Attaining atomic resolution from \textit{in situ} data collection at room temperature using counter-diffusion-based low-cost microchips

Jose A. Gavira, Isaac Rodriguez-Ruiz, Sergio Martinez-Rodriguez, Shibom Basu, Sébastien Teychené, Andrew A. McCarthy and Christoph Mueller-Dieckman

\textit{Acta Cryst.} (2020). D\textbf{76}, 751–758
Attaining atomic resolution from \textit{in situ} data collection at room temperature using counter-diffusion-based low-cost microchips

Jose A. Gavira,a,*, Isaac Rodriguez-Ruiz,b Sergio Martinez-Rodriguez,a,c Shibom Basu,d Sébastien Teychené,b Andrew A. McCarthyd and Christoph Mueller-Dieckmane

\textsuperscript{a}Laboratorio de Estudios Cristalográficos, IACT, CSIC–Universidad de Granada, Avenida Las Palmeras 4, 18100 Armilla, Spain, \textsuperscript{b}Laboratoire de Génie Chimique, Université de Toulouse, CNRS, INP, INSA, UPS Toulouse, Toulouse, France, \textsuperscript{c}Departamento de Bioquímica, Biología Molecular III e Inmunología, University of Granada, Avenida de la Investigación 11, 18071 Granada, Spain, \textsuperscript{d}EMBL Grenoble, 71 Avenue des Martyrs, CS 90181, 38042 Grenoble, France, and \textsuperscript{e}Structural Biology Group, ESRF, 6 Rue Jules Horowitz, 38043 Grenoble, France. *Correspondence e-mail: jgavira@iact.ugr-csic.es

Sample handling and manipulation for cryoprotection currently remain critical factors in X-ray structural determination. While several microchips for macromolecular crystallization have been proposed during the last two decades to partially overcome crystal-manipulation issues, increased background noise originating from the scattering of chip-fabrication materials has so far limited the attainable resolution of diffraction data. Here, the conception and use of low-cost, X-ray-transparent microchips for \textit{in situ} crystallization and direct data collection, and structure determination at atomic resolution close to 1.0 Å, is presented. The chips are fabricated by a combination of either OSTEMER and Kapton or OSTEMER and Mylar materials for the implementation of counter-diffusion crystallization experiments. Both materials produce a sufficiently low scattering background to permit atomic resolution diffraction data collection at room temperature and the generation of 3D structural models of the tested model proteins lysozyme, thaumatin and glucose isomerase. Although the high symmetry of the three model protein crystals produced almost complete data sets at high resolution, the potential of in-line data merging and scaling of the multiple crystals grown along the microfluidic channels is also presented and discussed.

1. Introduction

With the advent of ultrabright microfocus X-ray synchrotron sources (Flot \textit{et al.}, 2010) and femtosecond X-ray free-electron lasers (XFELs), new technologies and methodologies are in use to extract the most from these facilities and the capabilities offered to users (Spence, 2017, 2020; Tolstikova \textit{et al.}, 2019). One of their main advantages is that diffraction data can (again) be collected from samples at room temperature (RT), versus the standard diffraction data collection at cryogenic temperatures. In parallel, much effort has been devoted to ease the challenges of data collection at synchrotron sources, from sample mounting and centering to crystal screening (Bowler \textit{et al.}, 2015) and data processing (Aller \textit{et al.}, 2015; Hirata \textit{et al.}, 2019; Yu \textit{et al.}, 2019; Sparta \textit{et al.}, 2016; Basu \textit{et al.}, 2019; Monaco \textit{et al.}, 2013). However, sample handling remains a major issue. Cryoprotection of samples is another difficulty that is directly related to a loss of diffraction power and crystal quality (Pflugrath, 2015; Moreau \textit{et al.}, 2019). RT data collection, on the other hand, is a clear way to
avoid this procedure-dependence of data quality, as has recently been highlighted and reviewed (Broecker et al., 2018) for the diffraction of crystals in situ in their crystallization device. Furthermore, RT diffraction data can reveal motions crucial for catalysis, ligand binding and allosteric regulation, which are not always accessible using standard cryogenic data collection (Fraser et al., 2011; van den Bedem & Fraser, 2015).

For many years, our group has focused on the use of the capillary counter-diffusion technique, a particular application of free-interface diffusion (FID), for the screening and refinement of crystallization conditions and to gather in situ X-ray diffraction data either directly from crystals in capillaries at RT or from crystals at cryogenic temperatures (Otalora et al., 2009; Gavira, 2016). By definition, and as a base condition, the counter-diffusion technique requires the use of an environment with reduced convection, such as that provided by narrow capillaries. In the case of performing screening for crystallization conditions (González-Ramírez et al., 2017), counter-diffusion allows a wider region of the phase diagram to be explored while offering a stable solution environment that facilitates sample transportation. As radiation damage is an issue for data collection at room temperature at synchrotron sources (Garman & Owen, 2006), and to avoid crystal extraction for cryoprotection, we also investigated and demonstrated that the diffusion of cryoprotectant could be performed concomitantly with crystallization, even allowing the flash-cooling of crystals within capillaries (Ng et al., 2003; Gavira et al., 2002). Nevertheless, the technique does not seem to be user-friendly enough for adoption in structural biology laboratories, even though a variety of hardware has been developed and commercialized to ease its implementation (Ng et al., 2008; Dhouib et al., 2009; Pinker et al., 2013).

In parallel, the automation and miniaturization of crystallization hardware have been much improved and developed (Gavira, 2016). As summarized in early 2003 by Pusey and coworkers: a microfluidic chip for protein crystallization should be easy to use, be able to maintain a sealed environment over several months and provide access for the observation/visualization and harvesting of crystals as they are grown, whenever required (van der Woerd et al., 2003). Over the last two decades, abundant new microfluidic-based crystallization hardware has been proposed for the crystallization of biomacromolecules (Zhang et al., 2017; Sui & Perry, 2017; Gicquel et al., 2018). From the first chips tested by Hansen et al. (2002), much effort was concentrated on facilitating crystal extraction for subsequent cryoprotection (Gerds et al., 2008; Li et al., 2009; Hansen et al., 2006). However, current developments are focused on systems that not only allow in situ data collection (Gerds et al., 2008; Heymann et al., 2014), even for membrane-protein crystallization (Khvostichenko et al., 2014; Schieferstein et al., 2018), but also allow monitoring of protein crystal nucleation and growth by dynamic light scattering within each isolated nanolitre-size drop (Gicquel et al., 2018). Among those developments, several systems based on the FID technique have been successfully designed for the collection of in situ X-ray diffraction data for structure determination.

The first reported FID structure of thaumatin was collected using a polydimethylsiloxane (PDMS) chip, with the crystals diffracting to 1.25 Å resolution (Hansen et al., 2006). Later, the use of cyclic olefin copolymer (COC) chips designed for the counter-diffusion technique were reported, with the collection of diffraction data from thaumatins, and lysozyme crystals to 2.0 and 1.76 Å resolution, respectively (Ng et al., 2008). The combination of different materials [PDMS/poly(methyl methacrylate) (PMMA)/COC] to optimize the attainable resolution limit by reducing the background and the adsorption of diffracted reflections has also been explored (Dhouib et al., 2009). Resolution limits of 2.8 Å (PDMS), 1.9 Å (PMMA) and 1.65 Å (COC) for thaumatin and 1.5 Å for lysozyme were reported using a thinner chip made of COC (Dhouib et al., 2009). A significant advance was demonstrated with the production of a new generation of chips (ChipX) made of COC with the potential to be commercialized at an affordable price. ChipX was used to determine phases from data collected at RT from insulin and thaumatin crystals, and from lysozyme crystals derivatized with ytterbium, with the best crystals diffracting to 1.43, 1.55 and 2.3 Å resolution for lysozyme, thaumatin and insulin, respectively (Pinker et al., 2013). More recently, we have participated as members of the testing group for the commercial version of ChipX (ChipX3), which incorporates several improvements in terms of fabrication and both sample and reservoir loading. ChipX3 is also produced using COC to minimize X-ray scattering (de Wijn et al., 2019). The general applicability of this concept is not only as a screening tool based on the counter-diffusion method; it also allows fluorescent labeling, soaking of ligands or phasing of molecules, among other applications/advantages.

In this work, we present a low-cost approach to optimally collect RT data from multiple crystals grown in microchips using the counter-diffusion technique. Our objective was to demonstrate the feasibility of obtaining 3D structural models at the highest attainable resolution from crystals grown in microfluidic systems by collecting diffraction data at RT (i.e. without harvesting individual crystals). We have demonstrated this approach using three model proteins, glucose isomerase, lysozyme and thaumatin, generating the corresponding models at atomic resolution (near 1.0 Å). The proposed system can be incorporated into a fully automated workflow on a synchrotron beamline (in our case ID30B at the European Synchrotron Research Facility) for use as a high-throughput approach, facilitating the collection of users’ data with no intervention required by the end user. Moreover, these chips can easily be shipped by regular mail from the users’ laboratory to a synchrotron site to allow completely hands-off diffraction data collection and processing.

2. Materials and methods

2.1. Microchip fabrication

Hybrid OSTEMER–Kapton/Mylar microchips were specifically fabricated using 25 μm thick Kapton and Mylar windows (DuPont, France) to minimize X-ray background
noise. Briefly, OSTEMER (Mercenelabs, Sweden; Carlborg et al., 2012), which is based on thiol–ene click chemistry (Hoyle & Bowman, 2010), is an off-stoichiometry thiol–alkene–epoxy polymer with two different curing steps. In the first step, the OSTEMER formulation in the liquid state was poured beside a PDMS mold (fabricated with a low-cost soft-lithographic procedure described elsewhere; Rodriguez-Ruiz et al., 2017) placed against a first Kapton/Mylar film (Fig. 1).

The liquid formulation slowly diffuses by capillary action, filling the PDMS mold placed against the Kapton/Mylar film, which is later removed and which imprints the shape and determines the thickness of the reservoirs and channel structures. Once the full structure had been filled, OSTEMER was subsequently UV-cured, becoming solid, but flexible and yet sticky, so that the structure could be removed from the PDMS mold. Finally, a second Kapton/Mylar film was added on top and was auto-glued to the OSTEMER material using the remaining free epoxy groups. A final temperature-curing step completed the fabrication of the chips.

The microchip design consists of six rectangular-shaped microchannels with dimensions of 250 μm in width by 300 μm in depth (Fig. 1b) and a total length of approximately 4.0 cm. The cross section presented to the X-ray beam would thus consist of a 300 μm thick channel with 25 μm thick windows of Kapton or Mylar on each side. On one side of the microchip, the channels are joined in a small reservoir connected to an extra 400 × 300 μm channel, allowing the easy insertion of a 30G gauge syringe needle or protein solution injection (Fig. 2). On the opposite side of each channel, a precipitant reservoir is located at one end of the microchip, connected to an extra needle-inlet channel. Once the protein solution had been injected and the channels filled with the protein solution, a buffering plug of agarose was created at the end of each channel by injecting the necessary amount of agarose sol. Once a gel had been formed, the precipitant solution was injected to create a supersaturation gradient by precipitant counter-diffusion (detailed below), and the needle-inlet channels were sealed with Crystal Clear tape. The sequential steps are summarized in Fig. 2.

2.2. Protein crystallization

Lysozyme (hen egg-white lysozyme) and thaumatin (from Thaumatococcus daniellii) were purchased as lyophilized powders from Sigma (catalog Nos. L6876 and T7638, respectively), and glucose isomerase (β-xylose ketol-isomerase) from Streptomyces rubiginosus was purchased as a crystal suspension from Hampton Research (catalog No. HR7-100). Lysozyme was dissolved and dialyzed in 50 mM sodium acetate pH 4.5. Thaumatin was directly dissolved in deionized water prior to use. Glucose isomerase crystals were dissolved in deionized water and extensively dialyzed against 100 mM HEPES pH 7.0. Lysozyme and glucose isomerase were concentrated using Amicon Ultra-0.5 devices (Merck Millipore). The final concentrations of the protein solutions were determined spectrophotometrically by measuring their absorbance at 280 nm.

The crystallization conditions used to produce the crystals of the three model proteins were extracted from a previous
study (Conejero-Muriel et al., 2015). The initial experimental conditions are summarized in Table 1. Microchips were loaded following the previously described protocol.

2.3. Data collection and processing

A pair of microchips (Kapton and Mylar) containing the crystallization experiment for each protein were held in a modified crystallization microplate that served as a frame-holder for both chips at the same time (Fig. 3a). Diffraction experiments were performed on the ID30B beamline (McCarthy et al., 2018) at ESRF with the microplate held by the plate-gripper device mounted directly on the goniometer head of ID30B. Each chip was inspected with the help of a microscope equipped with a polarizer, and crystal positions were marked (with a red marker as shown in Fig. 3) to facilitate the identification of crystal locations. For each chip, data collection was started from the precipitant reservoirs to the protein-injection point (see Fig. 2). All identified crystals were tested for X-ray diffraction, including those grown in the precipitant reservoir, as observed in the case of glucose isomerase (Supplementary Fig. S1). A data-collection protocol of 40–45° wedges with a 0.1° oscillation range at 9.8% transmission was established using 12.4 keV energy photons and was applied for each protein using a beam size of 20 × 21 μm. In order to minimize crystal radiation damage, we followed a helical protocol (Flot et al., 2010) with the translation path corresponding to the length. Exceptions made to this protocol are listed below.

The collected partial diffraction data were input into the automatic data-processing pipeline as implemented on all ESRF MX beamlines (Monaco et al., 2013). There are six different pipelines that are run as data are collected: XDSAPP (Kabsch, 2010), GrenADES fastproc, GrenADES parallelproc, xia2 DIALS (Winter et al., 2018), autoPROC and EDNAproc, which makes use of POINTLESS (Evans, 2006) and AIMLESS (Evans & Murshudov, 2013). Subsequently, the results are scored and deposited in the ISPyB database (Delagenière et al., 2011), which is accessible through the EXI web service (https://exi.esrf.fr). The data shown in Supplementary Tables S1–S6 correspond to the non-anomalous data output from BEST (Bourenkov & Popov, 2010). The statistics compiled in the tables are intended to show the outputs derived from the different processing pipelines and are not for

---

### Table 1
Protein characteristics and initial crystallization conditions.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Concentration (mg ml⁻¹)</th>
<th>Molecular weight (kDa)</th>
<th>pI</th>
<th>Crystallization condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose isomerase</td>
<td>72</td>
<td>43.2</td>
<td>5.2</td>
<td>0.1 M MgCl₂, 10% (w/v) PEG 1K, 0.1 M HEPES pH 7.0</td>
</tr>
<tr>
<td>Lysozyme</td>
<td>10</td>
<td>14.3</td>
<td>9.0</td>
<td>5% (w/v) NaCl, 50 mM sodium acetate pH 4.5</td>
</tr>
<tr>
<td>Thaumatin</td>
<td>200</td>
<td>22.2</td>
<td>7.9</td>
<td>20% (w/v) sodium potassium tartrate</td>
</tr>
</tbody>
</table>

---

Figure 3
(a) Setup mounted for diffraction experiments on the ID30B beamline at ESRF. The two microchips are attached to a standard crystallization microplate serving as a holder for the plate-gripper goniometer head. Images of Kapton (center column) and Mylar (right column) microchips are shown containing crystals of thaumatin (b, c), glucose isomerase (d, e) and lysozyme (f, g). As a reference, the short-length channels have a dimension of 250 μm. The red marks on the chips ease crystal location when cross-polarized light is not available.
comparison purposes. Mosaicity and B factors were extracted from XDSAPP or from AIMLESS in all cases. The most compelling data sets are included in the automatic data processing described below. All other data sets can be found in the Supplementary Tables.

Multiple small wedges (each of 40–45°) were required to be merged and scaled together to give a complete data set, which in turn was used for structure determination. Thereby, isomorphous and statistically equivalent data sets were selected and merged using the automated workflow sxdm (SSX data merging), which is described in detail elsewhere (Basu et al., 2019). The sxdm package provides automatic scaling and merging of subsets, allowing the identification of the sets of subsets that result in the final merged data with the best quality. There were three different steps for data selection in the sxdm suite. Firstly, the data sets were selected based on ISa values \[I/\sigma(I)_{\text{asymptotic}}\] as reported from XSCALE (Kabsch, 2010), followed by two independent hierarchical cluster analyses: (i) based on unit-cell parameters and (ii) based on pair-CC (the cross-correlation between any two data sets). During the ISa-based selection of data sets, a cutoff value of 3.0 for \(I/\sigma(I)\) was applied. In this work, ISa-selected data sets were subjected to hierarchical cluster analysis of pairwise cross-correlation (pair-CC). CC values based on symmetrically equivalent common reflections for each possible pair of a given number of data sets were calculated in sxdm. Subsequently, these pair-CC values were clustered hierarchically, resulting in a dendrogram (data not shown). The most populated cluster was selected based on a cutoff value of 0.8. Finally, data sets selected based on the most populated cluster of the pair-CC selection method were scaled together using XSCALE and used for structure determination. The resolution cutoff for all cases was decided mainly based on a CC\(_{1/2}\) cutoff of 0.3.

2.4. Structure determination

All structural models were determined by molecular replacement in Phaser (Bunkóczi et al., 2013) using as initial models the following clean (no waters or ligands) PDB entries: 1lee (Sauter et al., 2001), 1thw (Ko et al., 1994) and 5i7g (J. Sandy, unpublished work) for lysozyme, thaumatin and glucose isomerase, respectively. Model refinement was finalized with phenix.refine (Afonine et al., 2012), and Coot (Emsley et al., 2010) was used for inspection and manual reconstruction. Model quality was assessed using MolProbity (Chen et al., 2010) implemented within the Phenix suite (Liebschner et al., 2019). The reflection files used for refinement were the output of sxdm. In the case of the thaumatin crystals obtained in the Kapton chip, which resulted in only four data sets, manual merging of the data using AIMLESS (Evans & Murshudov, 2013) was performed. Two data sets were sufficient to obtain a complete data set. Model graphics, including electron-density maps, were prepared using PyMOL (version 1.3r1; Schrodinger).

Coordinates and structure factors have been deposited in the PDB with accession codes 6ybf, 6ybi, 6ybo, 6ybr, 6ybx and 6yc5.

3. Results

Counter-diffusion crystallization techniques are based on facilitating the contact of a protein solution and a precipitating agent in a convention-free environment, such as in microgravity, gels, capillaries or microchannels. Under these conditions, a supersaturation gradient is formed from the precipitant incorporation point to the opposite end of the microchannel. This gradient screens the phase diagram from higher to lower values of supersaturation, producing fewer crystals of larger size as we follow the direction of the gradient along the channel (Otálora et al., 2009). Besides the screening advantage of this technique, it also concomitantly allows the gentle incorporation of other substances such as ligands, cryoprotectants, heavy atoms (Schieferstein et al., 2018) etc., providing (i) high stability of the media for crystal transportation (i.e. the high concentration of precipitant at the equilibrium provides a very stable environment for the crystals that can even withstand changes in temperature etc.) and (ii) the sequence of events occurring along the protein chamber serves as a record of the evolution of the experiment. Indeed, the chips containing the crystals of the three proteins were transported personally, with no special care, in a hand-luggage bag directly to beamline ID30B at ESRF, Grenoble, France prior to data collection.

Counter-diffusion microfluidic chips have already been produced and tested, but in all cases the materials used to fabricate the channels have strongly influenced the final attainable resolution. ChipX3 is the best example of a benchmarked chip designed for in situ characterization, and is optimized to achieve a good compromise between material rigidity and X-ray absorption/scattering (de Wijn et al., 2019). In this study, we aimed to produce counter-diffusion microfluidic chips fabricated with two highly X-ray-transparent materials: Kapton and Mylar. We tested them with three different model proteins, lysozyme, thaumatin and glucose isomerase, in an attempt to reach the maximum diffraction resolution from the crystals with a minimum background from the material. As expected, we obtained crystals of all three proteins along the microfluidic channels. In many cases the crystals fully filled the width of the channel (250 μm) and in some cases they continued growing to reach a dimension of greater than 1 mm, facilitating helical data collection. We observed this behavior for lysozyme and glucose isomerase, although in the case of thaumatin the crystals were on average smaller.

As shown in Fig. 3, each pair of chips (Kapton/Mylar) for each protein was mounted in a home-made frame holder using the plate-gripper mounted directly on the beamline rotation axis. Meanwhile, an aluminium metal holder for two chips has been designed and prepared at the ESRF for easy data collection at ID30B. It is mounted in the plate gripper (McCarthy et al., 2018) and is compatible with both the transversal movements.
**Table 2**

Data and refinement statistics for each deposited structural model: glucose isomerase (Glucy), lysozyme (Lzm) and thaumatin (Thau) obtained in Kapton (K) or Mylar (M) chips.

<table>
<thead>
<tr>
<th>Data set</th>
<th>Glucy (K)</th>
<th>Glucy (M)</th>
<th>Lzm (K)</th>
<th>Lzm (M)</th>
<th>Thau (K)</th>
<th>Thau (M)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of data sets</td>
<td>11</td>
<td>14</td>
<td>3</td>
<td>4</td>
<td>2†</td>
<td>19</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>41.12–1.06</td>
<td>46.99–1.20</td>
<td>35.38–1.13</td>
<td>39.58–1.12</td>
<td>35.48–1.35</td>
<td>41.42–1.14</td>
</tr>
<tr>
<td>(1.09–1.06)</td>
<td>(1.24–1.20)</td>
<td>(1.17–1.13)</td>
<td>(1.16–1.12)</td>
<td>(1.40–1.35)</td>
<td>(1.18–1.14)</td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P22</td>
<td>P22</td>
<td>P42,2</td>
<td>P42,2</td>
<td>P42,2</td>
<td>P42,2</td>
</tr>
<tr>
<td>a, c (Å)</td>
<td>93.94, 103.04</td>
<td>93.99, 103.1</td>
<td>79.11, 38.02</td>
<td>79.17, 37.96</td>
<td>58.38, 151.45</td>
<td>58.57, 151.52</td>
</tr>
<tr>
<td>Total No. of reflections</td>
<td>356629</td>
<td>349672</td>
<td>231847</td>
<td>3018386</td>
<td>237303 (3906)</td>
<td>2846737</td>
</tr>
<tr>
<td>No. of unique reflections</td>
<td>200000 (8688)</td>
<td>149874 (14856)</td>
<td>44967 (3710)</td>
<td>39456 (1152)</td>
<td>52807 (4028)</td>
<td>89815 (5094)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>7.93</td>
<td>8.86</td>
<td>11.56</td>
<td>20.14</td>
<td>4.4</td>
<td>31.69</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>92.32 (40.20)</td>
<td>99.95 (99.72)</td>
<td>98.12 (82.08)</td>
<td>84.57 (25.13)</td>
<td>90.14 (70.00)</td>
<td>92.34 (52.26)</td>
</tr>
<tr>
<td>Mean I/σ(I)</td>
<td>8.4 (1.14)</td>
<td>17.0 (0.9)</td>
<td>12.8 (0.7)</td>
<td>16.6 (1.2)</td>
<td>11.4 (1.2)</td>
<td>16.9 (0.64)</td>
</tr>
<tr>
<td>Wilson B factor (Å²)</td>
<td>13.05</td>
<td>16.95</td>
<td>16.46</td>
<td>19.55</td>
<td>15.50</td>
<td>22.5</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>13.0 (74.0)</td>
<td>8.4 (10.7)</td>
<td>10.9 (98.4)</td>
<td>12.5 (31.0)</td>
<td>8.4 (33.6)</td>
<td>14.3 (120)</td>
</tr>
<tr>
<td>CC1/2 (%)</td>
<td>98.7 (61.4)</td>
<td>99.4 (41.0)</td>
<td>99.7 (70.0)</td>
<td>99.5 (55.1)</td>
<td>99.5 (78.6)</td>
<td>99.8 (79.1)</td>
</tr>
<tr>
<td>Reflections in working/test sets</td>
<td>1990589998</td>
<td>1498427490</td>
<td>4491022848</td>
<td>394491973</td>
<td>527382648</td>
<td>896884491</td>
</tr>
<tr>
<td>Rwork (%)</td>
<td>12.30 (36.46)</td>
<td>12.10 (31.04)</td>
<td>13.67 (24.07)</td>
<td>13.32 (32.77)</td>
<td>12.68 (23.19)</td>
<td>13.52 (42.65)</td>
</tr>
<tr>
<td>Rfree (%)</td>
<td>13.10 (37.55)</td>
<td>13.05 (30.09)</td>
<td>15.02 (25.50)</td>
<td>15.25 (34.58)</td>
<td>14.82 (27.05)</td>
<td>15.71 (41.33)</td>
</tr>
<tr>
<td>No. of non-H atoms</td>
<td>Total</td>
<td>3726</td>
<td>3703</td>
<td>1282</td>
<td>1207</td>
<td>1958</td>
</tr>
<tr>
<td></td>
<td>Macromolecules</td>
<td>3439</td>
<td>3403</td>
<td>1168</td>
<td>1113</td>
<td>1811</td>
</tr>
<tr>
<td></td>
<td>Ligands</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Solvent</td>
<td>284</td>
<td>296</td>
<td>109</td>
<td>91</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>Total protein residues</td>
<td>385</td>
<td>385</td>
<td>129</td>
<td>129</td>
<td>207</td>
</tr>
<tr>
<td></td>
<td>R.m.s.d., bond lengths (Å)</td>
<td>0.005</td>
<td>0.005</td>
<td>0.007</td>
<td>0.012</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>R.m.s.d., angles (°)</td>
<td>0.91</td>
<td>0.89</td>
<td>1.02</td>
<td>1.21</td>
<td>1.31</td>
</tr>
<tr>
<td>Ramachandran statistics (%)</td>
<td>Favorable</td>
<td>96.87</td>
<td>96.87</td>
<td>99.21</td>
<td>99.19</td>
<td>98.05</td>
</tr>
<tr>
<td></td>
<td>Allowed</td>
<td>2.87</td>
<td>2.87</td>
<td>0.79</td>
<td>0.81</td>
<td>1.95</td>
</tr>
<tr>
<td></td>
<td>Outliers</td>
<td>0.26</td>
<td>0.26</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>Rotamer outliers (%)</td>
<td>1.12</td>
<td>1.73</td>
<td>4.35</td>
<td>0.80</td>
<td>3.54</td>
</tr>
<tr>
<td></td>
<td>Clashscore</td>
<td>1.02</td>
<td>1.34</td>
<td>3.42</td>
<td>2.25</td>
<td>3.32</td>
</tr>
<tr>
<td></td>
<td>Average B factor (Å²)</td>
<td>Overall</td>
<td>20.18</td>
<td>22.93</td>
<td>24.16</td>
<td>24.00</td>
</tr>
<tr>
<td></td>
<td>Biomolecules</td>
<td>19.27</td>
<td>22.00</td>
<td>22.95</td>
<td>22.89</td>
<td>25.19</td>
</tr>
<tr>
<td></td>
<td>Ligands</td>
<td>15.31</td>
<td>20.57</td>
<td>41.91</td>
<td>40.00</td>
<td>23.60</td>
</tr>
<tr>
<td></td>
<td>Solvent</td>
<td>31.22</td>
<td>33.69</td>
<td>36.32</td>
<td>37.02</td>
<td>39.45</td>
</tr>
<tr>
<td></td>
<td>PDB code</td>
<td>6ybo</td>
<td>6ybr</td>
<td>6ybf</td>
<td>6yce</td>
<td>6ybx</td>
</tr>
</tbody>
</table>

† Merged using **AIMLESS** (Evans & Murshudov, 2013).

ChipX3 (de Wijn et al., 2019) and OSTEMER–Mylar/Kapton chips.

We were able to rotate 45° (±22.5° from the vertical direction) for each data set (Supplementary Movie S1). Besides the inherent limitation of the total rotation angle, we were able to reach an overall completeness of greater than 90%, and of near 90% for the highest resolution shell, regardless of the protein (Supplementary Tables S1–S6). To minimize the radiation damage by spreading the dose, we used the helical procedure along the longest dimension of the crystals whenever possible. Supplementary Figs. S1–S3 show examples of the trajectories. Some of the data sets collected could be considered as a full data set by themselves (owing to the high crystal symmetry) and could be used individually to solve the structure. Of course, this would not be the case for proteins crystallizing in less symmetrical space groups or when data collection suffered from partial loss of the data set. One example of this situation is shown in Supplementary Fig. S4, in which the helical data collection lost the extreme of the crystals in the plane perpendicular to the X-ray beam. In these cases, multiple data collections from different crystals in a single protein channel (or even the full chip) could be used to provide a full data set.

We have applied the data-selection method as implemented in *sxdm* (Basu et al., 2019) in five out of the six cases studied in this work. In the case of glucose isomerase, the *sxdm* program selected 11 small-wedged data sets out of 14 collected sweeps for the Kapton chip and 14 out of 17 collected sweeps for the Mylar chip based on pair-CC-based hierarchical clustering (Basu et al., 2019). For lysozyme crystals, the program selected three out of 11 and four out of 14 small-wedged data sets as statistically equivalent from the Kapton chip and the Mylar chip, respectively. In the case of thaumatin crystals grown in the Mylar chip, the *sxdm* program merged 19 out of 22 collected sweeps based on pair-CC-based hierarchical clustering. The thaumatin crystals contained in the Kapton chip accidentally suffered from dehydration and thus only four data sets were collected, three of which were at a resolution better than 1.5 Å. We used this case as an example for regular data merging (two data sets) using **AIMLESS** (Evans & Murshudov, 2013). In the other five cases we merged data sets from different crystals to generate the final merged reflections file, which was then used to determine the structural models (from 19 data sets in the case of thaumatin crystals grown in the Mylar chip to only three data sets in the case of lysozyme grown in the Kapton chip). Even though our approach is far

756 Gavira et al. • Atomic resolution from in situ data collection

from being considered as a serial crystallography methodology (Tolstikova et al., 2019), the number of crystals obtained per chip allowed us to obtain the best from the multiple data-collection merging protocols, such as those used at XFELs, while keeping the highest attainable resolution. The final statistics for the three model proteins refined with the Phenix suite (Liebschner et al., 2019) are shown in Table 2. We could not detect any influence of the film composition, Kapton or Mylar, on the final quality of the models. We obtained 3D models at a resolution better than 1.2 Å in half of the cases. In all cases, the deposited models were better than any previous reported structures determined from data collected using microfluidic devices (Supplementary Table S7).

4. Conclusion

Counter-diffusion-based crystallization microchips, fabricated with a combination of either OSTEMER and Kapton or of OSTEMER and Mylar, allow direct structure determination at atomic resolution limits from data collected at room temperature. Helical data acquisition along the longest dimension of the crystals, in combination with the high symmetry of the three model-protein crystals, can allow the production of high-resolution 3D models from a single data set. Automatic merging of the best and highly isomorphous subsets of the data sets collected represents by far the best solution. This easy approach describes an inexpensive and effective approximation to serial crystallography experiments. If fabricated with standardized dimensions, these low-cost microchips would permit automated mail-in workflows on synchrotron beamlines to search for crystals along the microfluidic channels and to either collect a full data set from a single crystal or merge several data sets prior to entering a pipeline for further automatic procedures for structural determination of this thin-film microfluidic configuration to approximate time-resolved crystallography to be envisioned.

Acknowledgements

We are very grateful to the European Synchrotron Radiation Facility, Grenoble, France for provision of beam time and we would like to thank all of the beam staff for assistance and support during data collection. We would like to thank Ms R. Fernandez for setting up the crystallization experiments.

Funding information

This study was supported by project BIO2016-74875-P (MINECO), Spain co-funded by the Fondo Europeo de Desarrollo Regional (FEDER funds), European Union.

References


