

AutoLEI: An XDS-based pipeline with graphical user interface for automated real-time and offline batch 3D ED/microED data processing

Lei Wang,^{a,‡} Yinlin Chen,^{a,‡} Emma Scaletti Hutchinson,^b Pål Stenmark,^b Gerhard Hofer,^{a,*} Hongyi Xu^{a,c,*} and Xiaodong Zou^{a,d,*}

Received 22 September 2025

Accepted 2 December 2025

Edited by J. Sun, Peking University, People's Republic of China

‡ These authors contributed equally to this work.

Keywords: 3D electron diffraction; 3D ED; microcrystal electron diffraction; microED; electron crystallography; real-time data processing; offline batch data processing; data analysis; beam-sensitive materials.

CCDC references: 2512820; 2517260

Supporting information: this article has supporting information at www.iucrj.org

^aDepartment of Chemistry, Stockholm University, 106 91 Stockholm, Sweden, ^bDepartment of Biochemistry and Biophysics, Stockholm University, 106 91 Stockholm, Sweden, ^cResearch School of Chemistry, Australian National University, Acton, Australia, and ^dWallenberg Initiative Materials Science for Sustainability, Stockholm University, SE-106 91 Stockholm, Sweden. *Correspondence e-mail: gerhard.hofer@su.se, hongyi.xu@su.se, xiaodong.zou@su.se

Three-dimensional electron diffraction (3D ED), also known as microcrystal electron diffraction (microED), is an emerging method for determining structures from submicron-sized crystals. With the development of rapid and convenient data collection protocols, acquiring dozens of datasets in a single 3D ED/microED session has become routine. A fast and automated workflow for processing, scaling and merging a large number of 3D ED/microED datasets can significantly accelerate the structure determination process. Herein, we present an XDS-based pipeline with a graphical user interface for automated real-time and offline batch 3D ED/microED data processing (*AutoLEI*). We demonstrate the functionality and applications of the pipeline through four examples, using both offline and real-time data processing capabilities. The samples include small organic molecules, metal–organic frameworks (MOFs) and proteins, showcasing the versatility and efficiency of *AutoLEI* in various applications.

1. Introduction

Three-dimensional electron diffraction (3D ED), also known as microcrystal electron diffraction (microED), has been developed into a robust method for crystal structure determination over the last few decades (Kolb *et al.*, 2007; Zhang *et al.*, 2010; Shi *et al.*, 2013; Gemmi *et al.*, 2019). The method enables rapid and accurate structure determination using commercially available transmission electron microscopes (TEMs). A wide range of structures, including zeolites (Guo *et al.*, 2015; Cho *et al.*, 2023; Lu *et al.*, 2024), metal–organic frameworks (MOFs) (Huang, Willhammar & Zou, 2021; Huang, Grape *et al.*, 2021), pharmaceuticals (van Genderen *et al.*, 2016; Lightowler *et al.*, 2022) and macromolecules (Shi *et al.*, 2013; Xu *et al.*, 2019; Martynowycz *et al.*, 2022), have been successfully solved with 3D ED/microED.

Benefiting from the strong interaction between electrons and matter, 3D ED/microED data can be collected from crystals that are too small for single-crystal X-ray diffraction (Tremlett *et al.*, 2025). An ideal TEM grid typically contains hundreds of diffracting crystals or crystal fragments. With the latest 3D ED/microED protocol, it is possible to determine structures of high-symmetry stable crystals through a single dataset within a few minutes (Cichočka *et al.*, 2018; Plana-Ruiz *et al.*, 2020; Ito *et al.*, 2021; Unge *et al.*, 2024). However, owing to the inherent limitations of the goniometer, data merging for crystals of low symmetry is usually required to achieve high data completeness. For highly beam sensitive material, a small-wedge data collection and merging strategy can also be



OPEN ACCESS

Published under a CC BY 4.0 licence

employed to increase the signal-to-noise ratio and data completeness (Clabbers *et al.*, 2021). Furthermore, selectively merging 3D ED/microED datasets can improve data quality by increasing redundancy (Xu *et al.*, 2018; Samperisi *et al.*, 2021; Bengtsson *et al.*, 2022). More recently, high-throughput data collection has been achieved using software such as *Instamatic* (Cichocka *et al.*, 2018), *Automated EM Data Acquisition with SerialEM* (de la Cruz *et al.*, 2019), *Leginon* (Cheng *et al.*, 2021), *EPUD* (Thermo Fisher Scientific, TFS) and *Latitude D* (GATAN) or a dedicated electron diffractometer, such as Synergy-ED (Rigaku) and ELDICO ED-1 (ELDICO Scientific). It is feasible to perform qualitative phase analysis by collecting and processing well populated datasets (Luo *et al.*, 2023; Unge *et al.*, 2024; Lightowler *et al.*, 2024). In all the above cases, an efficient, automated and even real-time data processing solution is highly desirable to enhance the efficiency of structure determination and phase analysis using 3D ED/microED.

A large selection of well established software, such as X-ray detector software (*XDS*) (Kabsch, 2010), *PETS2* (Palatinus *et al.*, 2019), *CrysAlisPro-ED* (Truong *et al.*, 2023), *xia2.multiplex* (Gildea *et al.*, 2022) and *DIALS* (Winter *et al.*, 2018, 2022), is available for processing 3D ED/microED data. Data processing can be performed via either command-line instructions or graphical user interfaces (GUIs) (Powell *et al.*, 2021; Brehm *et al.*, 2023), offering flexibility to users. However, format conversion, sorting, processing, analyzing and merging a large number of 3D ED/microED datasets can be time consuming and laborious. Automated batch processing and real-time processing capabilities are still underdeveloped for 3D ED/microED. In our previous work, *Edtools* (Wang *et al.*, 2019) was developed to process batch datasets in offline mode, which only works in a command line. A processing pipeline *Scipion-ED* (Bengtsson *et al.*, 2022) embedded in *Scipion* was also developed. However, the complex operation and specialized platform requirements make these programs difficult for beginners to use.

In this work, we utilized *XDS* as the data processing engine and developed a pipeline with a graphical user interface for automated real-time and offline batch 3D ED/microED data

processing (*AutoLEI*). Based on the fast data processing performance of *XDS*, *AutoLEI* enables both real-time and offline batch 3D ED/microED data processing. This Python-based pipeline is straightforward to install and provides users with the options and parameters that are essential for working with electron diffraction data. Meanwhile, data processing statistics and quality indicators are also optimized for 3D ED/microED. Here, we demonstrate how to use *AutoLEI* to process data collected from different types of crystals, including an organic molecule, a metal–organic framework and two proteins, on different microscopes and detectors. An example of real-time data processing using *AutoLEI* is also presented. Furthermore, a detailed tutorial with training datasets is provided.

2. *AutoLEI* implementation

2.1. *AutoLEI* installation and supported data formats

AutoLEI is an open-access, Python-based 3D ED/microED data processing pipeline, available through PyPI, GitLab and Zenodo. The software operates on Linux systems or within the Windows Subsystem for Linux (WSL) environment with Python version 3.8 or greater, and requires *XDS* to be pre-installed and configured globally. *AutoLEI* also provides a one-click installation command that automatically installs all required dependencies on GitLab. Detailed installation instructions are available on corresponding GitLab wiki (see *Data availability*). In version 1.0.0, the native image format is SMV for *XDS* data processing. *AutoLEI* can convert MRC images collected by *EPUD* (TFS), and TIFF images collected on CheeTah (Amsterdam Scientific Instruments, ASI), OneView (GATAN) and Timepix hybrid pixel detectors (ASI). Other formats need to be converted in advance to SMV format using other software, for example *REDp* (Wan *et al.*, 2013) and *RosettaSciIO* (Prestat *et al.*, 2025). Real-time data processing is designed for raw MRC image data collected using *EPUD* (TFS) as well as data collected using *Instamatic* (Cichocka *et al.*, 2018). Future updates are expected for improved compatibility and increased support, while users can

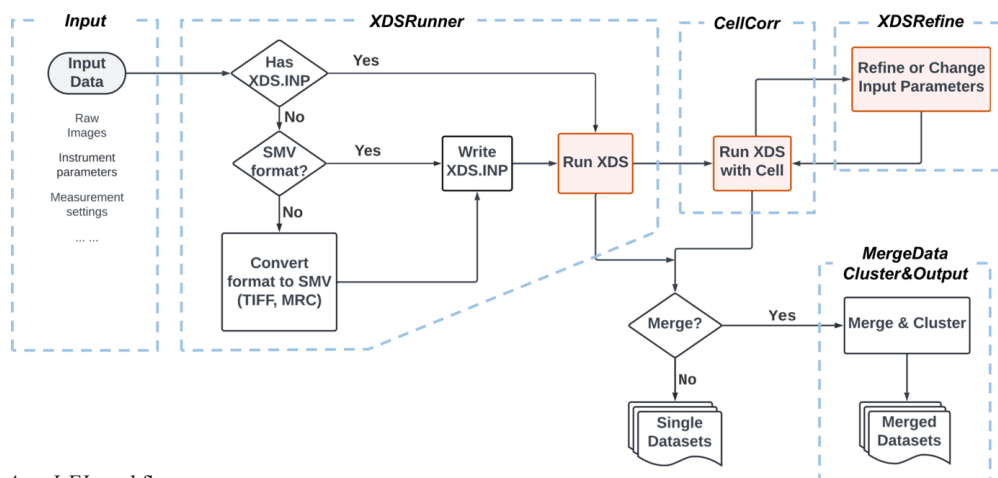


Figure 1
An overview of the *AutoLEI* workflow.

also add new functionality through the programming interface.

2.2. *AutoLEI* data processing workflow

The primary aim of *AutoLEI* is to process and analyze a batch of 3D ED/microED datasets using *XDS* offline or in real time. *AutoLEI* builds upon the experience from our earlier software, *Edtools* (Wang *et al.*, 2019). While *Edtools* relies on command-line input and requires an existing *XDS.INP* file generated by *Instamatic* (Cichocka *et al.*, 2018), *AutoLEI* introduces a graphical user interface (GUI) and supports image data collected using various acquisition software. Moreover, *AutoLEI*'s workflow has been specifically optimized for rotational electron diffraction data, allowing the processing to be significantly accelerated and enabling efficient real-time analysis. Fig. 1 shows the workflow of *AutoLEI*. Data processing goes through the steps of *XDSRunner*, *CellCorr*, *XDSRefine*, *MergeData* and *Cluster&Output*. These steps are implemented as tabs in the *AutoLEI* interface, as shown in Figures S1–S8 in the supporting information. By following this workflow, final HKL files are obtained for downstream structural determination. Videos 1–3 in the supporting information showcase the offline data processing of three examples: tyrosine, MOF SU-100 and the protein *H. sapiens* MutT homolog 1 (MTH1) while Video 4 shows the real-time data processing of triclinic lysozyme.

2.2.1. *Input*

Input allows users to set up the work directory, instrument parameters and measurement settings (Figure S1). The instrument parameters can be loaded from a predefined instrument model or existing input files. Alternatively, they can be given manually when datasets collected from different instruments and detectors are used. *Input* creates the *Input_parameters.txt* file (see an example in Section S3.2 of the supporting information), which will be used for generating input files *XDS.INP* for both real-time and offline 3D ED/microED data processing.

2.2.2. *XDSRunner*

XDSRunner generates the corresponding *XDS.INP* files for each dataset (Figure S2). It then launches *XDS* to process these datasets. No prior knowledge of the sample is required. *AutoLEI* processes all datasets sequentially under the work directory and parses statistics from *XDS* outputs. After data processing, *AutoLEI* will generate a summary table named *xdsrunner.xlsx*. The table contains key parameters and quality indicators such as space group (SG), unit-cell parameters (Unit cell), cell volume (Vol.), indexing rate (Index%), $I/\sigma(I)^{\text{asymptotic}}$ (ISa), redundancy-independent merging R factor (R_{meas}), correlation between reflection intensities of random half datasets ($CC_{1/2}$), completeness and estimated resolution (Reso.) of each dataset for inspection (Diederichs, 2010; Karplus & Diederichs, 2012). SG, unit cell and ISa are read from *XDS* outputs directly while others are calculated by *AutoLEI* based on *XDS* HKL files. The high-resolution cut-off

of each dataset is estimated by checking $CC_{1/2}$, R_{int} and the signal-to-noise ratio of each resolution shell.

XDSRunner also contains supporting functions as described below:

Format converter: Converts raw image files to SMV format and writes the metadata into SMV files.

Find beam center: Finds the beam center automatically with or without a beam stop.

Estimate symmetry & cell-cluster: Clusters data based on unit-cell parameters and estimates the Laue group. The Laue group is estimated based on R_{meas} and $CC_{1/2}$ using four indicators. The unit-cell clustering is performed based on a distance matrix. These ensure that data are processed in the correct Laue group. Details of these functions can be found in Section S1.1 of the supporting information.

2.2.3. *CellCorr*

CellCorr allows users to run *XDS* with given unit-cell parameters and space group (Figure S3). It allows re-indexing of the original HKL file using the space group and unit-cell parameters suggested by either *Estimate symmetry & cell-cluster* in *XDSRunner* or determined by other software packages such as *REDp* (Wan *et al.*, 2013) or *PETS2* (Palatinus *et al.*, 2019). This ensures that the datasets will be processed, scaled and merged in the correct Laue group and unit cell. *CellCorr* outputs a summary table named *xdsrunner2.xlsx* containing the same data quality indicators as *xdsrunner.xlsx*.

2.2.4. *XDSRefine*

XDSRefine provides the opportunity to update *XDS* keywords in the input files for single, selected or all datasets if users are not satisfied with the outputs from *CellCorr* (Figure S4). The step is useful for problematic data that cannot be processed using default settings. In *AutoLEI* version 1.0.0, *XDSRefine* also provides automatic refinement of the rotation axis, and excludes problematic frames defined by outliers; see Section S1.2 for more details. The reconstructed reciprocal lattice of each dataset can also be visually inspected. *XDSRefine* updates *xdsrunner2.xlsx* and the table will then be used for downstream data scaling and merging.

2.2.5. *MergeData*

MergeData analyzes statistics in *xdsrunner2.xlsx* and makes initial selections of datasets for merging (Figure S5). Users can then inspect key indicators listed in the table and apply a filter for scaling/merging. The default value will be provided as a reference. The information about the filtered data is written to the file *xdspicker.xlsx*. *MergeData* will then initiate *XSCALE* and output scaled data as an HKL file.

2.2.6. *Cluster&Output*

Cluster&Output performs cluster analysis based on pairwise correlation coefficients of reflection intensities in the file *XSCALE.Lp* generated from the last step (Figure S6). A dendrogram is produced for visual inspection for further

clustering (Wang *et al.*, 2019). *Cluster&Output* will then perform scaling/merging of each cluster, outputting corresponding HKL files and data processing statistics (Figure S7). Web-page-based reports will also be generated for inspection of the data reduction.

2.3. Real-time processing

Real-time 3D ED/microED data processing is a fully automated pipeline in *AutoLEI* (Figure S8). With default or customized settings, it provides crucial feedback to users for data quality analysis, evaluates data collection parameters and generates result HKL files in real time. The setup of the real-time processing function is found on the *RealTime* tab of *AutoLEI*, as shown in Fig. 2. Real-time processing reads *Input_parameters.txt* from *Input* and continuously monitors the input folder into which the data are being saved. Once a single dataset collection is completed, real-time processing will try to process the data in either screening mode or merging mode:

(a) *Screening mode*: When the unit-cell parameters and space group are not provided, the data are processed without performing data merging.

(b) *Merging mode*: When the unit-cell parameters and space group are provided, the data will be processed in the specified

cell and space group. If the data quality satisfies the filtering condition, it will be assigned as a valid dataset and merged with other valid datasets.

In the *RealTime* tab, estimated resolution, $CC_{1/2}$ and ISa for individual data sets or merged data are plotted and updated live, as shown in Figs. 2(a) and 2(b), respectively. Real-time processing currently supports the data collected using *EPUD* (TFS) and *Instamatic* (Cichocka *et al.*, 2018).

3. Examples

To demonstrate the functions and applications of *AutoLEI*, we chose four different samples (Fig. 3): an organic compound, tyrosine ($P2_12_12_1$, $a = 6.92 \text{ \AA}$, $b = 21.15 \text{ \AA}$, $c = 5.84 \text{ \AA}$) (Boggs & Donohue, 1971), a bismuth metal–organic framework, SU-100 ($I2/a$, $a = 17.85 \text{ \AA}$, $b = 9.61 \text{ \AA}$, $c = 21.05 \text{ \AA}$, $\beta = 96.77^\circ$) (Grape *et al.*, 2020), a protein, MutT homolog 1 (MTH1) ($P2_12_12_1$, $a = 59.5 \text{ \AA}$, $b = 67.1 \text{ \AA}$, $c = 80.1 \text{ \AA}$) (Svensson *et al.*, 2011), and triclinic lysozyme ($P1$, $a = 27.07 \text{ \AA}$, $b = 31.25 \text{ \AA}$, $c = 33.76 \text{ \AA}$, $\alpha = 88.0^\circ$, $\beta = 108.0^\circ$, $\gamma = 112.1^\circ$) (Wang *et al.*, 2007). 3D ED/microED data were collected on a variety of microscopes and detectors using different software and processed using *AutoLEI*, as summarized in Table 1. Among them, tyrosine (Video 1), SU-100 (Video 2) and MTH1 (Video 3) use the offline data processing protocol, while lysozyme (Video 4)



Figure 2 *RealTime* microED work page. (a) Screening mode. (b) Merging mode. ‘Iters’ in (b) denotes the number of merging iterations.

Table 1Summary of data used for demonstrating *AutoLEI* processing.

JEOL: Japan Electron Optics Laboratory Co., Ltd. ASI: Amsterdam Scientific Instruments. TFS: Thermo Fisher Scientific.

Sample	Microscope	Detector	Data collection software	Total datasets	Data collection time	Data processing time†
Tyrosine ($P2_12_12_1$)	JEM-2100 LaB ₆ (JEOL)	Timepix hybrid pixel detector (ASI)	<i>Instamatic</i>	12	~1 h 20 min	~5 min
SU-100 ($I2/a$)	JEM-2100 LaB ₆ (JEOL)	Timepix hybrid pixel detector (ASI)	<i>Instamatic</i>	16	~2 h 15 min	~7 min
MTH1 ($P2_12_12_1$)	Krios G3i (TFS)	Ceta-D (TFS)	<i>EPUD</i> (TFS)	38	~2 h 14 min	~12 min
Lysozyme ($P1$)	Krios G2 (TFS)	Ceta-D (TFS)	<i>EPUD</i> (TFS)	72	~6 h	real time (< 1 min)

† Timing is from the start of data processing to obtaining usable HKL files (Videos 1–3).

showcases the online data processing capability. The sample preparation details are given in Section S2.

3.1. Rapid data quality assessment and selection of a high-quality dataset for the structure determination of tyrosine

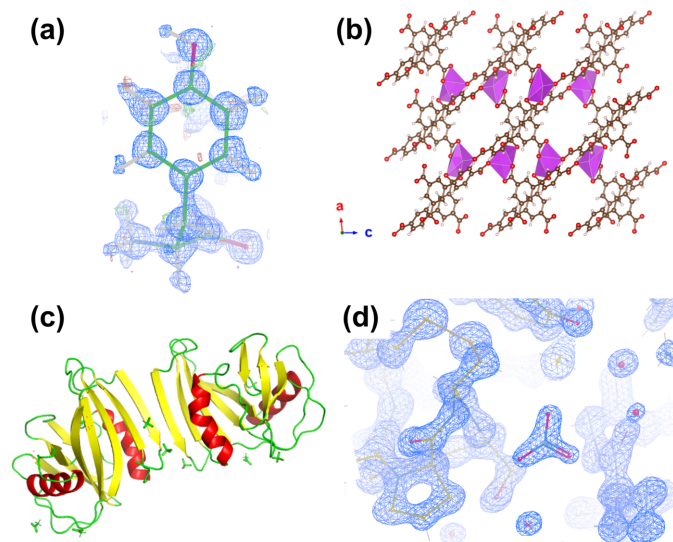
Because of the rapid 3D ED/microED data collection, many datasets can often be collected from different crystals. It is therefore important to make a rapid data quality assessment, in particular to identify datasets with the highest and lowest quality. Here we use tyrosine crystals as an example to demonstrate the capability of *AutoLEI*.

Tyrosine is a naturally occurring amino acid and an essential protein building block. Owing to its high orthorhombic symmetry and relatively high stability under the electron beam, it is possible to determine the structure of tyrosine crystals using a single 3D ED/microED dataset collected over a large tilt range. A total of 12 tyrosine datasets were collected by *Instamatic* on a JEOL JEM2100 microscope at 200 kV using a Timepix detector. The crystal tracking function was used to

ensure that most of the data were collected over a large tilt range ($\sim 130^\circ$) (Cichocka *et al.*, 2018).

In the *XDSRunner* tab, *XDS* first ran without a specified space group and automatically determined the Laue group and Bravais lattice. A default resolution range from 20 to 0.8 Å was set for the initial data processing. Unit-cell parameters, space group, ISa, $CC_{1/2}$, completeness and estimated resolution were extracted from *XDS* outputs, as shown in Fig. 4(a). Most of the datasets exhibit the estimated resolution of 0.8 Å, which equals the pre-defined high-resolution limit. This suggests that the resolution range should be extended. Notably, *XDS* assigned the space group $P222$ (No. 16) to 10 out of the 12 datasets while giving wrong space groups for two datasets (No. 7 and 12). By applying the built-in *Estimate symmetry & cell clustering* function – optimized for electron diffraction data – it was confirmed that these two datasets share the same Laue group and have similar unit-cell parameters to the other datasets. This allows new *XDS* runs to be performed in the *CellCorr* tab with the specified space group ($P222$) and unit cell ($a = 6.000$ Å, $b = 7.027$ Å, $c = 21.718$ Å, $\alpha = \beta = \gamma = 90^\circ$) [Fig. 4(a)]. Next, we applied *XDSrefine* to refine the rotation axis, divergence and mosaicity, remove scale outliers, and change the resolution range to 30–0.5 Å, as indicated in Fig. 4(b). After the refinement, statistic values such as $CC_{1/2}$ and estimated resolution are improved for most of the datasets. For example, the resolution of experiment 5 is improved from 1.0 to 0.74 Å, where the $CC_{1/2}$ increases from 91.78% to 96.11%.

The choice of optimal datasets can vary depending on which key indicators are prioritized. From our experience, we recommend datasets with ISa > 5, $CC_{1/2}$ > 95% and completeness > 80% as being promising for single-data-set-based structure determination. Here we selected experiment 10 because it has the highest completeness (87.61%) and also a high ISa (7.51), $CC_{1/2}$ (98.18%) and resolution (0.68 Å). *AutoLEI* also enables the systematic absences in reciprocal space to be visually inspected. By rapidly checking the reciprocal lattices of all datasets using *AutoLEI*, the presence of systematic absences could be identified on $h00$, $00l$ in experiment 10 and $0k0$ in experiment 12 (Figure S9). The space group was deduced as $P2_12_12_1$. The time used for data processing and analysis was approximately 5 minutes in total. The integrated data were then used for structure solution and kinematic refinement. The refinement converged to a final R_1 of 11.90% for 1253 reflections with $I > 2\sigma(I)$ and 162 parameters (Table S1). All hydrogen atoms, including the proton

**Figure 3**

Structures used for the demonstration of *AutoLEI*. (a) L-tyrosine obtained using a selected high-quality dataset. The electrostatic potential map (in blue, contoured at 1 RMSD) and difference map [positive in green and negative in red, contoured at 3 RMSD drawn in *COOT* (Emsley *et al.*, 2010)] are superimposed. (b) The framework of the bismuth-containing MOF SU-100 viewed along the b axis (BiO₇ polyhedra shown in purple). (c) Apo-MTH1. (d) Lysozyme model near a nitrate anion with no-filled $2F_o - F_c$ electrostatic potential map (blue, contoured at 1.5 RMSD) superimposed.

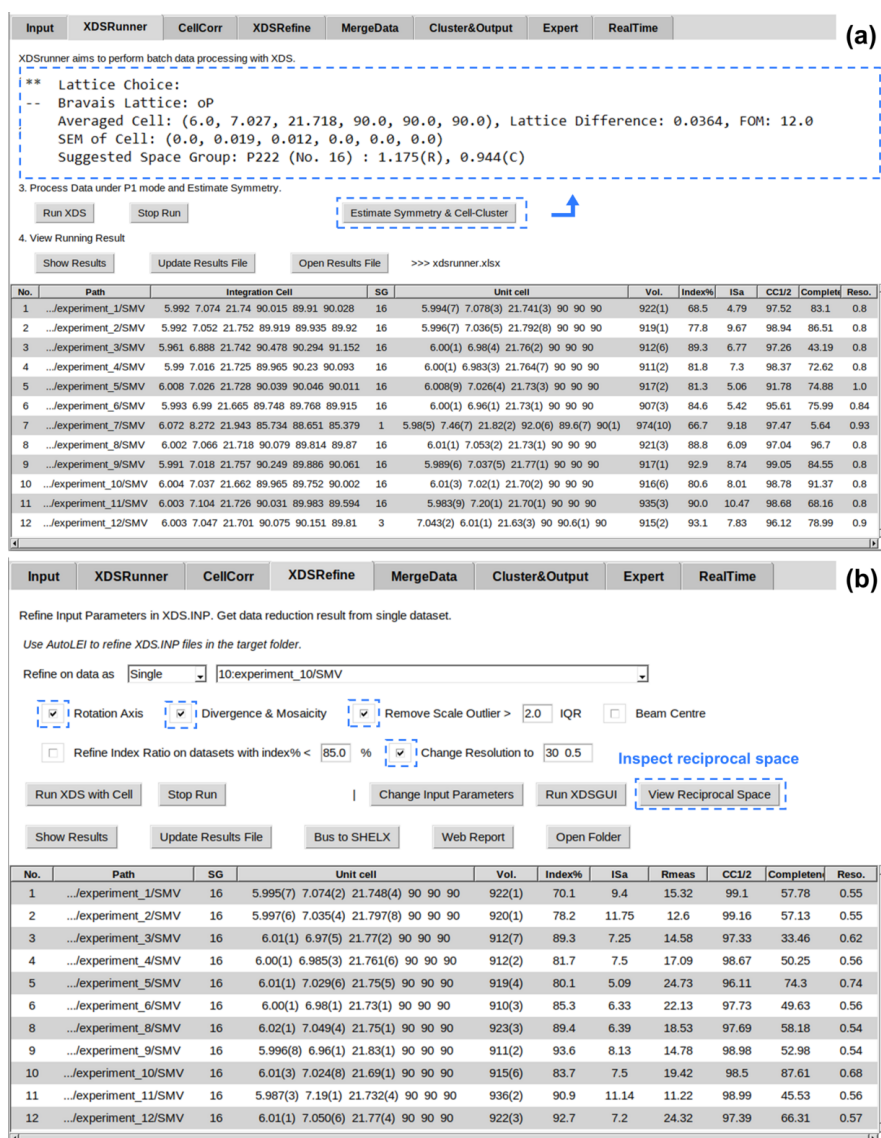


Figure 4

Batch data processing of 12 L-tyrosine datasets using *AutoLEI*. (a) Result table after running *XDSRunner*. The suggested unit-cell parameters with specific Laue group and lattice type setting were extracted from the function *Estimate Symmetry & Cell Cluster*. (b) Result table after running *XDSRefine*. *View Reciprocal Space* was designed to inspect the reciprocal space of a single dataset.

near the N atom, could be found as shown in Fig. 3(a). The structure could also be solved and refined using other datasets (Table S1) except for two datasets (experiment 3 and experiment 7) that had too low completeness (43.19% and 5.64%, respectively) [Fig. 4(a)] due to beam blockage by the grid.

In conclusion, *AutoLEI* can directly generate data information tables so that users can make their own assessment of the data quality. This can help users to identify the best datasets for structure determination or to decide whether data merging is needed in cases where the completeness is low. It can also help to eliminate low-quality datasets.

3.2. Interactive clustering and merging of multiple datasets from a monoclinic bismuth metal–organic framework, SU-100

For crystals with low symmetry, merging datasets collected from crystals from different orientations is often needed to

achieve high completeness. In addition, data merging can minimize the effects of dynamical scattering and improve the data quality (Xu *et al.*, 2018). Here we chose a monoclinic bismuth metal–organic framework (*I2/a*, $a = 17.85 \text{ \AA}$, $b = 9.61 \text{ \AA}$, $c = 21.05 \text{ \AA}$, $\beta = 96.77^\circ$) (Grape *et al.*, 2020). In this study, a total of 16 datasets were collected using *Instamatic*, and all data were processable after *XDSRefine*, as shown in Fig. 5(a). The data resolution ranges from 0.54 (experiment 8) to 0.78 Å (experiment 11), and the completeness ranges from 20.82% (experiment 7) to 82.61% (experiment 2). Fourteen out of the 16 datasets have an ISa higher than 5 and were then merged during the *XDSRefine* step after specifying the space group and unit-cell parameters [Fig. 5(b)]. We further conducted cluster analysis based on the pairwise correlation coefficients of reflection intensities (Wang *et al.*, 2019). As shown in Fig. 5(c), a threshold of 0.5 was selected that gives a significant distance between the four clusters. The final

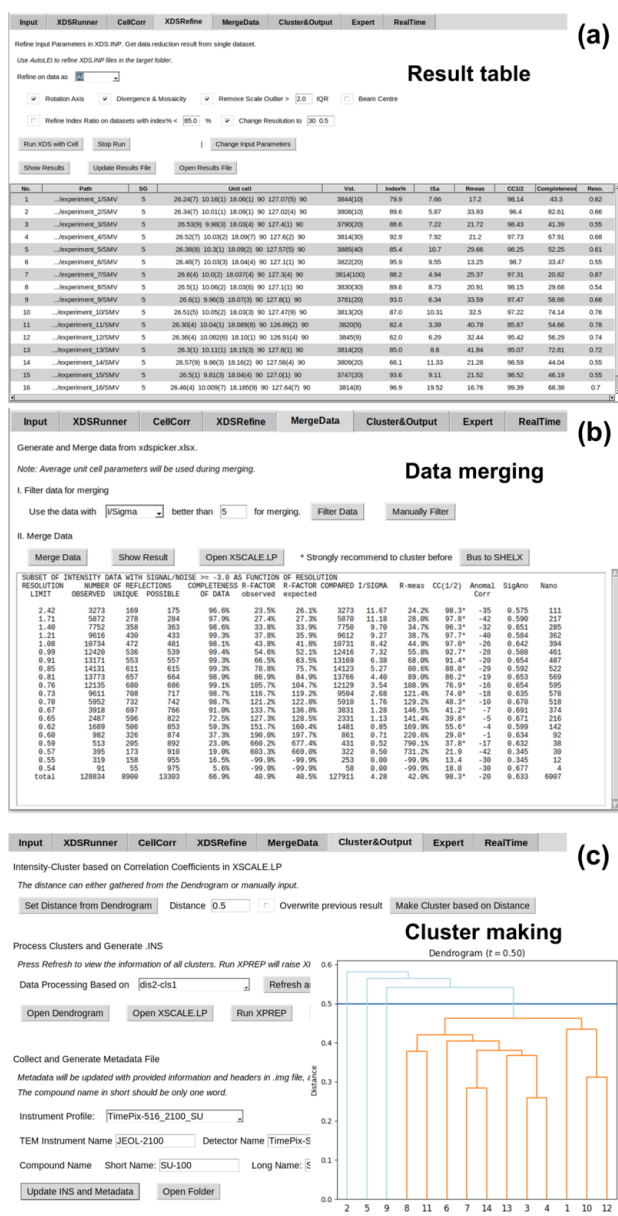


Figure 5

Data processing, analysis and merging of datasets of SU-100 by *AutoLEI*. (a) *XDSRefine* work page showing the result table of 16 datasets. (b) *MergeData* work page. Fourteen out of 16 datasets with $I/\sigma > 5$ were used for initial merging after several clicks. (c) *Cluster&Output* work page. Clusters were created based on pairwise reflection intensity correlations and a final HKL file for structure determination was prepared that contained merged data from the largest cluster of 11 datasets (Wang *et al.*, 2019).

merged HKL file from 11 datasets was obtained in 7 minutes. The resolution cut-off of the merged data estimated by *AutoLEI* was 0.62 Å. All non-H atom positions could be directly located and refined anisotropically using the merged data. The refinement converged to an R_1 of 19.46% for 4231 unique reflections with $I > 2\sigma(I)$ and 200 parameters, as shown in Table S2.

It is important to mention that data merging can significantly improve the data quality (Xu *et al.*, 2018; Samperisi *et al.*, 2021). We compared the structure refinement using the merged dataset and a single dataset from experiment 2, which has the highest completeness of all single datasets, as shown in Figure S10. While the structure could be refined anisotropically using the merged data, anisotropic refinement using

a single dataset of experiment 2 gave non-positive-definite atomic displacement parameters for 19 atoms.

3.3. Processing small-wedge 3D ED/microED data from the protein MutT homolog 1 (MTH1)

For structure determination of very beam sensitive crystals, such as protein crystals (Shi *et al.*, 2013), a small-wedge data collection strategy (Clabbers *et al.*, 2021) can be used, where a large number of 3D ED/microED datasets are collected from different crystals, each with a small tilt range (10–20°) and a low completeness. *AutoLEI* is ideal for conducting repetitive data processing, data scaling and merging of such batch data.

MTH1 plays an important role in sanitizing oxidized dNTPs from the free nucleotide pool, preventing their incorporation into DNA, which reduces genotoxicity (Yoshimura *et al.*, 2003; Gad *et al.*, 2014; Jemth *et al.*, 2018). Here, we used apo-MTH1 as an example to demonstrate how small-wedge 3D ED/microED data can be effortlessly processed and merged with *AutoLEI*. A total of 38 datasets were collected using *EPUD* (TFS) and the tilt range was limited to 15 or 20° in order to collect data with a high signal-to-noise ratio before the crystals were damaged by the electron beam. Only 28 out of 38 datasets were processable after *XDSRefine*. Fig. 6(a) shows the dendrogram of the clustering analysis after the *MergeData* step in *AutoLEI*. A stepwise clustering analysis using different distance thresholds can be performed in *AutoLEI* to find the optimal one. As shown in Fig. 6(a), five thresholds were tested, resulting in five groups containing 27, 26, 23, 18 and 13 datasets, respectively. Each group is called dis X -cls Y , where X is the threshold and Y is the main cluster under the threshold. The first four groups showed comparable completeness (74.93–75.66%), while the last group had a slightly lower completeness (70.23%). We chose dis4-cls1 for data merging because it has the highest $CC_{1/2}$ and lowest R_{meas} among the first four groups. A web-page-based report was then generated and used for data reduction of dis4-cls1. Fig. 6(b) indicates the relationships between $CC_{1/2}$, R_{meas} and completeness against resolution drawn from the report of dis4-cls1. The estimated resolution was 2.76 Å and it took less than 12 minutes to

obtain the final HKL file using *AutoLEI*. The structure of MTH1 could be solved by *Phaser* (McCoy *et al.*, 2007) and refined using *Phenix.refine* (Afonine *et al.*, 2012; Liebschner *et al.*, 2019). The final R_{work} and R_{free} converged to 21.07% and 26.82%, respectively, as shown in Table S3.

3.4. Real-time batch data processing for triclinic lysozyme

Real-time data processing for single-crystal X-ray diffraction is implemented at most synchrotron X-ray facilities. This capability is also desirable for 3D ED/microED, as it provides live processing statistics on the current batch data collection. This allows users to inspect the data quality in order to decide how to optimize the data collection strategy and to stop data collection when enough data are collected. By combining *AutoLEI* with automated data collection protocols, rapid structure determination and phase analysis become possible. Here, we used triclinic lysozyme crystals to demonstrate the real-time 3D ED/microED data processing capability of *AutoLEI*.

Triclinic lysozyme crystals diffract to a very high resolution (Wang *et al.*, 2007). However, triclinic lysozyme is challenging for 3D ED/microED because of its low symmetry and preferred crystal orientation. Here, a small-wedge data collection strategy was employed to increase the data resolution. The tilt range of individual crystals was restricted to 15° to keep beam damage minimal. To achieve real-time data

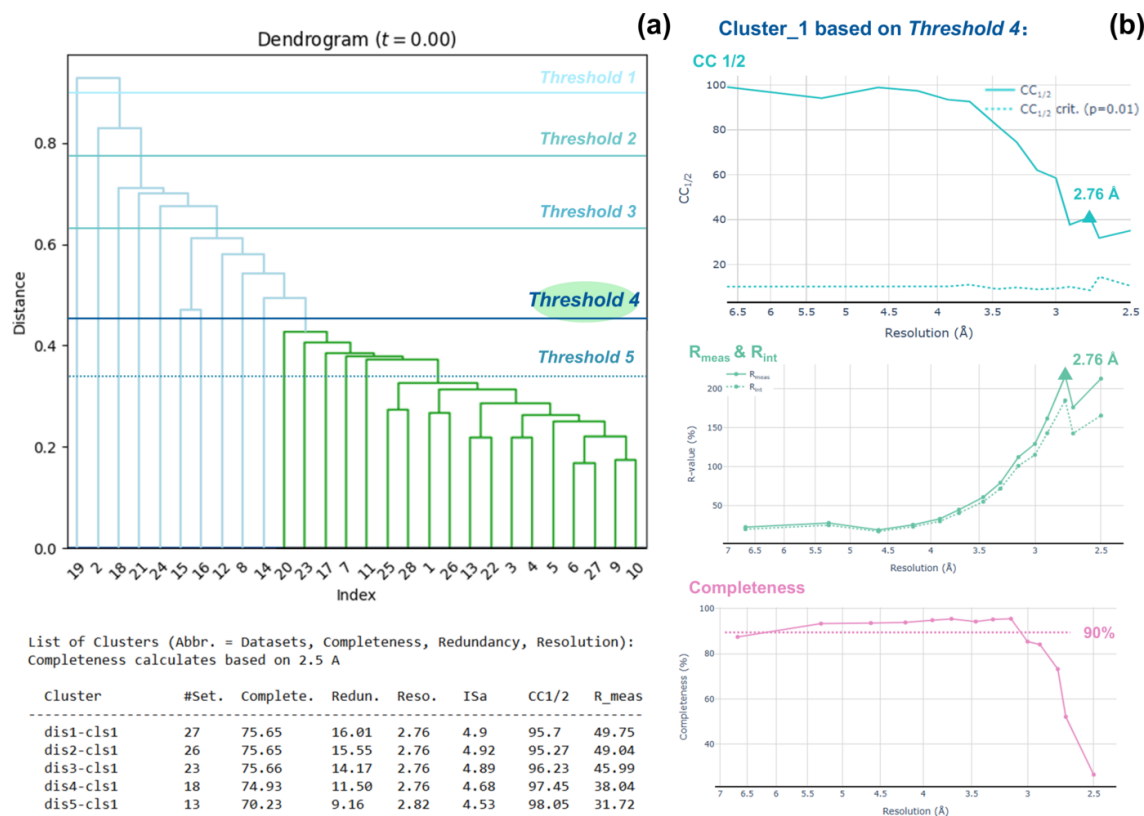


Figure 6 Data processing, analysis and merging of 38 MTH1 datasets by *AutoLEI*. (a) Clustering analysis after data merging in the *AutoLEI* interface. A total of 5 cluster selections were made based on different distances. A summary of the key statistics of these 5 clusters was tabulated in the command window. (b) Web-page-based report plots of Cluster_1 based on threshold 4 (18 datasets in total). Plots were captured from the report. From top to bottom: $CC_{1/2}$ versus resolution, R_{meas} and R_{int} versus resolution, and completeness versus resolution.

merging (merge mode), the unit-cell parameters and space group need to be given as prior knowledge. These two input parameters can easily be obtained by first collecting a high-tilt 3D ED/microED dataset. Data with both $CC_{1/2}$ higher than 90% and ISa higher than 5 were considered to be good and subsequently used for real-time merging iterations. These parameters could also be customized by editing the strategy file settings. Furthermore, the resolution of each good dataset as well as the statistics of merged data completeness and $CC_{1/2}$ were updated live. Users can then use this information to check whether the current data collection strategy or data filtering strategy should be modified.

A total of 71 datasets were collected on lysozyme crystals in 6 h. The completeness of the data reached 95.8%, while the $CC_{1/2}$ was 98.68% at the resolution cut-off of 1.3 Å, as shown in Fig. 7(a). Owing to preferred orientation, the completeness

only reached 50% and 78% when merging the first 10 and 21 datasets, respectively. In order to reach a higher completeness, we had to change the data collection strategy and look for crystals with different morphologies on the TEM grids. The data collection was terminated after the merged data reached a completeness higher than 95% at the pre-set resolution cut-off of 1.3 Å and a $CC_{1/2}$ above 98%. The web-page-based report was then generated to estimate the real high-resolution cut-off, which was set at 1.1 Å at the $CC_{1/2}$ of 30%, as shown in Figs. 7(b) and 7(c). The final HKL file merged from 56 datasets was used immediately for structure solution using *Phaser* (McCoy *et al.*, 2007) and refinement using *Phenix* (Afonine *et al.*, 2012; Liebschner *et al.*, 2019). The final refinement converged with R_{work} of 20.42% and R_{free} of 23.45%. High R values are common for electron diffraction data due to dynamical effects. The no-filled $2F_o - F_c$ elec-

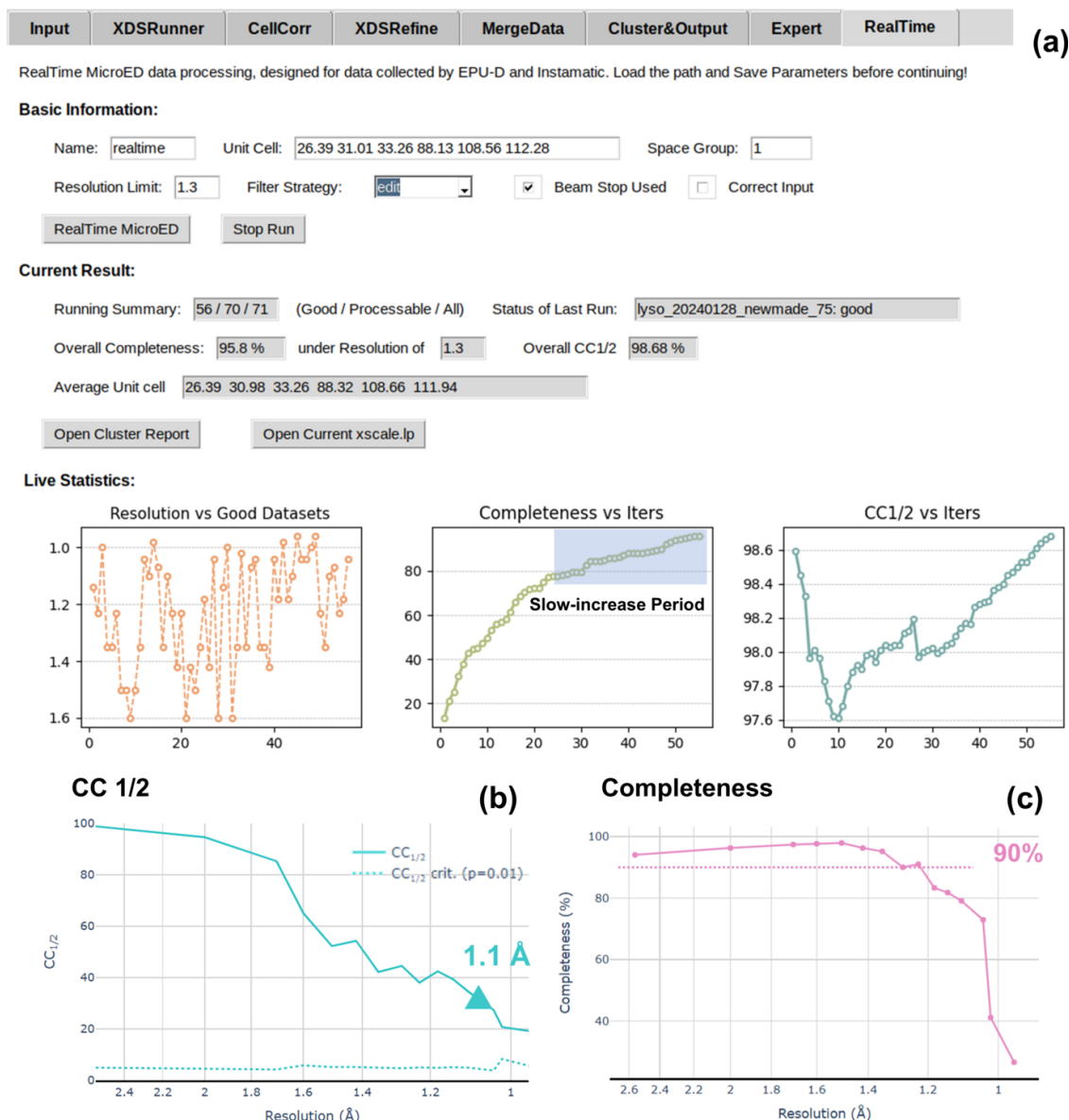


Figure 7

Real-time 3D ED/microED data processing of triclinic lysozyme datasets using *AutoLEI*. (a) The live processing results are displayed under *Live Statistics*. (b) $CC_{1/2}$ and (c) completeness against resolution plots captured from the web-page-based report.

trostatic potential map for the lysozyme structure is presented in Fig. 3(d).

4. Conclusion and outlook

We developed an *XDS*-based pipeline with a GUI, *AutoLEI*, for offline and real-time batch 3D ED/microED data processing. The visualization interface enables data import, processing, selection and merging with only a few clicks. We demonstrated the applications of *AutoLEI* on four different samples with various processing strategies to show its simplicity and ease of use. For stable or high-symmetry crystals, *AutoLEI* enables rapid identification of promising single datasets for structure determination. For low symmetry or electron-sensitive materials, *AutoLEI* can assist in selecting and merging datasets, thereby generating a final HKL file with high completeness. In addition, real-time data processing provides direct feedback, allowing users to adjust data collection strategies ‘on the fly’.

Further developments are planned for *AutoLEI*, including incorporation of additional data processing engines, enabling comparative analyses and data integration. These developments will further simplify 3D ED/microED data processing and lower the entry barriers for researchers new to the field of 3D ED/microED.

5. Related literature

The following references are cited in the supporting information: Burnley *et al.* (2017), Dolomanov *et al.* (2009), Giordano *et al.* (2012), Grant & Pickup (1995), Grosse-Kunstleve (1999), Grosse-Kunstleve *et al.* (2004), Jain & Dubes (1988), Knudsen *et al.* (2013), Kolb *et al.* (2012), Niggli (1928), Palatinus (2013), Schönemann (1966), Sheldrick (2008), Storn & Price (1997), Van Der Walt *et al.* (2014).

Acknowledgements

We thank Dr Stef Smeets and Dr Taimin Yang for sharing the code of *Edtools*. The 3D ED/microED data for the protein samples were collected at the Cryo-EM Swedish National Facility funded by KAW, Family Erling Persson and Kempe Foundations, SciLifeLab, Stockholm University and Umeå University. We also thank Dr Mathieu Coinçon, Mr Dustin Morado and Dr Marta Carroni for TEM support at SciLifeLab.

Data availability

The source code is open-source and available on GitLab (<https://gitlab.com/tristonewang/autolei>), Zenodo (10.5281/zenodo.15155987) and PyPI (<https://pypi.org/project/autolei/>). The data used as examples can be downloaded from Zenodo (<https://zenodo.org/records/14536385>).

Funding information

This work was supported by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956099 (NanED – Electron Nanocrystallography-H2020-MSCAITN). We also acknowledge financial support from the Swedish Research Council [VR, 2019-00815 (XZ), 2022-03681 (PS) and 2022-03596 (HX)], the Swedish Cancer Society 24-3848-Pj to PS and the Knut and Alice Wallenberg Foundation (KAW) 2019.0124 to XZ. XZ acknowledges support from the Wallenberg Initiative Materials Science for Sustainability (WISE) funded by the Knut and Alice Wallenberg Foundation (KAW).

References

- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* **D68**, 352–367.
- Bengtsson, V. E. G., Pacoste, L., de la Rosa-Trevin, J. M., Hofer, G., Zou, X. & Xu, H. (2022). *J. Appl. Cryst.* **55**, 638–646.
- Boggs, R. & Donohue, J. (1971). *Acta Cryst.* **B27**, 247–247.
- Brehm, W., Triviño, J., Krahn, J. M., Usón, I. & Diederichs, K. (2023). *J. Appl. Cryst.* **56**, 1585–1594.
- Burnley, T., Palmer, C. M. & Winn, M. (2017). *Acta Cryst.* **D73**, 469–477.
- Cheng, A., Negro, C., Bruhn, J. F., Rice, W. J., Dallakyan, S., Eng, E. T., Waterman, D. G., Potter, C. S. & Carragher, B. (2021). *Protein Sci.* **30**, 136–150.
- Cho, J., Willhammar, T. & Zou, X. (2023). *Microporous Mesoporous Mater.* **358**, 112400.
- Cichocka, M. O., Ångström, J., Wang, B., Zou, X. & Smeets, S. (2018). *J. Appl. Cryst.* **51**, 1652–1661.
- Clabbers, M. T. B., Holmes, S., Muusse, T. W., Vajjhala, P. R., Thygesen, S. J., Malde, A. K., Hunter, D. J. B., Croll, T. I., Flueckiger, L., Nanson, J. D., Rahaman, Md. H., Aquila, A., Hunter, M. S., Liang, M., Yoon, C. H., Zhao, J., Zatsepin, N. A., Abbey, B., Sieracki, E., Gambin, Y., Stacey, K. J., Darmanin, C., Kobe, B., Xu, H. & Ve, T. (2021). *Nat. Commun.* **12**, 2578.
- de la Cruz, M. J., Martynowycz, M. W., Hattne, J. & Gonen, T. (2019). *Ultramicroscopy* **201**, 77–80.
- Diederichs, K. (2010). *Acta Cryst.* **D66**, 733–740.
- Dolomanov, O. V., Bourhis, L. J., Gildea, R. J., Howard, J. A. K. & Puschmann, H. (2009). *J. Appl. Cryst.* **42**, 339–341.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Gad, H., Koolmeister, T., Jemth, A.-S., Eshtad, S., Jacques, S. A., Ström, C. E., Svensson, L. M., Schultz, N., Lundbäck, T., Einarsdottir, B. O., Saleh, A., Göktürk, C., Baranczewski, P., Svensson, R., Berntsson, R. P.-A., Gustafsson, R., Strömberg, K., Sanjiv, K., Jacques-Cordonnier, M.-C., Desroses, M., Gustavsson, A.-L., Olofsson, R., Johansson, F., Homan, E. J., Loseva, O., Bräutigam, L., Johansson, L., Höglund, A., Hagenkört, A., Pham, T., Altun, M., Gaugaz, F. Z., Vikingsson, S., Evers, B., Henriksson, M., Vallin, K. S. A., Wallner, O. A., Hammarström, L. G. J., Wiita, E., Almlöf, I., Kalderén, C., Axelsson, H., Djureinovic, T., Puigvert, J. C., Häggblad, M., Jeppsson, F., Martens, U., Lundin, C., Lundgren, B., Granelli, I., Jensen, A. J., Artursson, P., Nilsson, J. A., Stenmark, P., Scobie, M., Berglund, U. W. & Helleday, T. (2014). *Nature* **508**, 215–221.
- Gemmi, M., Mugnaioli, E., Gorelik, T. E., Kolb, U., Palatinus, L., Boullay, P., Hovmöller, S. & Abrahams, J. P. (2019). *ACS Cent. Sci.* **5**, 1315–1329.
- Gildea, R. J., Beilsten-Edmands, J., Axford, D., Horrell, S., Aller, P., Sandy, J., Sanchez-Weatherby, J., Owen, C. D., Lukacik, P., Strain-

- Damerell, C., Owen, R. L., Walsh, M. A. & Winter, G. (2022). *Acta Cryst.* **D78**, 752–769.
- Giordano, R., Leal, R. M. F., Bourenkov, G. P., McSweeney, S. & Popov, A. N. (2012). *Acta Cryst.* **D68**, 649–658.
- Grant, J. A. & Pickup, B. T. (1995). *J. Phys. Chem.* **99**, 3503–3510.
- Grape, E. S., Xu, H., Cheung, O., Calmels, M., Zhao, J., Dejoie, C., Proserpio, D. M., Zou, X. & Inge, A. K. (2020). *Cryst. Growth Des.* **20**, 320–329.
- Guo, P., Shin, J., Greenaway, A. G., Min, J. G., Su, J., Choi, H. J., Liu, L., Cox, P. A., Hong, S. B., Wright, P. A. & Zou, X. (2015). *Nature* **524**, 74–78.
- Grosse-Kunstleve, R. W. (1999). *Acta Cryst.* **A55**, 383–395.
- Grosse-Kunstleve, R. W., Sauter, N. K. & Adams, P. D. (2004). *Acta Cryst.* **A60**, 1–6.
- Huang, Z., Grape, E. S., Li, J., Inge, A. K. & Zou, X. (2021). *Coord. Chem. Rev.* **427**, 213583.
- Huang, Z., Willhammar, T. & Zou, X. (2021). *Chem. Sci.* **12**, 1206–1219.
- Ito, S., White, F. J., Okunishi, E., Aoyama, Y., Yamano, A., Sato, H., Ferrara, J. D., Jasnowski, M. & Meyer, M. (2021). *CrystEngComm* **23**, 8622–8630.
- Jain, A. K. & Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice Hall.
- Jemth, A.-S., Gustafsson, R., Bräutigam, L., Henriksson, L., Vallin, K. S. A., Sarno, A., Almlöf, I., Homan, E., Rasti, A., Warpman Berglund, U., Stenmark, P. & Helleday, T. (2018). *Nucleic Acids Res.* **46**, 10888–10904.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Karplus, P. A. & Diederichs, K. (2012). *Science* **336**, 1030–1033.
- Knudsen, E. B., Sørensen, H. O., Wright, J. P., Goret, G. & Kieffer, J. (2013). *J. Appl. Cryst.* **46**, 537–539.
- Kolb, U., Gorelik, T., Kübel, C., Otten, M. T. & Hubert, D. (2007). *Ultramicroscopy* **107**, 507–513.
- Kolb, U., Shankland, K., Meshi, L., Avilov, A. & David, W. I. F. (2012). *Uniting electron crystallography and powder diffraction*. Dordrecht: Springer Netherlands.
- Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J. & Adams, P. D. (2019). *Acta Cryst.* **D75**, 861–877.
- Lightowler, M., Li, S., Ou, X., Cho, J., Liu, B., Li, A., Hofer, G., Xu, J., Yang, T., Zou, X., Lu, M. & Xu, H. (2024). *Angew. Chem. Int. Ed.* **63**, e202317695.
- Lightowler, M., Li, S., Ou, X., Zou, X., Lu, M. & Xu, H. (2022). *Angew. Chem. Int. Ed.* **61**, e202114985.
- Lu, P., Xu, J., Sun, Y., Guillet-Nicolas, R., Willhammar, T., Fahda, M., Dib, E., Wang, B., Qin, Z., Xu, H., Cho, J., Liu, Z., Yu, H., Yang, X., Lang, Q., Mintova, S., Zou, X. & Valtchev, V. (2024). *Nature* **636**, 368–373.
- Luo, Y., Wang, B., Smeets, S., Sun, J., Yang, W. & Zou, X. (2023). *Nat. Chem.* **15**, 483–490.
- Martynowycz, M. W., Clabbers, M. T. B., Hattne, J. & Gonen, T. (2022). *Nat. Methods* **19**, 724–729.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Niggli, P. (1928). *Handbuch der Experimentalphysik*, Vol. 7, pp. 108–176. Leipzig: Akademische Verlagsgesellschaft.
- Palatinus, L. (2013). *Acta Cryst.* **B69**, 1–16.
- Palatinus, L., Brázda, P., Jelínek, M., Hrdá, J., Steciuk, G. & Klementová, M. (2019). *Acta Cryst.* **B75**, 512–522.
- Plana-Ruiz, S., Krysiak, Y., Portillo, J., Alig, E., Estradé, S., Peiró, F. & Kolb, U. (2020). *Ultramicroscopy* **211**, 112951.
- Powell, S. M., Novikova, I. V., Kim, D. N. & Evans, J. E. (2021). *bioRxiv*, <https://doi.org/10.1101/2021.12.13.472146>.
- Prestat, E., de la Peña, F., Lähnemann, J., Jokubauskas, P., Francis, C., Fauske, V. T., pietsjoh, Ostasevicius, T., Nemoto, T., Furnival, T., Johnstone, D. N., Cautaeys, N., Somnath, S., pquinn-dls, Caron, J., MacArthur, K. E., Nord, M., Burdet, P., Tappy, N., Aarholt, T., Poon, T., actions-user, Ribet, S., Attolight-NTappy, Slater, T., Taillon, J., Migunov, V. & noemiebonnet (2025). hyperspy/roset-tasciio: v0.10.0. <https://zenodo.org/records/16503042>.
- Samperisi, L., Jaworski, A., Kaur, G., Lillerud, K. P., Zou, X. & Huang, Z. (2021). *J. Am. Chem. Soc.* **143**, 17947–17952.
- Schönemann, P. H. (1966). *Psychometrika* **31**, 1–10.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. (2013). *eLife* **2**, e01345.
- Storn, R. & Price, K. (1997). *J. Glob. Optim.* **11**, 341–359.
- Svensson, L. M., Jemth, A.-S., Desroses, M., Loseva, O., Helleday, T., Högbom, M. & Stenmark, P. (2011). *FEBS Lett.* **585**, 2617–2621.
- Tremlett, C. J., Stubbs, J., Stuart, W. S., Shaw Stewart, P. D., West, J., Orville, A. M., Tews, I. & Harmer, N. J. (2025). *IUCrJ* **12**, 262–279.
- Truong, K.-N., Ito, S., Wojciechowski, J. M., Göb, C. R., Schürmann, C. J., Yamano, A., Del Campo, M., Okunishi, E., Aoyama, Y., Mihira, T., Hosogi, N., Benet-Buchholz, J., Escudero-Adán, E. C., White, F. J., Ferrara, J. D. & Bücker, R. (2023). *Symmetry* **15**, 1555.
- Unge, J., Lin, J., Weaver, S. J., Sae Her, A. & Gonen, T. (2024). *Adv. Sci.* **11**, 2400081.
- Van Der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E. & Yu, T. (2014). *PeerJ* **2**, e453.
- van Genderen, E., Clabbers, M. T. B., Das, P. P., Stewart, A., Nederlof, I., Barentsen, K. C., Portillo, Q., Pannu, N. S., Nicolopoulos, S., Gruene, T. & Abrahams, J. P. (2016). *Acta Cryst.* **A72**, 236–242.
- Wan, W., Sun, J., Su, J., Hovmöller, S. & Zou, X. (2013). *J. Appl. Cryst.* **46**, 1863–1873.
- Wang, B., Zou, X. & Smeets, S. (2019). *IUCrJ* **6**, 854–867.
- Wang, J., Dauter, M., Alkire, R., Joachimiak, A. & Dauter, Z. (2007). *Acta Cryst.* **D63**, 1254–1268.
- Winter, G., Beilsten-Edmands, J., Devenish, N., Gerstel, M., Gildea, R. J., McDonagh, D., Pascal, E., Waterman, D. G., Williams, B. H. & Evans, G. (2022). *Protein Sci.* **31**, 232–250.
- Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K. & Evans, G. (2018). *Acta Cryst.* **D74**, 85–97.
- Xu, H., Lebrette, H., Clabbers, M. T. B., Zhao, J., Griese, J. J., Zou, X. & Högbom, M. (2019). *Sci. Adv.* **5**, eaax4621.
- Xu, H., Lebrette, H., Yang, T., Srinivas, V., Hovmöller, S., Högbom, M. & Zou, X. (2018). *Structure* **26**, 667–675.e3.
- Yoshimura, D., Sakumi, K., Ohno, M., Sakai, Y., Furuichi, M., Iwai, S. & Nakabeppu, Y. (2003). *J. Biol. Chem.* **278**, 37965–37973.
- Zhang, D., Oleynikov, P., Hovmöller, S. & Zou, X. (2010). *Z. Kristallogr.* **225**, 94–102.