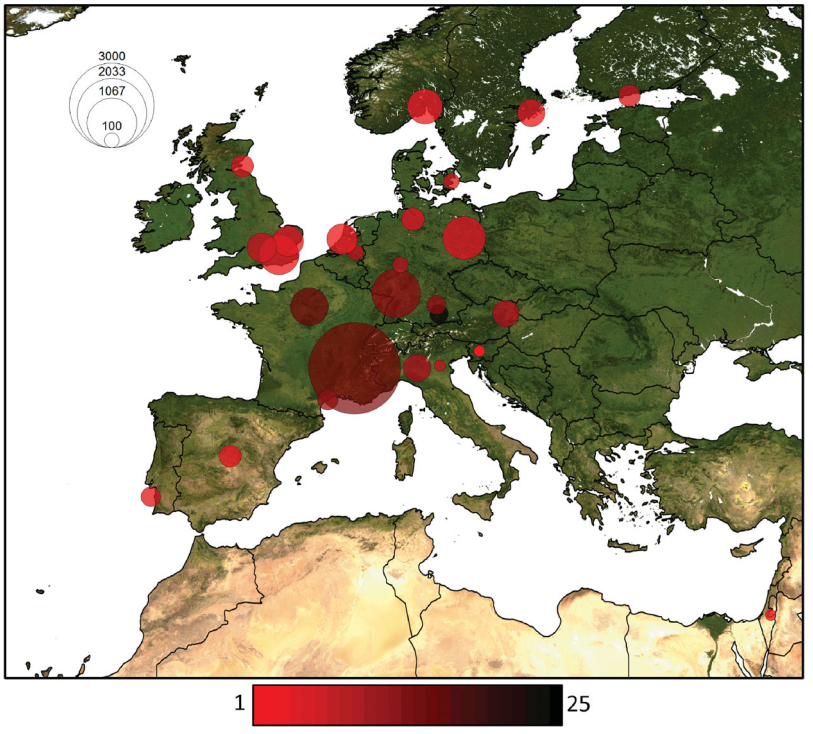


Figure 2. Geographical distribution of the 9872 crystals sent to MASSIF-1 from across Europe in 2015. Circles are scaled to the number of crystals and are centred on the city where the laboratory responsible is based. The colour is scaled to the number of different projects (protein acronyms) from each BAG.

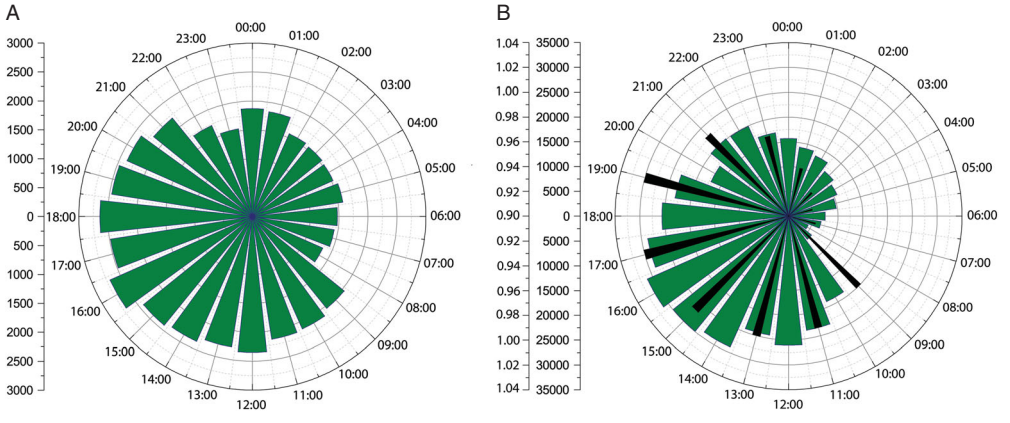


Figure 3. Robots do not sleep. Data collection totals by the hour in which they were started are shown as a 24 h clock for MASSIF-1 (left) and the other 7 MX beamlines at the ESRF (right) in 2015. There is a significant contrast in the proportion of data collections started after midnight, which drop continually on human-operated beamlines but remains steady on MASSIF-1. The drop in productivity maps closely to the decrease in mean relative performance in cognitive tests performed after hours of sustained wakefulness (right, black bars, central scale) assuming awakening at 07:00 (data from [49]). Dips at mealtimes (13:00 and 20:00) are also absent for MASSIF-1. It should be noted that the period 08:00 to 09:30 is reserved for beamline startup on the human operated beamlines.

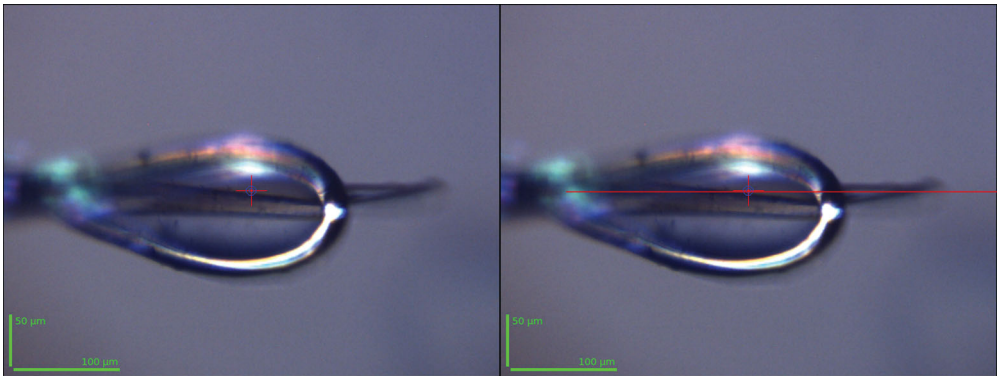


Figure 4. Seeing is not believing. A crystal is shown after an X-ray centring procedure; visual inspection suggests that the crystal is not centred as the beam appears to be above the crystal (left). However, the collection of four diffraction images over the full rotation range confirmed that the crystal remained in the beam over the full rotation range. Fortunately, in this case part of the crystal extends beyond the sample support where refractive and reflective effects are diminished, showing the true position of the crystal (right). Using this as a reference (red line), it can be seen that the crystal is indeed correctly centred to the beam position, demonstrating a 10 µm optical vertical offset.

analysis of each sample, regardless of the time, allowing researchers to analyse the data once collected.

3.2. *X-ray vision*

The most critical step in the whole process is the X-ray centring routine. Not only does this locate the crystal, it also defines the relative quality of diffraction and provides information on the dimensions of the crystal that are useful in downstream applications (see next section). One of the greatest advantages of consistently running centring using X-rays is that visible light plays no role. This is particularly useful when crystals are mounted in an opaque medium (such as lipidic cubic phase or after a poor cryoprotection protocol) but less obvious problems occur due to refractive and reflective effects that are difficult to resolve optically. Figure 4 shows a crystal after an X-ray centring routine has been performed. On inspection of the images it seems quite clear that the crystal is positioned below the X-ray beam (blue circle). In this case, part of the crystal is protruding from the nylon loop and free of mother liquor. This provides a clear reference for the actual position of the crystal in the loop. If a line is drawn using the centre of the exposed part of the crystal as a reference position (Figure 4, right) it can be seen that the crystal is indeed well centred and that a 10 µm optical vertical offset is observed. Four diffraction images, 90° apart in ω, confirmed that the crystal remained in the beam over the full rotation range. What is striking is that this sample is well mounted and is surrounded by minimal cryo-solution. As the sample is quite small ($\sim$25 µm vertically) the problem is magnified, with larger crystals, the problems encountered with optical effects may not always be noticed. The light that is incident on the video microscope travels 5 cm, with large changes in refractive index from the cooled nitrogen gas stream and the solution in the sample mount. This will lead to optical distortions from Schlieren and refractive effects. These *Fata Morgana* mean the true size and position of crystals cannot be determined optically, even when all liquid is removed.[50] Using consistent X-ray centring is the only way to ensure that a crystal is correctly centred over the full rotation range and to the position of highest diffraction signal.

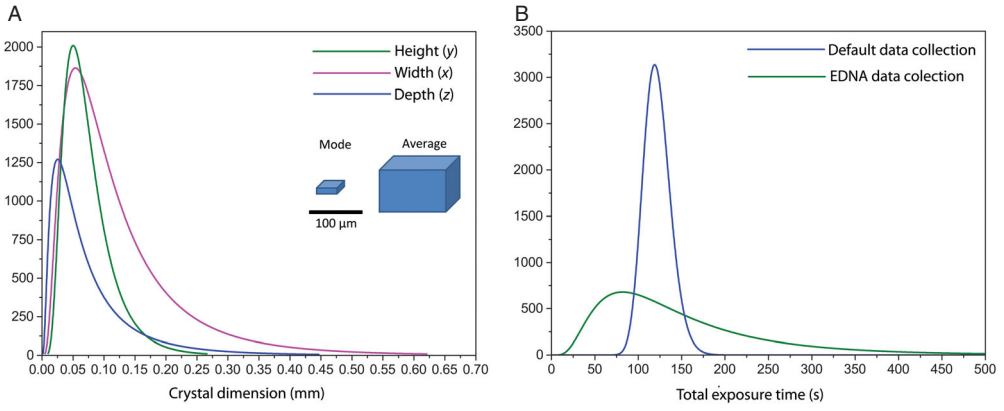


Figure 5. Distribution of measured crystal dimensions and exposure times for data collections on MASSIF-1. (a) Lognormal distributions of the observed crystal dimensions from 2015 crystals. The dimensions are: x the measured crystal length parallel to the spindle axis, y the length orthogonal to the spindle axis and z the length orthogonal to the spindle axis 90° away in ω. Representations of an average and modal crystal are shown for illustrative purposes. (b) Lognormal distributions of total data collection times in 2015. A much wider distribution is observed for strategies calculated using EDNA where the crystal volume, diffraction quality and beam properties are used in the calculation.

3.3. *An expert system uses all the information available to optimize data collection*

A distinct advantage of an automated system is the level of information that is known about each sample. The simple fact that all actions are chained means that the system knows which sample is being processed at any time and any information already gathered can then be used in the subsequent steps. This is a very simple process, but the unpredictability of human operators makes it troublesome to implement on a normal beamline. The X-ray centring procedure provides the most information on the sample, critically its position relative to the beam. However, it quickly became apparent that other information is available, such as the variability in diffraction quality [51] and the exact dimensions of the crystal, the latter being particularly useful. Having the measured dimensions for a large number of samples has allowed us to look at the distribution of sizes that are encountered on the beamline (Figure 5(a)). What is striking is that the most commonly observed size is much smaller than expected, with modal values in the order of 20–40 µm. This led to the decision to set the default beam size to 50 µm, as it best reflects the size of the samples being studied.[40] It is also clear from these data that most protein crystals form plates or needles rather than regular cuboids (Figure 5(a)). In addition to global trends, the availability of dimensions for each individual crystal allows two important adaptations: first, the beam size can be adjusted to best match the size of the crystal. This will reduce background scatter and improve signal to noise.[52] Second, accurate knowledge of the volume of the crystal allows the correct dose absorbed by the crystal during data collection to be calculated.[53] When the beam is much larger than the crystal the dose distributions can vary enormously as a function of rotation [54]; having the shape and size of the crystal allows the full dose allowable before radiation damage for a data set to be calculated, leading to a better data collection strategy. In cases where crystals are smaller than the beam the dose must be decreased to avoid radiation damage; here again the volume is essential to collecting the best possible data. This is demonstrated in the distribution of the exposure times required for data collection strategies (Figure 5(b)). When the software is unable to index crystals, or users specifically require it, a default data collection is launched. This strategy collects a 180° rotation limiting the dose to the Garman limit ($\sim$ 30 MGy [55]) for an average crystal. The distribution of these data collection times is tight due to small variations in flux from the synchrotron filling mode. However, when a strategy is calculated,

Table 2. The use of crystal volumes in strategy calculations.

	Crystal larger than beam		Crystal smaller than beam	
	Default crystal size	Measured crystal size	Default crystal size	Measured crystal size
Crystal dimensions (μm)	100 × 100 × 100	496 × 239 × 521	100 × 100 × 100	28 × 57 × 13
Beam diameter (μm)	50	50	50	50
Space group	*P*222	*P*222	*P*222	*P*222
Unit cell dimensions				
a, *b*, *c* (Å)	36.68, 49.88, 71.65	36.68, 49.88, 71.65	121.79, 186.2, 241.83	121.79, 186.2, 241.83
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Flux (ph s^{-1})	1.2×10^{12}	1.2×10^{12}	9×10^{11}	9×10^{11}
Transmission (%)	48.2	100	100	100
Dose (MGy)	**4.13**	**16.02**	**9.95**	**3.50**
Total exposure time (s)	**48.0**	**94.0**	**70.6**	**24.6**
Oscillation range (°)	0–360	0–360	27–160	24–160
Detector resolution (Å)	1.23	1.14	5.92	5.92

Notes: Differences in calculated strategy based on the measured crystal size or the default crystal size in RADDOSE and BEST. The strategies are taken from the largest and smallest crystals processed with a strategy on MASSIF-1 in 2015. The strategy for the largest crystal is optimized for structure solution by SAD protocols. The differences in dose and exposure time (bold) are of particular interest.

all parameters, such as the measured flux, beam profile and the crystal volume, are taken into account. This leads to a broad distribution as exposure times either increase or decrease given the specific crystal characteristics. The difference in the calculated strategy when either the measured crystal size or the default ($100 \times 100 \times 100\,\mu m^3$) is used is shown in Table 2. In cases where the crystal is much larger than the beam, longer exposure times can be used, leading to higher resolution and a greater signal to noise ratio. Where crystals are much smaller than the beam, the time taken to reach the dose limit is much shorter. By reducing the dose, excessive radiation damage should be avoided, again leading to better quality data. The use of consistent X-ray centring and the inclusion of all available sample information allow the calculation of optimized data collection strategies that directly translate into higher-quality data: plotting the distribution of signal to noise ratios ($\langle I/\sigma(I)\rangle$) for data sets that were automatically processed on MASSIF-1 and comparing them to those collected on the human-operated beamlines shows a very similar pattern but with a shift to significantly higher values for MASSIF-1 (Figure 6). This is reflected in the average and modal $\langle I/\sigma(I)\rangle$ values of 11 and 8.5 for MASSIF-1 compared to the human-operated beamlines, where the values are 9 and 6, respectively. There are many metrics that can be used to define the quality of processed data. We have selected the signal to noise as it allows us to directly compare the global quality of the data collected from all crystals collected at the ESRF in 2015. An enormous variety of crystals are tested at the ESRF but the similar distribution of values indicates that the types of project observed on MASSIF-1 are the same as those on the other beamlines (Figure 6). We attribute the higher signal to noise ratios

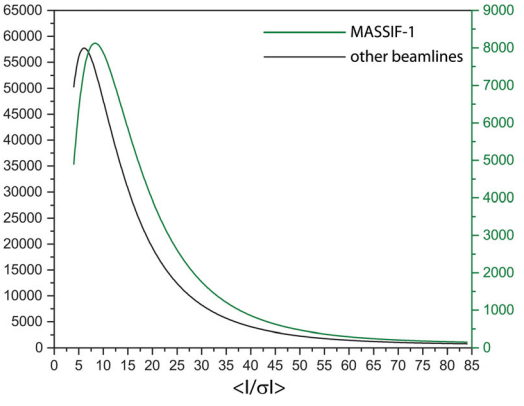


Figure 6. Distribution of $\langle I/\sigma(I)\rangle$ values for data sets automatically processed on MASSIF-1 and the ESRF human-operated beamlines in 2015. Lognormal distributions of the $\langle I/\sigma(I)\rangle$ values from data sets from MASSIF-1 (green) and other ESRF beamlines (black). The distributions are very similar in shape but there is a significant shift to higher values for MASSIF-1.

to the ability to correctly centre crystals and calculate optimized data collection strategies using information not usually available to users.

3.4. *How is the beamline used?*

The nearly 10,000 crystals processed on MASSIF-1 in 2015 represent a wide variety of projects composed of 180 different proteins. An analysis of the fate of all the samples received in 2015 is informative (Figure 7(a)). Almost all of the crystals received were examined; a small proportion could not be mounted. The crystals that were processed are equally split into three categories: a third contained empty loops or diffraction was absent; a third diffracted but were of poor quality and the final third yielded a processed data set (Figure 7(a)). Only 23% of crystals received could be indexed successfully by EDNA, showing that the beamline is used heavily for screening as well as for established projects. The collection of a default data set of 180° where indexing fails is also shown to be a useful procedure: for 678 crystals that could not be indexed, the default data set collected could be processed. Out of all samples only 2.8% were requested SAD protocols. However, these protocols led to the automatic solution of 20 unique structures in 2015.[47,56] While treating all user samples as a whole has led to important developments on the beamline,[40] it can mask the various trends that occur at the level of specific projects. The same analysis is therefore presented for three individual projects that represent different modes of using the beamline (Figure 7(b)–(d)). In the first case, 50 crystals were sent for SAD phasing. The crystals exhibited considerable variability in diffraction quality with only a 14% yield in processed data sets (Figure 7(b)). Of these processed data sets, two demonstrated a diffraction limit beyond 4 Å and one was used to solve the structure.[56] As these crystals grew in multiple plates, the X-ray centring routine allowed the best positions to be sampled consistently. Screening this number of crystals using manual centring on multiple positions is time consuming and samples fewer positions. The second case is a classical screening project (Figure 7(c)). For the 150 crystals of a membrane protein sent, none could be indexed from characterization images; however, three data sets collected could be processed. The routines allowed the rapid identification of additives that improved crystal quality. The final example shows the fate of crystals for a fragment-based drug design project. The crystals were thin plates that make manual centring challenging. Here, 95% of samples had an optimized data set collected from the best position; for the remaining 5% where indexing failed, a default data set was collected. One sample contained

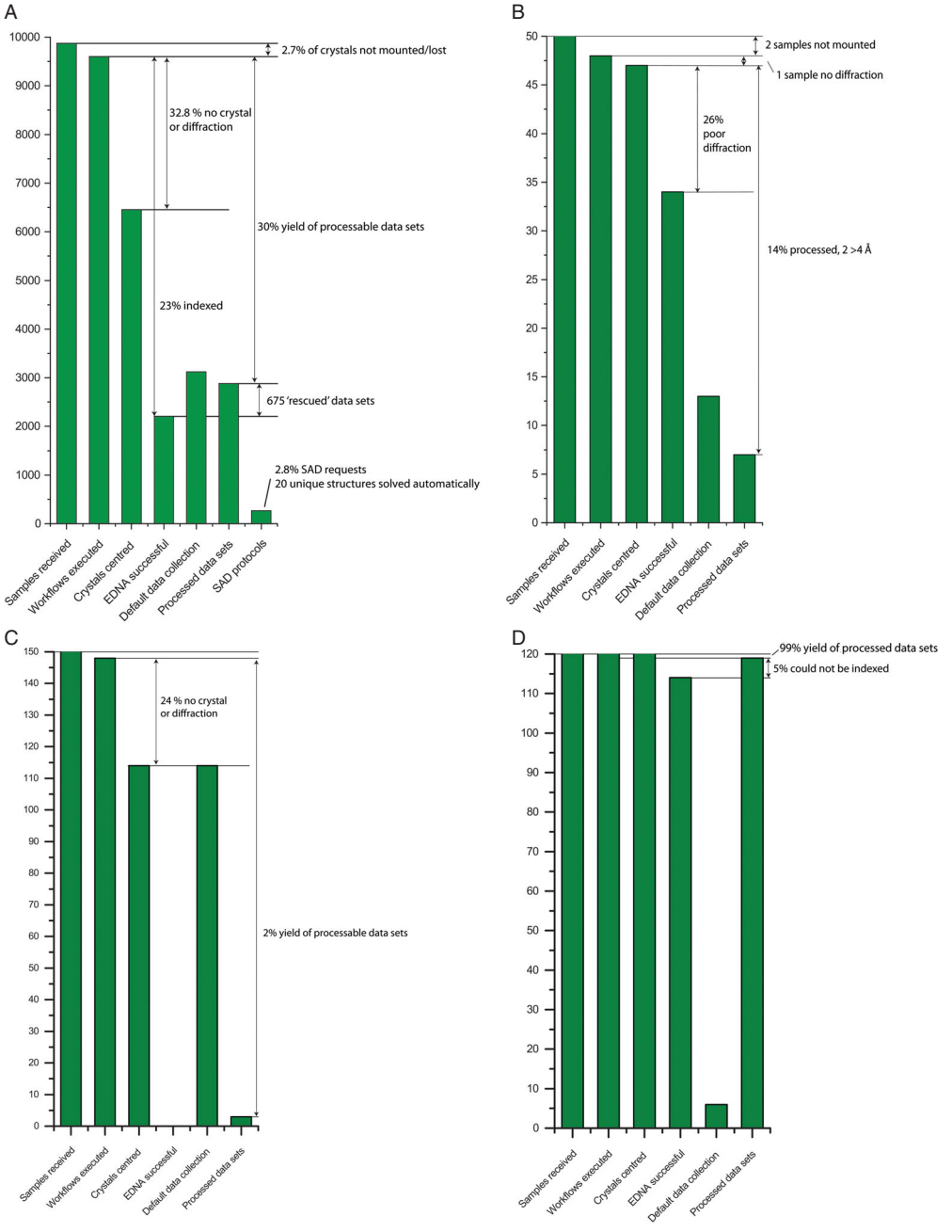


Figure 7. Crystal attrition on MASSIF-1 for all crystals and various types of individual project. (a) The fate of all crystals sent to MASSIF-1 in 2015 is shown. Of all samples sent, around one-third do not diffract or are empty loops; one-third diffract, but cannot be processed and the final third yield a processed data set. (b) A phasing project where crystals vary considerably in their quality, only 14% yielded a processed data set, two of these were $> d_{\min}$ 4 Å and 1 allowed the structure to be solved. (c) A screening project for a membrane protein. No crystals could be indexed initially, three of the default collections could be processed. (d) A fragment-based drug design project. A near 100% yield in data sets to look for small molecule binding.

no crystal (Figure 7(d)). These example projects are representative of the samples received at MASSIF-1. While all have different requirements, the automation in place has saved considerable time, not just by performing the experiment but by consistently scanning each sample to locate and optimize its position and use all available information to collect optimized data sets. This can be performed manually, but in a less consistent manner that would require far more time per sample.

3.5. *The impact of MASSIF-1 on data collection at other beamlines*

The technology developed for MASSIF has been directly exported to all the ESRF MX beamlines and has been extended into a wide variety of complex experiments. It was clear during the development of the instrumentation that the beamline would be highly dependent on on-line data analysis. This was pioneered at the ESRF MX beamlines [57] with the development of software

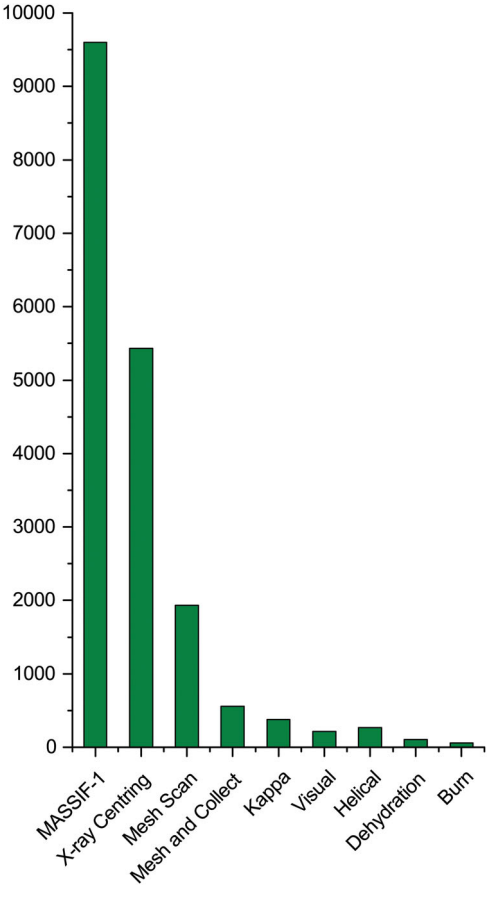


Figure 8. The impact of MASSIF-1 technology on the human-operated ESRF MX beamlines in 2015. The number of workflows launched by type is shown. The X-ray centring routine developed for MASSIF-1 has been highly successful on the other ESRF beamlines and is the most popular workflow run when users are present. 'Kappa' is a κ-goniometer reorientation, 'Visual' is a reorientation of the κ-goniometer along a user-defined line; 'Helical' takes two user-defined points on needle crystal and calculates an optimized strategy given translation of the crystal between the points; 'Dehydration' performs a dehydration gradient on a crystal while analysing the effects on the diffraction and 'Burn' determines a crystal's radiation sensitivity experimentally.

for automatic indexing and strategy calculation such as DNA, BEST and EDNA.[13,14,17,58] However, it became clear that for MASSIF to succeed, the analysis had to be taken a step further to allow the results of analysis to be used in further experiments involving instrument control. These experiments pose a challenge for both hardware (sample changers, goniometers) and software (data acquisition and data analysis). A classic example of this need is crystal reorientation, where characterization images must be collected in order to determine the orientation of the crystal, the goniometer κ angle must be moved and then the crystal re-characterized to determine the data collection strategy.[59] X-ray centring in particular clearly requires analysis coupled to instrument control. In order to automate such experiments a high-level software environment is required to coordinate data acquisition and data analysis. We selected a workflow manager that allowed the first steps to be taken in the coupling of on-line data analysis to instrument control.[60] The workflow manager allows complex workflows to be designed as modules that are independent from beamline control software and hardware. Workflows are then started by the beamline control GUI MXCuBE2 [61] that connects to a dedicated workflow server via a web-service call. The workflow then requests operations, such as data collections and motor movements, by connecting back to MXCuBE2 through an XML-RPC server. As each workflow is a module, they can be easily combined into more complex workflows, as was the case for MASSIF-1, where multiple processes such as centring, characterization and data collection are combined. The modules can also be easily exported to the other beamlines. This is the case for X-ray centring,[40] which is the most popular workflow run by users present at the beamline (Figure 8). Workflows have been expanded to cover a wide range of experiments such as dehydration of protein crystals,[62] kappa goniometer reorientations,[59] serial crystallography methods,[63] automatic strategies to determine the radiation sensitivity of crystals [64] and even solution scattering experiments.[65] In 2015, 8936 experiment workflows were launched by human operators on the ESRF beamlines (Figure 8), the most popular being those developed from MASSIF-1 technology; namely X-ray centring, mesh scans and mesh and collect (serial crystallography). The drive to automate data collection has been directly exported to the manually operated beamlines and applied to more complex experiments with great success.

4. Conclusions

The full automation of MX data collection is not just an exercise in efficiency but can also lead to better data by consistently selecting the best region or regions of crystals and using the information gathered to collect the best data possible from each sample. As the system is not subject to human weaknesses such as fatigue, emotional involvement with specific samples or the need to visualize samples, it is a powerful tool for the modern structural biologist. The scientist remains deeply involved through the ability to define specific experiment requirements *via* a database [66] and the availability of results in real time. Full automation also allows new data archiving policies to be implemented smoothly. The beamline collects all necessary metadata for complete data sets and this will allow us to automatically archive raw diffraction images for data sets. This is in line with the newly adopted data policy at the ESRF and the recommendations of the International Union of Crystallography (IUCr).[67,68] Once collected, data sets will be assigned a digital object identifier, archived and embargoed for three years. This will allow data sets to be linked to publications and make a valuable resource for the community open access, once released.

The beamline is not designed to completely replace user visits to the synchrotron but rather to do the hard work of screening crystals or collecting data sets through the night, freeing researchers to spend time on more challenging data collection problems or study the biology. The field of structural biology is currently undergoing incredible advances with the developments

being made in cryo-electron microscopy (Cryo-EM).[69,70] If the promise of Cryo-EM holds true, it will leave X-ray crystallography to determine the structures of components of complexes or proteins with a molecular weight < 100 kDa. This implies a significant shift towards service rather than the complex and time-consuming discipline that has defined MX over the last 40 years. Automated systems such as MASSIF-1 can then be used as the perfect complement to biological projects that will require greater effort in the production of complexes and the development of less mature techniques. Coupled to recent advances that have been made in the automatic mounting of crystals at the EMBL Grenoble [71] a full protein to structure pipeline can now be envisioned. The reduction in project lifetimes that this level of automation can enable should allow novel structures to be solved rapidly and new compounds of therapeutic or biological interest to be discovered and quickly developed. We hope that MASSIF-1 heralds a new age of fast and automated structure solution as a service for structural biologists.

Acknowledgements

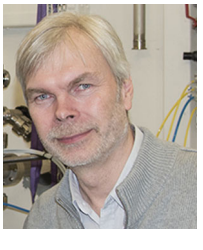
The authors would like to thank the support groups at the ESRF that provide outstanding services that allow complex beamlines to be constructed and maintained. Matias Guijarro, Pascal Theveneau and the technicians of the Structural Biology Group were instrumental in the construction of the beamline and are thanked for their invaluable contributions. The ESRF-EMBL Joint Structural Biology Group and the ESRF Automation Task Force (ATF) supported the developments.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Matthew W. Bowler studied Biological and Medicinal Chemistry at Exeter University receiving a BSc (Hons) in 1998. He then obtained an MSc from the Laboratory of Molecular Biophysics, Oxford University, and went on to read for a PhD with Prof. Sir John Walker, at the MRC and Cambridge University, on the structure and function of ATP synthase. After post-doctoral research, he moved to the ESRF, Grenoble in 2007 where he was Project Manager for the MASSIF sample evaluation beamlines, developing the scientific case and beamline design. In 2012 he moved to the EMBL Grenoble Outstation where he is responsible, with Didier Nurizzo, for the development of the world leading, fully automatic beamline MASSIF-1. His research centres on method developments, particularly the automation of complex experiments with Olof Svensson, and combining techniques, such as EM and SAXS, with high-resolution X-ray crystallography and NMR, to investigate the control of protein phosphorylation and its use and misuse in eukaryotic signalling networks.



Olof Svensson read Engineering Physics at Chalmers University of Technology (Gothenburg, Sweden) where he obtained an MSc in 1992. He obtained his PhD at the ESRF on the materials science beamline in 1994. After post-doctoral research on BM14, he joined the ESRF Scientific Software Group where he worked on experiment automation. In 2008 he received the BESSY Innovation Award, together with Vicente Rey, for his work concerning the automation of MX experiments. He has been the EDNA project manager since 2007 and is now responsible for workflow developments for experiment automation on the ESRF MX beamlines in general and for MASSIF-1 in particular.

Didier Nurizzo studied Biochemistry and Biophysics at the University of Marseille and obtained his PhD from the University of Paris XI, Orsay, working under the supervision of Prof. Christian Cambillau (AFMB, Marseille) on oxydo-reduction proteins. After post-doctoral research in Prof. Ted Baker's lab in New-Zealand, he moved to England to work in the York Structural Biology Laboratory under the supervision of Prof. Gideon Davies. At that time his main focus was on crystallography techniques. In 2003, he joined the ESRF MX group to build the ID23-1 MAD beamline and since then has been dedicated to beamline developments with a strong interest in automation and robotics. He is now responsible for MASSIF-1, in collaboration with Matthew W. Bowler, and the development of robotics for the ESRF MX beamlines.

References

[1] Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. Nature. 1958;181:662–666.

[2] Blake CC, Koenig DF, Mair GA, North AC, Phillips DC, Sarma VR. Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. Nature. 1965;206: 757–761.

[3] Hart DJ, Waldo GS. Library methods for structural biology of challenging proteins and their complexes. Curr Opin Struct Biol. 2013;23:403–408.

[4] Vijayachandran LS, Viola C, Garzoni F, et al. Robots, pipelines, polyproteins: enabling multiprotein expression in prokaryotic and eukaryotic cells. J Struct Biol. 2011;175:198–208.

[5] Camper DV, Viola RE. Fully automated protein purification. Anal Biochem. 2009;393:176–181.

[6] Ferrer JL, Larive NA, Bowler MW, Nurizzo D. Recent progress in robot-based systems for crystallography and their contribution to drug discovery. Exp Opin Drug Disc. 2013;8:835–847.

[7] Elsliger M-A, Deacon AM, Godzik A, et al. The JCSG high-throughput structural biology pipeline. Acta Crystallogr Sect F Struct Biol Crystal Commun. 2010;66:1137–1142.

[8] Heinemann U, Bussow K, Mueller U, Umbach P. Facilities and methods for the high-throughput crystal structural analysis of human proteins. Acc Chem Res. 2003;36:157–163.

[9] Joachimiak A. High-throughput crystallography for structural genomics. Curr Opin Struct Biol. 2009;19:573–584.

[10] Beteva A, Cipriani F, Cusack S, et al. High-throughput sample handling and data collection at synchrotrons: embedding the ESRF into the high-throughput gene-to-structure pipeline. Acta Crystallogr Sect D Biol Crystallogr. 2006;62:1162–1169.

[11] Bowler MW, Guijarro M, Petitdemange S, et al. Diffraction cartography: applying microbeams to macromolecular crystallography sample evaluation and data collection. Acta Crystallogr Sect D Biol Crystallogr. 2010;66:855–864.

[12] Cipriani F, Felisaz F, Launer L, et al. Automation of sample mounting for macromolecular crystallography. Acta Crystallogr Sect D Biol Crystallogr. 2006;62:1251–1259.

[13] Leslie AG, Powell HR, Winter G, et al. Automation of the collection and processing of X-ray diffraction data – a generic approach. Acta Crystallogr Sect D Biol Crystallogr. 2002;58:1924–1928.

[14] Bourenkov GP, Popov AN. Optimization of data collection taking radiation damage into account. Acta Crystallogr Sect D Biol Crystallogr. 2010;66:409–419.

[15] Cohen AE, Ellis PJ, Miller MD, Deacon AM, Phizackerley RP. An automated system to mount cryo-cooled protein crystals on a synchrotron beamline, using compact sample cassettes and a small-scale robot. J Appl Crystallogr. 2002;35:720–726.

[16] Holton J, Alber T. Automated protein crystal structure determination using ELVES. Proc Natl Acad Sci USA. 2004;101:1537–1542.

[17] Incardona MF, Bourenkov GP, Levik K, Pieritz RA, Popov AN, Svensson O. EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. J Synch Rad. 2009;16:872–879.

[18] Jacquamet L, Joly J, Bertoni A, et al. Upgrade of the CATS sample changer on FIP-BM30A at the ESRF: towards a commercialized standard. J Synch Rad. 2009;16:14–21.

[19] Snell G, Cork C, Nordmeyer R, et al. Automated sample mounting and alignment system for biological crystallography at a synchrotron source. Structure. 2004;12:537–545.

[20] Zhang Z, Sauter NK, van den Bedem H, Snell G, Deacon AM. Automated diffraction image analysis and spot searching for high-throughput crystal screening. J Appl Crystallogr. 2006;39:112–119.

[21] Berman HM, Coimbatore Narayanan B, Costanzo LD, et al. Trendspotting in the Protein Data Bank. FEBS Lett. 2013;587:1036–1045.
[22] Ben-Shem A, Jenner L, Yusupova G, Yusupov M. Crystal structure of the eukaryotic ribosome. Science. 2010;330:1203–1209.
[23] Selmer M, Dunham CM, Murphy FV, et al. Structure of the 70S ribosome complexed with mRNA and tRNA. Science. 2006;313:1935–1942.
[24] Sibanda BL, Chirgadze DY, Blundell TL. Crystal structure of DNA-PKcs reveals a large open-ring cradle comprised of HEAT repeats. Nature. 2010;463:118–121.
[25] Warne T, Serrano-Vega MJ, Baker JG, et al. Structure of a β1-adrenergic G-protein-coupled receptor. Nature. 2008;454:486–491.
[26] Deller MC, Rupp B. Approaches to automated protein crystal harvesting. Acta Crystallogr Sect F Struct Biol Commun. 2014;70:133–155.
[27] Pothineni SB, Strutz T, Lamzin VS. Automated detection and centring of cryocooled protein crystals. Acta Crystallogr Sect D Biol Crystallogr. 2006;62:1358–1368.
[28] Monaco S, Gordon E, Bowler MW, et al. Automatic processing of macromolecular crystallography X-ray diffraction data at the ESRF. J Appl Crystallogr. 2013;46:804–810.
[29] Winn MD, Ballard CC, Cowtan KD, et al. Overview of the CCP4 suite and current developments. Acta Crystallogr Sect D Biol Crystallogr. 2011;67:235–242.
[30] Winter G. xia2: an expert system for macromolecular crystallography data reduction. J Appl Crystallogr. 2010;43:186–190.
[31] Perrakis A, Morris R, Lamzin VS. Automated protein model building combined with iterative structure refinement. Nat Struct Biol. 1999;6:458–463.
[32] Joosten RP, Joosten K, Murshudov GN, Perrakis A. PDB_REDO: constructive validation, more than just looking for errors. Acta Crystallogr Sect D Biol Crystallogr. 2012;68:484–496.
[33] Malbet-Monaco S, Leonard GA, Mitchell EP, Gordon EJ. How the ESRF helps industry and how they help the ESRF. Acta Crystallogr Sect D Biol Crystallogr. 2013;69:1289–1296.
[34] Okazaki N, Hasegawa K, Ueno G, Murakami H, Kumasaka T, Yamamoto M. Mail-in data collection at SPring-8 protein crystallography beamlines. J Synch Rad. 2008;15:288–291.
[35] Robinson H, Soares AS, Becker M, Sweet R, Heroux A. Mail-in crystallography program at Brookhaven National Laboratory's National Synchrotron Light Source. Acta Crystallogr Sect D Biol Crystallogr. 2006;62:1336–1339.
[36] Tsai Y, McPhillips SE, Gonzalez A, et al. AutoDrug: fully automated macromolecular crystallography workflows for fragment-based drug discovery. Acta Crystallogr Sect D Biol Crystallogr. 2013;69:796–803.
[37] Wasserman SR, Koss JW, Sojitra ST, Morisco LL, Burley SK. Rapid-access, high-throughput synchrotron crystallography for drug discovery. Trends Pharm Sci. 2012;33:261–267.
[38] Bowler MW, Nurizzo D, Barrett R, et al. MASSIF-1: a beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules. J Synch Rad. 2015;22:1540–1547.
[39] Nurizzo D, Bowler MW, Theveneau P, Guichard N, Guijarro M, Leonard GA. RoboDiff: combining a sample changer and goniometer for highly automated macromolecular crystallography experiments. J Appl Crystallogr.; submitted for publication.
[40] Svensson O, Malbet-Monaco S, Popov A, Nurizzo D, Bowler MW. Fully automatic characterization and data collection from crystals of biological macromolecules. Acta Crystallogr Sect D Biol Crystallogr. 2015;71:1757–1767.
[41] Theveneau P, Baker R, Barrett R, et al. The upgrade programme for the structural biology beamlines at the European Synchrotron Radiation Facility: high throughput sample evaluation and automation. J Phys: Conf Ser. 2013;425:012001.
[42] Ben-Shem A, Frolow F, Nelson N. Crystal structure of plant photosystem I. Nature. 2003;426: 630–635.
[43] Hulsen G, Broennimann C, Eikenberry EF, Wagner A. Protein crystallography with a novel large-area pixel detector. J Appl Crystallogr. 2006;39:550–557.
[44] Wakatsuki S, Belrhali H, Mitchell EP, et al. ID14 'Quadriga', a beamline for protein crystallography at the ESRF. J Sync Rad. 1998;5:215–221.
[45] Snigirev A, Kohn V, Snigireva I, Lengeler B. A compound refractive lens for focusing high-energy X-rays. Nature. 1996;384:49–51.
[46] Vaughan GBM, Wright JP, Bytchkov A, et al. X-ray transfocators: focusing devices based on compound refractive lenses. J Sync Rad. 2011;18:125–133.

[47] Kharde S, Calvino FR, Gumiero A, Wild K, Sinning I. The structure of Rpf2-Rrs1 explains its role in ribosome biogenesis. Nucleic Acids Res; 2015;43:7083–7095.
[48] Accidents and sleepiness: a consensus statement from the international conference on work hours, sleepiness and accidents; 1994 Sept 8–10; Stockholm. J Sleep Res. 1994;3:195.
[49] Dawson D, Reid K. Fatigue, alcohol and performance impairment. Nature. 1997;388:235.
[50] Pellegrini E, Piano D, Bowler MW. Direct cryocooling of naked crystals: are cryoprotection agents always necessary? Acta Crystallogr Sect D Biol Crystallogr. 2011;67:902–906.
[51] Bowler MG, Bowler MW. Measurement of the intrinsic variability within protein crystals: implications for sample-evaluation and data-collection strategies. Acta Crystallogr Sect F Struct Biol Commun. 2014;70:127–132.
[52] Schneider TR. Synchrotron radiation: micrometer-sized x-ray beams as fine tools for macromolecular crystallography. HFSP J. 2008;2:302–306.
[53] Paithankar KS, Owen RL, Garman EF. Absorbed dose calculations for macromolecular crystals: improvements to RADDOSE. J Synch Rad. 2009;16:152–162.
[54] Zeldin OB, Gerstel M, Garman EF. Optimizing the spatial distribution of dose in X-ray macromolecular crystallography. J Sync Rad. 2013;20:49–57.
[55] Owen RL, Rudiño-Piñera E, Garman EF. Experimental determination of the radiation dose limit for cryocooled protein crystals. Proc Natl Acad Sci USA. 2006;103:4912–4917.
[56] Muir KW, Kschonsak M, Li Y, Metz J, Haering CH, Panne D. Structure of the Pds5–Scc1 complex and implications for cohesin function. Cell Rep. 2016.
[57] Mueller-Dieckmann C, Bowler M, Carpentier P, et al. The status of the macromolecular crystallography beamlines at the European Synchrotron Radiation Facility. Eur Phys J Plus. 2015;130:1–11.
[58] Arzt S, Beteva A, Cipriani F, et al. Automation of macromolecular crystallography beamlines. Prog Biophys Mol Biol. 2005;89:124–152.
[59] Brockhauser S, Ravelli RBG, McCarthy AA. The use of a mini-[kappa] goniometer head in macromolecular crystallography diffraction experiments. Acta Crystallogr Sect D Biol Crystallogr. 2013;69:1241–1251.
[60] Brockhauser S, Svensson O, Bowler MW, et al. The use of workflows in the design and implementation of complex experiments in macromolecular crystallography. Acta Crystallogr Sect D Biol Crystallogr. 2012;68:975–984.
[61] Gabadinho J, Beteva A, Guijarro M, et al. MxCuBE: a synchrotron beamline control environment customised for macromolecular crystallography experiments. J Synch Rad. 2010;17:700–707.
[62] Bowler MW, Mueller U, Weiss M, et al. Automation and experience of controlled crystal dehydration: results from the European synchrotron HC1 collaboration. Cryst Growth Des. 2015;15(3):1043–1054.
[63] Zander U, Bourenkov G, Popov AN, et al. MeshAndCollect: an automated multi-crystal data-collection workflow for synchrotron macromolecular crystallography beamlines. Acta Crystallogr Sect D Biol Crystallogr. 2015;71:2328–2343.
[64] Leal RM, Bourenkov GP, Svensson O, Spruce D, Guijarro M, Popov AN. Experimental procedure for the characterization of radiation damage in macromolecular crystals. J Synch Rad. 2011;18:381–386.
[65] De Maria Antolinos A, Pernot P, Brennich ME, et al. ISPyB for BioSAXS, the gateway to user autonomy in solution scattering experiments. Acta Crystallogr Sect D Biol Crystallogr. 2015;71:76–85.
[66] Delagenière S, Brenchereau P, Launer L, et al. ISPyB: an information management system for synchrotron macromolecular crystallography. Bioinformatics. 2011;27:3186–3192.
[67] Tanley SWM, Diederichs K, Kroon-Batenburg LMJ, Schreurs AMM, Helliwell JR. Experiences with archived raw diffraction images data: capturing cisplatin after chemical conversion of carboplatin in high salt conditions for a protein crystal. J Sync Rad. 2013;20:880–883.
[68] Tanley SWM, Schreurs AMM, Helliwell JR, Kroon-Batenburg LMJ. Experience with exchange and archiving of raw data: comparison of data from two diffractometers and four software packages on a series of lysozyme crystals. J Appl Crystallogr. 2013;46:108–119.
[69] Kühlbrandt W. The resolution revolution. Science. 2014;343:1443–1444.
[70] Subramaniam S, Kuhlbrandt W, Henderson R. CryoEM at IUCrJ: a new era. IUCrJ. 2016;3:3–7.
[71] Cipriani F, Rower M, Landret C, Zander U, Felisaz F, Marquez JA. CrystalDirect: a new method for automated crystal harvesting based on laser-induced photoablation of thin films. Acta Crystallogr Sect D Biol Crystallogr. 2012;68:1393–1399.