

Article

EM-*detwin*: A Program for Resolving Indexing Ambiguity in Serial Crystallography Using the Expectation-Maximization Algorithm

Yingchen Shi^{1,2} and Haiguang Liu^{1,3,*}

- ¹ Complex Systems Division, Beijing Computational Science Research Centre, 8 E Xibeiwang Rd, Haidian, Beijing 100193, China; shiyc12@csrc.ac.cn
- ² Department of Engineering Physics, Tsinghua University, 30 Shuangqing Rd, Haidian, Beijing 100084, China
- ³ Physics Department, Beijing Normal University, 19 Xinjiekouwai St, Haidian, Beijing 100875, China
- * Correspondence: hgliu@csrc.ac.cn

Received: 25 May 2020; Accepted: 7 July 2020; Published: 8 July 2020



Abstract: Serial crystallography (SX), first used as an application of X-ray free-electron lasers (XFELs), is becoming a useful method to determine atomic-resolution structures of proteins from micrometer-sized crystals with bright X-ray sources. Because of unknown orientations of crystals in SX, indexing ambiguity issue arises when the symmetry of Bravais lattice is higher than the space group symmetry, making some diffraction signals wrongly merged to the total intensity in twinned orientations. In this research, we developed a program within the *CrystFEL* framework, the *EM-detwin*, to resolve this indexing ambiguity problem based on the expectation-maximization algorithm. Testing results on the performance of the *EM-detwin* have demonstrated its usefulness in correctly indexing diffraction data as a valuable tool for SX data analysis.

Keywords: indexing ambiguity; crystallography; expectation-maximization; XFEL

1. Introduction

Serial crystallography (SX) is a new technology to solve protein structures from many crystals diffracted by X-rays in a serial fashion. SX was initially developed for X-ray free-electron lasers (XFELs), whose femtosecond X-ray pulses destroy crystal samples after a single exposure. The femtosecond X-ray pulses are considered to be short enough, such that the radiation damage to the crystals does not occur during the exposure [1]. Therefore, the term serial femtosecond crystallography (SFX) was coined due to the femtosecond X-ray pulses at XFELs. The SFX has been demonstrated to be a useful method to determine atomic-resolution structures of proteins from tiny crystals in the size of micrometers or even submicrometers [2–6]. The applications of the SX method have also been demonstrated at synchrotron radiation facilities [7–9]. A large number of crystals are required for diffraction measurements, and the signals need to be indexed and merged to obtain a set of full intensity for electron density model building. It is noted that crystals are in unknown orientations during the experiments, because they are either injected to the X-ray interception section, or delivered on a tape, or scanned in their mother liquor. As a result, the crystal orientations are not correlated with serial crystallography. This characteristic brings up the indexing ambiguity issue, which is unique for SX. Here, the indexing ambiguity means that more than one indexing solutions are equally good for a single diffraction pattern to satisfy peak position constraints. This happens when the Bravais symmetry of a crystal lattice is higher than the space group symmetry, and the diffraction pattern from a crystal may be indexed either in the correct orientation or in its twinned orientation(s). Even if no physical twinning occurs to the crystals in SX experiments, merging the data may result in a twined dataset for those crystals with indexing ambiguities. The procedure to solve this problem is called detwinning.



In conventional crystallography, the orientations of the crystals are controlled by goniometers, such that the relative orientation information between diffraction patterns is known. Furthermore, each pair of diffraction patterns from conventional crystallography usually share a good number of common Bragg peaks (or reflections) that are measured with full intensity, which facilitate unambiguous indexing. For each pattern, possible indexing solutions can be tried in turn and the one with the best intensity agreement to a reference pattern is selected as the correct solution for merging. However, the detwinning is more challenging in XFEL nanocrystal crystallography. The number of diffraction peaks in a single pattern is usually smaller in SFX than in conventional crystallography due to the "still" exposure in contrast to the oscillation exposure, and each pattern only records partial intensities for each reflection because of the wide angular spread of Bragg peaks in reciprocal space (due to the small size of crystal samples and micro-focused X-ray beams). In certain SX experiments, a crystal may survive multiple measurements at synchrotrons or with unfocused XFEL X-rays. Under such circumstances, each crystal results in a series of diffraction patterns that are orientationally related, such as in fine φ -slicing measurements [10–12]. This can facilitate the indexing by treating the series of patterns as groups. In general, autoindexing methods, such as MOSFLM [13], DIRAX [14] or LABELIT [15], cannot immediately distinguish between twin-related indexing solutions based on the locations of diffraction peaks alone. In a situation where two indexing solutions are equivalent, without applying detwinning analysis, nearly half of the patterns will be indexed as its twin and be merged incorrectly. As the intensity variation caused by partial reflection is significant, one needs robust algorithms for detwinning analysis on crystallography data with partial reflections.

Several methods have been developed to solve the indexing ambiguity problem. The BD algorithm first proposed by Brehm and Diederichs [16] and several variants [17,18] reduce the dimensions of data recorded in patterns based on pairwise similarities measured using Pearson's correlation and then do clustering analysis. The algorithm has been implemented in *CrystFEL* software [19] as the *ambigator* program. Another method is based on the expectation-maximization (EM) algorithm to iteratively improve the indexing consistency by maximizing the similarities between individual patterns and the merged reflection model [20]. Different from the BD algorithm, which evaluates the pairwise distances between diffraction patterns that record partial reflections, the EM algorithm utilizes the relationship between reflections from any pattern and the full reflection model merged in the previous iteration. The convergence of the indexing solutions and the merged reflection model will yield a solution that is optimally consistent with data in the whole dataset. In this paper, we present the implemented EM algorithm within the *CrystFEL* framework as an add-on program called *EM-detwin*. The program was tested with three experimental datasets in two distinct space groups to assess the quality of the merged results.

2. Materials and Methods

2.1. The Implementation of the EM-detwin

The *EM*-detwin algorithm is an iterative reconstruction algorithm and uses the Pearson's correlation coefficients between diffraction patterns and the reflection model obtained in the previous iteration as metrics to assign the crystal orientations, such that the associated diffraction patterns can be indexed consistently. The Pearson correlation between a diffraction pattern I_i (the term "crystal" will be used in the following text because there may be diffraction signals from multiple crystals recorded in a single pattern) and a model consisting of full reflections I_{full} defined as the following [20]:

$$r_{i} = \frac{\sum_{h} \left[I_{i}(h) - \bar{I}_{i} \right] \left[I_{full}(h) - \bar{I}_{full} \right]}{\left\{ \sum_{h} \left[I_{i}(h) - \bar{I}_{i} \right]^{2} \sum_{h} \left[I_{full}(h) - \bar{I}_{full} \right]^{2} \right\}^{1/2}}$$
(1)

where I_i is the diffraction intensities recorded in the crystal *i*, with $\{h\}$ as the miller indices of common reflections and r_i is the correlation coefficient corresponding to the diffraction data from crystal *i*. The mean values for the common dataset are denoted as the \overline{I}_i and \overline{I}_{full} .

The algorithm takes the indexing solutions from *CrystFEL* stream file as the input, and carries out the first round of merging without detwinning to get the initial model of the merged full reflections, the M(0). In the subsequent iterations, the merged reflection model is improved by updating the orientations of crystals based on the comparison between crystals and previously merged full reflection models. Specifically, in the *k*th iteration, the diffraction signals from the *i*th crystal { I_i } is transformed to its equivalent indexed intensities { I_i^t }, where *t* enumerates all of the possible twinning operators of corresponding space groups (i.e., the miller indices are updated based on the symmetry operations). Then, the correlation coefficients { r_i^t } are calculated between { I_i^t } and the merged full reflection model obtained in the previous iteration, the M(k-1). Based on the correlation coefficients, a new model M(k) is obtained by merging the data in the updated orientations. There are two merging strategies, one is the "winner-takes-all" and the other is the "weighted-merging". For the "winner-takes-all" mode, the winner twinning operator t_0 for crystal { I_i } is obtained by max_{ r_i^t }, and only { $I_i^{t_0}$ } is merged

to M(k). In the "weighted merging" strategy, the set of $\{I_i^t\}$ are merged to M(k) in a weighted manner, where the weights are calculated in the following way:

$$w_i^t = \frac{r_i^t}{\sum_t r_i^t} \tag{2}$$

Since each crystal can only take a single orientation in reality, at the final iteration, all crystals are merged in the "winner-takes-all" mode to get merged intensities.

The *EM-detwin* program is designed in a similar way as the *process_hkl* program in the *CrystFEL* software. Besides the final merged model from the whole dataset, the *EM-detwin* also generates the merged models from half datasets (even vs. odd numbered patterns) for consistency analysis. The new input parameters for *EM-detwin* are listed in Table 1, while other parameters are the same as the *process_hkl* program.

Parameters	Explanation
<pre>-spacegroupnum = <n></n></pre>	The space group number, used to determine twinning operators.
_winner_takes_all	If set, the program will use "winner-takes-all" mode to do merging; otherwise,
whiter takes an	"weighted-merging" will be used for intermediate model merging
<pre>-highres = <high></high></pre>	Reject reflections with resolution higher than <i>high</i> Å when calculating correlation coefficients.
-lowres = < low >	Reject reflections with resolution lower than low Å when calculating correlation coefficients.
-write-assignments = <file></file>	Write the final re-indexed results to <i>file</i> .
-add-operators = $<$ op $>$	Add specific twin operators.

Table 1. New input parameters for the *EM-detwin* program.

2.2. Test Datasets

Three datasets were used to assess the performance of the *EM-detwin* program, based on the availability of the diffraction data. The datasets include: (1) the HIV-1 Integrase Catalytic Core Domain (IN-CCD) data [5] collected in 2018 at the PAL-XFEL (Pohang, South Korea); (2) the Beta-lactamase (BLAC) data [21] collected in 2018 at the European-XFEL (Hamburg, Germany); and (3) the bacteriorhodopsin (BR) data [22] collected in 2017 at the LCLS (Stanford, CA, USA). The symmetry information is summarized in Table 2. The BLAC dataset was downloaded from the CXIDB [23] with the CXIDB id 83. The other two datasets were generously provided by the authors of each work. The input data are *CrystFEL* stream files generated by indexamajig before being processed by the *ambigator*. There are 27,311, 12,474 and 241,475 indexed crystals in IN-CCD, BLAC and BR stream files, respectively. The BR dataset is from a pump-probe SFX experiment with extraordinarily high redundancy in measurements that is not common in SX. In order to assess the performance for a typical

sized SX experimental dataset, the test was done with the first 20,000 chunks of the BR stream file, containing 15,847 indexed crystals.

Dataset	Space Group	Bravais Symmetry	Twinning Operators
IN-CCD	P3 ₁ 21	321_H	(-h, -k, l), (-k, -h, -l) *
BLAC	P3 ₂ 21	321_H	(-h, -k, l), (-k, -h, -l) *
BR	P63	6/m	(k, h, −l)

Table 2. Symmetry information of the test datasets.

* The two twinning operators lead to equivalent miller indices.

The atomic coordinates for these three proteins were downloaded from the RCSB Protein data bank (PDB) [24]. The PDB codes for the IN-CCD, BLAC and BR are 6JCG, 6GTH and 6G7H, respectively. These atomic models were used for molecular replacement phasing and subsequent model refinement with the phenix software (version 1.11.1) [25].

3. Results

3.1. The Convergence of EM-detwin Program

The three datasets were subject to the EM-detwin analysis to get the merged reflections. We applied 30 iterations to test the performance with two merging strategies independently. The merged data from "weighted" and "winner-takes-all" strategies are very similar, with the overall R-factors (see footnote of Table 3 for definition) of 0.36%, 0.32% and 0.25% for IN-CCD, BLAC and BR datasets, respectively. The small R-factors suggest that the differences between the merged results using these two strategies are essentially the same. Therefore, we used the results from the "weighted" strategy for the rest of the analysis.

Comparison	Overall R-Factor #/CC _{1/2} ⁺ /CC _{star} *			
Comparison –	IN-CCD	BLAC	BR	
ambigator vs. EM-detwin	14.21%/	9.47%/	8.77%/	
	0.935/0.983	0.989/0.997	0.999/0.999	
ambigator vs. twinned	42.09%/	32.60%/	35.02%/	
	0.575/0.854	0.852/0.959	0.905/0.975	
EM-detwin vs. twinned	37.73%/	33.23%/	30.92%/	
	0.489/0.810	0.846/0.957	0.915/0.978	

Table 3. Comparison between merged data from different algorithms.

[#]*R*-factor = $\frac{\Sigma ||F_1| - k|F_2||}{\Sigma ||F_1||}$, where k is the scaling factor between the two datasets. [†]*CC*_{1/2} is Pearson correlation coefficient. ^{*}*CC*_{star} = $\sqrt{\frac{2CC_{1/2}}{1+CC_{1/2}}}$

The convergence was examined by calculating the correlation coefficients between the final reflection model and the intermediately merged models at each iteration, as shown in Figure 1. Although the convergence speed varies among the three test dataset, the curves suggest that 30 iterations are sufficient for the merged model to converge. We also observed that the "winner-takes-all" strategy can converge faster. To visually assess the results of the *EM-detwin* program, we compared the merged models from original stream files (before detwinning) and the *EM-detwin* results. The reflections on planes with twinning symmetry were visualized using the CCP4 software [26], showing that the false symmetry disappeared after the analysis by the *EM-detwin* (Figure 2). This demonstrates that the *EM-detwin* program is capable of removing the extra symmetries due to twin-related orientations in all three datasets.



Figure 1. Correlation coefficients between intermediate and final merged models. The solid lines show the data with the "weighted" merging strategy, and the dashed lines show the data with the "winner-takes-all" (WTA) strategy.



Figure 2. Differences between merged diffraction intensities before and after being processed by *EM-detwin.* (**a**,**b**) Diffraction intensities on the (h,0,l) plane of the Integrase Catalytic Core Domain (IN-CCD), where the green dashed line represents the twinning symmetry "(h,0,l) = (-h,0,l)"; (**c**,**d**) (h,k,5) plane of Beta-lactamase (BLAC) reflections, where the green arrows indicate the twinning symmetry "(h,k,5) = (-h,-k,5)" (central symmetry in planes for fixed *l*); (**e**,**f**) (h,k,0) plane of BR reflections, where green dashed line represents the twinning symmetry "(h,k,0) = (k,h,0)" (rotational symmetry).

3.2. Merging Statistics Comparison

The pairwise R-factors were computed for three merged results: *ambigator*, *EM-detwin* and the twinned data. No resolution limits were set in the R-factor calculations. The results in Table 3 show that *EM-detwin* and *ambigator* results are very similar, and they are both significantly different from the twinned data. This is consistent with the quantification using the correlations, in terms of $CC_{1/2}$ and CC_{star} . The quality of the merged results is also assessed using the R-split values, which measures the differences between merged models from two subsets composed of even/odd numbered datasets, as shown in Table 4. The resolutions of published PDB structures were used as the cut-off values for R-split factor calculations. The R-split comparison results indicate that the *EM-detwin* and the *ambigator* are comparable in terms of self-consistency.

Dataset	Algorithm	Resolution Cut-Of	Overall R-Split #	CC _{1/2} ⁺ /CC _{star} *
IN-CCD	ambigator EM-detwin	2.5Å	9.26% 9.11%	0.989/0.997 0.996/0.999
BLAC	ambigator EM-detwin	1.7Å	29.68% 28.59%	0.853/0.959 0.872/0.965
BR	ambigator EM-detwin	1.5Å	10.33% 10.24%	0.985/0.996 0.985/0.996

Table 4. Merging consistency comparison.

 $\# R-\text{split} = \frac{1}{\sqrt{2}} \frac{\sum ||F_1| - k|F_2||}{0.5 \sum ||F_1| + k|F_2||}, \text{ where } k \text{ is the scaling factor between the two datasets. } \dagger, * \text{ the definitions of } CC_{1/2}/CC_{\text{star}}$ are the same as in Table 3.

3.3. Structure Determination and Refinement with Detwinned Data

The merged data after detwinning was used to build atomic models and refine protein structures. The *phenix.phaser* program was used to build the initial model using a molecular replacement method. Then the model was refined using the *phenix.refine* program. The default parameters in *phenix.phaser and phenix.refine* were applied. The resolution limits were set to be the same as in the published structures. Merged data from original stream files without detwinning were used for structure refinement as control groups. The final R-work and R-free values of model refinement are listed in Table 5 (see Supplementary Table S1 for the complete statistics of structure refinement). The electron density maps are shown in Figure 3 for the three structures, which are calculated based on the merged data by the *EM-detwin* program.

It is shown that for all datasets, both detwinning programs improve the refinement model quality, in contrast to the large R-factors resulting from the twinned data. The comparison also suggests that *ambigator* and *EM-detwin* have similar performance on the qualities of structure determination and refinement.

Dataset	Algorithm	Resolution	R-Work	R-Free
IN-CCD	ambigator EM-detwin None	2.5Å	0.1840 0.1828 0.2744	0.2202 0.2230 0.3351
BLAC	ambigator EM-detwin None	1.7Å	0.2308 0.2324 0.2993	0.2657 0.2620 0.3446
BR	ambigator EM-detwin None	1.5Å	0.1913 0.1927 0.2792	0.2116 0.2120 0.3077

Table 5. Model refinement results from *phenix* with default parameters.



Figure 3. Refined protein structures and electron density maps using merged intensity data from the *EM-detwin* program. (**a**) the IN-CCD structure, (**b**) the BLAC structure and (**c**) the bacteriorhodopsin (BR) structure. The 2Fo-mFc maps were shown at σ = 2.0.

3.4. Robustness of the Program

The EM-*detwin* program was evaluated for cases where data redundancy is low. We used the first 5000 crystals from each dataset as a reduced dataset to carry out the detwinning analysis. Using the indexing solutions obtained with the full dataset as a reference, the percentage of consistently indexed crystals was computed for the case with the reduced dataset. The similarity between the final models from the full and the reduced datasets was measured using the correlation coefficients. The results are summarized in Table 6. The performance of the *ambigator* program was also assessed with the reduced datasets in parallel for comparison. The data suggest that both programs yielded to the merged intensity models that are consistent with the results from the full datasets. However, the performance of the EM-*detwin* with the "winner-takes-all" merging strategy is not as good as the *ambigator* program in the case of reduced datasets, especially for the BLAC dataset. Based on this result, the "winner-takes-all" mode is not recommended for small datasets, as it is a more aggressive strategy which converges faster (see Figure 1) but with a risk of converging to local optimal points.

Algorithms -	Indexing Consistency			Correlation Coefficient		
	IN-CCD	BLAC	BR	IN-CCD	BLAC	BR
ambigator	99.28%	86.18%	99.72%	0.9854	0.9332	0.9905
EM-detwin	96.12%	88.12%	98.79%	0.9275	0.9179	0.9895
EM-detwin (WTA *)	95.64%	57.90%	98.58%	0.9280	0.7881	0.9895

Table 6. Performance of the detwinning algorithms for reduced datasets.

* WTA indicates the "winner-takes-all" merging strategy.

The EM-detwin algorithm has a computational complexity of O(MN), with N as the number of diffraction patterns and M as the number of iterations. The *ambigator* program needs to compute the pairwise similarities between all patterns, so it has a computational complexity of $O(N^2)$. Therefore, the EM-detwin program shall have speed advantages when N is big. This is confirmed by the execution time comparison for the three datasets. For smaller datasets of BR and BLAC, the EM-*detwin* program needs to run for 1260 s and 603 s respectively. It takes about a similar time for the *ambigator* program (3397 s and 649 s for BR and BLAC datasets respectively) to complete the detwinning analysis. For the IN-CCD dataset composed of 23,711 crystals, the EM-*detwin* execution time is 2150 s, faster than the ambigator that requires 11,873 s. All tests were done with a Macbook Pro A1706 computer utilizing a single 2.9 GHz Core i5 CPU. Therefore, we recommend using the EM-*detwin* program for larger dataset analysis.

In the case that partial intensities are measured due to the experimental limitations (such as monochromatic X-rays, ultrashort exposure time, crystal mosaicity, etc.), a post-refinement procedure may improve the quality of merged signals [27–29]. The post-refinement on partial reflections is a separate procedure from the detwinning, therefore we resort it to a specialized algorithm to handle it.

To facilitate the further analysis of post-refinement, the EM-detwin program has the option to save the re-indexed results in the format of CrystFEL stream file to allow partiality modeling and corrections. To test the impact of post-refinement on the three test datasets, we applied the *partialator* program in CrystFEL on the detwinned data using the xsphere model [27] and found small improvements in the model R-factors for these three datasets (see Table S1).

4. Conclusions

As the SX method is becoming more advanced, it attracts broader applications, providing an alternative for the conventional crystallography. We expect structures of proteins packed in various space groups to be determined using this approach. The indexing ambiguity issue needs to be resolved in order for this method to be useful for the crystals belonging to the space groups that possess twinning symmetries. In this research, we implemented an expectation-maximization algorithm-based program, the *EM-detwin*, to address this issue. Testing results on experimental data show that the diffraction patterns can be correctly indexed and merged for model building and structure refinement. The program of *EM-detwin* has been tested within the framework of *CrystFEL* software for versions 0.6.3, 0.7.0, 0.8.0 and 0.9.0. The source codes and installation guide are available on Github (https://github.com/LiuLab-CSRC/detwin).

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4352/10/7/588/s1, Table S1: Data collection and refinement statistics.

Author Contributions: Conceptualization, H.L.; methodology, Y.S. and H.L.; software, Y.S.; validation, Y.S., writing Y.S. and H.L.; funding acquisition, H.L. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by the funding from National Natural Science Foundation of China (grant numbers: 31971136, 11811540392, U1930402).

Acknowledgments: The authors are grateful to Tobias Weinert from Paul Scherrer Institute (PSI) for providing the BR test dataset.

Conflicts of Interest: The authors declare no conflict of interest.

References

- 1. Neutze, R.; Wouts, R.; van der Spoel, D.; Weckert, E.; Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **2000**, *406*, 752–757. [CrossRef] [PubMed]
- Chapman, H.N.; Fromme, P.; Barty, A.; White, T.A.; Kirian, R.A.; Aquila, A.; Hunter, M.S.; Schulz, J.; DePonte, D.P.; Weierstall, U.; et al. Femtosecond X-ray protein nanocrystallography. *Nature* 2011, 470, 73–77. [CrossRef]
- Liu, W.; Wacker, D.; Gati, C.; Han, G.W.; James, D.; Wang, D.; Nelson, G.; Weierstall, U.; Katritch, V.; Barty, A.; et al. Serial femtosecond crystallography of G protein-coupled receptors. *Science* 2013, 342, 1521–1524. [CrossRef]
- Kupitz, C.; Basu, S.; Grotjohann, I.; Fromme, R.; Zatsepin, N.A.; Rendek, K.N.; Hunter, M.S.; Shoeman, R.L.; White, T.A.; Wang, D.; et al. Serial time-resolved crystallography of photosystem II using a femtosecond X-ray laser. *Nature* 2014, *513*, 261–265. [CrossRef]
- Park, J.-H.; Yun, J.-H.; Shi, Y.; Han, J.; Li, X.; Jin, Z.; Kim, T.; Park, J.; Park, S.; Liu, H.; et al. Non-Cryogenic Structure and Dynamics of HIV-1 Integrase Catalytic Core Domain by X-ray Free-Electron Lasers. *Int. J. Mol. Sci.* 2019, 20, 1943. [CrossRef] [PubMed]
- Yun, J.H.; Li, X.; Park, J.H.; Wang, Y.; Ohki, M.; Jin, Z.; Lee, W.; Park, S.Y.; Hu, H.; Li, C.; et al. Non-cryogenic structure of a chloride pump provides crucial clues to temperature-dependent channel transport efficiency. *J. Biol. Chem.* 2019, 294, 794–804. [CrossRef] [PubMed]
- Nogly, P.; James, D.; Wang, D.; White, T.A.; Zatsepin, N.; Shilova, A.; Nelson, G.; Liu, H.; Johansson, L.; Heymann, M.; et al. Lipidic cubic phase serial millisecond crystallography using synchrotron radiation. *IUCrJ* 2015, 2, 168–176. [CrossRef] [PubMed]
- 8. Rossmann, M.G. Serial crystallography using synchrotron radiation. *IUCrJ* 2014, *1*, 84–86. [CrossRef] [PubMed]

- 9. Grünbein, M.L.; Kovacs, G.N. Sample delivery for serial crystallography at free-electron lasers and synchrotrons. *Acta Crystallogr. Sect. D Struct. Biol.* **2019**, *75*, 178–191.
- Weinert, T.; Olieric, N.; Cheng, R.; Brünle, S.; James, D.; Ozerov, D.; Gashi, D.; Vera, L.; Marsh, M.; Jaeger, K.; et al. Serial millisecond crystallography for routine room-temperature structure determination at synchrotrons. *Nat. Commun.* 2017, *8*, 542. [CrossRef]
- Wierman, J.L.; Paré-Labrosse, O.; Sarracini, A.; Besaw, J.E.; Cook, M.J.; Oghbaey, S.; Daoud, H.; Mehrabi, P.; Kriksunov, I.; Kuo, A.; et al. Fixed-target serial oscillation crystallography at room temperature. *IUCrJ* 2019, *6*, 305–316. [CrossRef] [PubMed]
- 12. de la Mora, E.; Coquelle, N.; Bury, C.S.; Rosenthal, M.; Holton, J.M.; Carmichael, I.; Garman, E.F.; Burghammer, M.; Colletier, J.-P.; Weik, M. Radiation damage and dose limits in serial synchrotron crystallography at cryo- and room temperatures. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 4142–4151. [CrossRef]
- 13. Leslie, A.G.W. Integration of macromolecular diffraction data. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 1999, 55, 1696–1702. [CrossRef] [PubMed]
- 14. Duisenberg, A.J.M. Indexing in single-crystal diffractometry with an obstinate list of reflections. *J. Appl. Crystallogr.* **1992**, *25*, 92–96. [CrossRef]
- 15. Sauter, N.K.; Grosse-Kunstleve, R.W.; Adams, P.D. Robust indexing for automatic data collection. *J. Appl. Crystallogr.* **2004**, *37*, 399–409. [CrossRef]
- Brehm, W.; Diederichs, K. Breaking the indexing ambiguity in serial crystallography. *Acta Crystallogr. Sect. D* 2014, 70, 101–109. [CrossRef]
- 17. Gildea, R.J.; Winter, G. Determination of Patterson group symmetry from sparse multi-crystal data sets in the presence of an indexing ambiguity. *Acta Crystallogr. Sect. D Struct. Biol.* **2018**, *74*, 405–410. [CrossRef]
- Diederichs, K. Dissecting random and systematic differences between noisy composite data sets. *Acta Crystallogr.* Sect. D Struct. Biol. 2017, 73, 286–293. [CrossRef]
- 19. White, T.A.; Kirian, R.A.; Martin, A.V.; Aquila, A.; Nass, K.; Barty, A.; Chapman, H.N. CrystFEL: A software suite for snapshot serial crystallography. *J. Appl. Crystallogr.* **2012**, *45*, 335–341. [CrossRef]
- 20. Liu, H.; Spence, J.C.H. The indexing ambiguity in serial femtosecond crystallography (SFX) resolved using an expectation maximization algorithm. *IUCrJ* 2014, *1*, 393–401. [CrossRef]
- 21. Wiedorn, M.O.; Oberthür, D.; Bean, R.; Schubert, R.; Werner, N.; Abbey, B.; Aepfelbacher, M.; Adriano, L.; Allahgholi, A.; Al-Qudami, N.; et al. Megahertz serial crystallography. *Nat. Commun.* **2018**, *9*, 4025. [CrossRef]
- Nogly, P.; Weinert, T.; James, D.; Carbajo, S.; Ozerov, D.; Furrer, A.; Gashi, D.; Borin, V.; Skopintsev, P.; Jaeger, K.; et al. Retinal isomerization in bacteriorhodopsin captured by a femtosecond x-ray laser. *Science (80-)* 2018, 361, eaat0094. [CrossRef]
- 23. Maia, F.R.N.C. The Coherent X-ray Imaging Data Bank. Nat. Methods 2012, 9, 854–855. [CrossRef] [PubMed]
- 24. Rose, P.W.; Bi, C.; Bluhm, W.F.; Christie, C.H.; Dimitropoulos, D.; Dutta, S.; Green, R.K.; Goodsell, D.S.; Prlić, A.; Quesada, M.; et al. The RCSB Protein Data Bank: New resources for research and education. *Nucleic Acids Res.* **2013**, *41*, 475–482. [CrossRef] [PubMed]
- Liebschner, D.; Afonine, P.V.; Baker, M.L.; Bunkoczi, G.; Chen, V.B.; Croll, T.I.; Hintze, B.; Hung, L.W.; Jain, S.; McCoy, A.J.; et al. Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol.* 2019, 75, 861–877. [CrossRef]
- Winn, M.D.; Ballard, C.C.; Cowtan, K.D.; Dodson, E.J.; Emsley, P.; Evans, P.R.; Keegan, R.M.; Krissinel, E.B.; Leslie, A.G.W.; McCoy, A.; et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2011, 67, 235–242. [CrossRef] [PubMed]
- Ginn, H.M.; Brewster, A.S.; Hattne, J.; Evans, G.; Wagner, A.; Grimes, J.M.; Sauter, N.K.; Sutton, G.; Stuart, D.I. A revised partiality model and post-refinement algorithm for X-ray free-electron laser data. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2015, *71*, 1400–1410. [CrossRef]
- White, T.A. Post-refinement method for snapshot serial crystallography. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 2014, 369, 20130330. [CrossRef]
- Sauter, N.K. XFEL diffraction: Developing processing methods to optimize data quality. J. Synchrotron Radiat. 2015, 22, 239–248. [CrossRef]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).