Autonomous MicroED data collection enables compositional analysis

Johan Unge[1#], Jieye Lin[1#], Sara J Weaver[1], Ampon Sae Her[1] and Tamir Gonen[1,2,3*]

[1]Department of Biological Chemistry, University of California, Los Angeles, 615 Charles E. Young Drive South, Los Angeles, California 90095, United States
[2] Department of Physiology, University of California, Los Angeles, 615 Charles E. Young Drive South, Los Angeles, California 90095, United States
[3] Howard Hughes Medical Institute, University of California, Los Angeles, Los Angeles, California 90095, United States

Abstract

MicroED is an effective method for analyzing the structural properties of sub-micron crystals, which are frequently found in small-molecule powders. By developing and using an autonomous and high throughput approach to MicroED, we demonstrate the expansion of capabilities and the possibility of performing complete compositional analysis of complex samples. With the use of SerialEM for data collection of thousands of datasets from thousands of crystals and an automated processing pipeline, compositional analysis of complex mixtures of organic and inorganic compounds can be accurately executed. Quantitative analysis suitable for compounds having similar chemical properties can be made on the fly. These compounds can be distinguished by their crystal structure properties prior to structure solution. Additionally, with sufficient statistics from the autonomous approach, even small amounts of compounds in mixtures can be reliably detected. Finally, atomic structures can be determined from the thousands of data sets.

Introduction

Microcrystal electron diffraction (MicroED) is a cryoEM method for determining the 3D structure of inorganic, organic, or biological macromolecules.[1,2] Compared to X-rays or neutrons, electrons exhibit a stronger interaction with the sample and cause considerably less damage per useful elastic scattering event. Thus, the optimal crystal size for MicroED is well below 1 μm^3, and even crystals consisting of only a few layers can be used for structure determination.[3] In fact, electron diffraction is currently the only method that can routinely produce a complete diffraction dataset for samples of this size and can be acquired with only picograms of a sample. This contrasts with X-ray free-electron laser (XFEL) serial crystallography and serial synchrotron crystallography (SSX), which require hundreds of thousands of crystals, with the crystals needing to be larger than 1 µm and 5 µm, respectively.[4,5] Ultimately, crystallographic techniques that use electrons, X-rays, or neutrons are complementary because the beams interacting with the sample have different sensitivities for various elements. Moreover, MicroED and neutron crystallography can visualize hydrogen atoms, which are often not possible with X-ray crystallography due to

their relatively small cross-section with X-rays.[6] Finally, electron diffraction can provide information on charge.

Determining the composition of mixtures of compounds is a common and essential task that can be challenging as the components may have similar physical and chemical properties. There are several methods for determining the composition of mixtures, including chromatography, spectroscopy, and mass spectrometry[7-9]. In recent years, MicroED has emerged as a promising technique for identifying and quantifying the components of complex mixtures, offering high resolution and sensitivity. Crystallization naturally distinguishes between different isomers, including constitutional, conformational, geometric, diastereomer, and enantiomer isomers, based on their crystal structure properties[10]. Even without solving the full crystal structure, the unit cell parameters and space group can describe the crystal structure and its content. This means that most compounds can be distinguished by their unit cells, making compositional analysis of crystalline samples possible without collecting a full set of crystallographic data. Since determining the unit cell parameters is one of the first steps in structure determination and does not depend on solving the phase problem or collecting high-resolution data, this approach can be used even when the diffraction qualities of the compounds are limited.

High throughput automation is increasingly essential in structural biology methods, particularly in MicroED, where multiple data sets are often required. Autonomous data collection reduces manual labor, increases instrument usage, and allows larger data sets to be collected in the same timeframe. This is especially important when many data sets must be merged to achieve higher completeness or when crystals are oriented preferentially on the grid. Additionally, studies have shown that automated approaches are effective in analyzing multiple phase systems and distinguishing different crystal forms. For instance, Wang *et al*. demonstrated the automated analysis of two zeolites using SerialRED for rotation electron diffraction.[11] Smeets *et al*. used Instamatic to determine two structures from four phases in a multiple phase system.[12] Jones *et al*. distinguished four natural products from a heterogeneous powder mixture,[13] while Ge *et al*. identified two different zeolitic-imidazolate frameworks from phase mixtures.[14] Broadhurst *et al*. identified four forms of carbamazepine based on their unit-cell dimensions and processed them to structures of the respective forms.[15] Luo *et al*. automatically examined hundreds of crystals using SerialRED and identified five similar zeolite phases,[16] and Sasaki *et al*. found two synthesized phases using SerialEM in diffraction mode, [17] one of which led to a MicroED structure. These studies demonstrate the potential of high throughput automation in collecting and analyzing data in MicroED and other structural biology methods and highlight the ability of automated methods to accelerate the discovery of new crystal forms and improve our understanding of complex material structures.

Here we present a method for compositional analysis of complex mixtures comprising nine compounds per sample at varying percentages. The compounds in each group possess comparable physical and chemical characteristics such as weight, hydrophilicity, and charge distribution, which are standard distinguishing factors in most analytical methods. Some compounds have the same chemical formula but differ in stereochemistry (four compounds) or constitutional isomers (three compounds), making separation challenging for typical analysis methods. We developed a high throughput autonomous MicroED approach to enable identification of all constituents and their relative ratios, mostly with high accuracy of just a few percent of their total composition. To demonstrate the extension of MicroED's capabilities using this approach, we further applied the method to an additional mixture. We identified all the ingredients in a mixture of nine compounds with a wide distribution of amounts per weight in the composition, with the smallest amount of constituents being around 3% by total mass.

Results & Discussion

Identifying compounds in a complex mixture using high throughput autonomous MicroED

We developed new procedures in SerialEM and coupled them to a Python script that we developed to autonomously collect and process MicroED data generated from them in real time. The method involving collecting an atlas montage of the entire grid, followed by medium-magnification montages of the most promising grid squares and finally applying the data collection script to the crystals selected from the medium-magnification montages (Figure 1). We used this approach to determine the components in complex mixtures. We used several mixtures in this study as proof of principle. Mixture A contained a several inorganic salts while mixtures B and C contained several saccharides and amino acids, respectively. These mixtures were analyzed using the pipeline to identify all their components.

Mixture A was prepared with varying amounts of salts ranging from 3.7 to 29.3 mg, totaling to 112.0 mg (Table S1). The mixture composed of Sodium bicarbonate (14.4% *v/v*), Sodium sulfate (21.3% *v/v*), Potassium sulfate (14.2% *v/v*), Magnesium sulfate heptahydrate (6.7% *v/v*), Sodium chloride (5.4% *v/v*), Calcium gluconate (6% *v/v*), Sodium citrate dihydrate (15.4% *v/v*), Calcium acetate monohydrate (4.7% *v/v*), Magnesium acetate tetrahydrate (11.9% *v/v*). When this mixture was loaded to the microscope, several crystals appeared ”melted” as their edges were not sharp suggesting that somehow in the process hydration may have occurred which may lead to a loss in crystal order and resolution. Hydration could have happened in the test tube prior to grid preparation. We also observed that several crystals clumped together so they were not suitable for analysis. Regardless of these concerns we selected 1961 crystals for autonomous diffraction and out of these only 913 (47%) delivered sufficiently good data for identification using unit cell parameters as discriminators (Table S2). Since we also recorded the images of the crystals, we could go back and look at the morphologies of the crystals that did not diffract. Indeed, most of the problematic crystals had the issues described above suggesting that sample preparation should be improved in the future (Figure S1 in Supplementary Information). Despite this concern we could identify all components in the mixture and even identify the constituent with lowest percentage. Calcium acetate monohydrate which was added at ~3.3% in the input was identified in 40 crystals corresponding to a 4.4% counting ratio (Figure 2).

In the second set of mixed compounds, a total of 9 different saccharides, along with a similar scaffold (L-Ascorbic acid), were weighed in with relative amounts ranging from 3.0 to 27.5% (Mixture B). The components of this mixture were: D-Glucose (4.2% *v/v*), D-Sucrose (3% *v/v*), D-Maltose monohydrate (10.3% *v/v*), L-Arabinose (4.2% *v/v*), L-Ascorbic acid (13.1% *v/v*), D-Galactose (6% *v/v*), D-Trehalose dihydrate (19.7% *v/v*), D-Xylose (11.9% *v/v*) and L-Rhamnose monohydrate (27.5% *v/v*) (Table S3, Supplementary Information). Separating some of the saccharides using size or affinity chromatography based on hydrophobicity or charge is challenging due to their chemical similarities. For instance, four of the saccharides, L-Arabinose, D-Xylose, D-Galactose, and D-Glucose, are pairwise chemically identical and differ only in the position of a hydroxyl group (diastereomers), which requires stereochemical separation. Trehalose and Maltose are also chemically identical, consisting of two glucose entities and differing only at the glycosidic linkage sites of the two pyranose rings. Similarly, sucrose is only one methylene group larger than maltose, which is similar in size. However, the crystallographic unit cells for all of these compounds are distinct from each other and easily distinguished in MicroED, making crystallographic analysis of chemically identical but stereochemically different compounds possible.

For the saccharides mixture (B), we collected 1374 data sets autonomously, of which 1000 (73%) were successfully identified as having one of the unit cells in the mixture. This was a substantial increase in success rate from 47% in mixture A to 73% in mixture B. As with Mixture A, the remaining 27% either did not diffract well possibly due to hydration, or originated from crystal clumps or were not crystalline to begin with. Importantly, all components of the mixture could be identified and for most the composition was determined within a 2% error. Only two compounds, L-Ascorbic acid and D-Trehalose dihydrate, demonstrated larger errors of 5.1% and 7.2%, respectively. Although the reason for these discrepancies is not known, it was noted in other work that crystals from a few compounds seem to have a tendency to lose their diffraction properties upon grinding (personal communication)[17]. We speculate that this could be linked to variations in crystal water or hygroscopic properties of the compound. The carboxylic group of citric acid is slightly hygroscopic (can attract the equivalent of one water per molecule in equilibrium in), while D-Trehalose already contains crystalline water, whereas many sugar compounds are less hygroscopic. However, we also note that deviations in the ratios of the constituents of the mixtures analyzed may come from differences in the size distributions of the sample crystals after grinding, and damages to the crystals during sample preparation. It is also likely that materials adhere differently to the grid, and this could also lead to minor errors. Overall, using MicroED in an automatic setting, only one grid containing the mixture was prepared and mounted, from which after processing the relative ratios could be derived (Figure 3A; Table S4 in Supplementary Information).

We next applied the approach to a third class of compounds – amino acids. Mixture C contained L-Glutamic acid (17.7% v/v), L-Alanine (5.4% v/v), L-Tyrosine (8.6% v/v), L-Serine (5% v/v), L-Valine (12.7% v/v), L-Cysteine (23% v/v), L-Threonine (5% v/v), L-Aspartic acid (10.2% v/v), L-Glutamine (12.4% v/v) (Table S3, Supplementary Information). We selected 1401 crystals for autonomous MicroED analyses and this time the success rate was even higher than for Mixture B as 1121 out of 1401 (80%) crystals were identified. The area-corrected ratios of compounds in mixture C were found to be similar to the ratios obtained by counting the number of crystals, with most ratios within 2% in relative amounts. However, for L-Tyrosine, L-Aspartic acid, and L-Glutamine, which had larger relative errors, the area corrected ratios tended to be better than the uncorrected ones, particularly for L-Tyrosine, where the overestimated ratio dropped from 10.6% to 8.2% after area correction. The estimated ratios are surprisingly close to the weighed-in ratios in this analysis (Figure 3B; Table S4 in Supplementary Information). These results suggest that sample preparation and the material used for grid preparation are the key determinants of success. It is also possible that the amorphous carbon support used in these experiments is more suitable for biological material like amino acids.

Although only a small wedge of data is collected from each crystal, data sets can be merged, and structures solved as needed. We applied this approach to mixture C and were able to solve all components in the mixture to sub atomic resolution (Figure 4; Table S5 in Supplementary Information), with completeness levels greater than 80% and more than 2000 observed reflections in most cases even for these small molecules.

Conclusion

In summary, the high throughput autonomous MicroED approach established and demonstrated in this study has great potential for expanding the applications of MicroED as an analytical tool well beyond a structural determination tool. With the ability to collect and analyze vast amounts of data, MicroED can be used for compositional analysis, providing a reliable and statistically significant analysis of the relative composition of a sample.

Methods such as chromatography and mass spectrometry[7,8] separate molecules based on difference in their sizes or chemical properties such as charge or hydrophobicity. For structures similar in size and chemical properties, like with isomers, separation is difficult. For crystalline samples the unit cell of a compound is a result of the exact packing of the molecules in order to lower the solid-state energy and is a stereoselective 3D process[10]. Therefore, chemically similar compounds usually result in a very different crystal packing and sometimes also a different symmetry between the molecules, resulting in differences in space groups and unit cell parameters. As these can be determined rather accurately from processing the diffraction movies, the cells can easily be distinguished. Should the unit cell parameters accidentally be very close to each other for any two crystals, as a next step the difference between the individual intensities of two data sets can be used to discriminate between two structures. The intensity of a reflection hkl is a result of the content of the unit cell, and the corresponding hkl of another data set can vary substantially with only a minor difference in the unit cell content[24].

However, further improvements in sample preparation[13] are necessary to fully exploit the potential of MicroED for analytical purposes. The experiments conducted using the three mixtures have provided us with valuable insights into the factors that may impact compositional analysis. It was observed during sample preparation that the crystals had different physical properties, which influenced the efficiency of grid preparation. For instance, some materials were harder and less prone to breaking into the required size for MicroED, while others were brittle and more easily prepared. Secondly, electrostatic charging of the powder during grinding can cause an excess of powder to attach to the interior of the vial rather than onto the EM grid. Thirdly, the physical shearing of the sample may impact the crystallinity, although this is more often experienced with protein samples rather than small molecules or salts.

Notably, we observed that around 70-80% of the selected crystals underwent automatic data processing and were successfully identified. We could conclude that a few crystals were too small to produce a high-quality diffraction pattern or were damaged prior to the experiment. We further noticed that the carbon support was not always flat even within a single grid square which could have led to errors in the estimation of the eucentric height of individual crystals. As a result a few crystals partially rotated out of the beam center during data collection. It is likely that with improved crystal selection criteria and a local eucentric height determination, the throughput could be further improved to the limitations of the sample itself, determined by variations in the diffraction properties of individual crystals within the sample.

The main purpose of the data collection here was to analyze the constituents rather than to solve the structures, which is why a short rotation range of only 50° per crystal was used instead of the typical larger rotation range that would be more appropriate for structure determination. This decision facilitated the collection of more data sets within the same time frame, increasing the statistical significance of the analysis. Attempts to use much shorter rotation wedges underlined the difficulties to index the data sets reliably due to the flatness of the Ewald sphere. As an example at the extreme end, no rotation was used in another study were instead prior knowledge of the unit cell parameters was used to compensate for the lack of rotation data[**27**]. Our approach which involved continuous rotation of at least 20° per crystal) allows unit cell determination without any *a priori* knowledge[1,2] which is more suitable for analytical purposes. Our study also demonstrates that it is not necessary to solve the structures for compositional analysis, as proper unit cell parameters are sufficient to distinguish between the compounds in the mixtures. This approach also allows for a more comprehensive analysis of the sample when the structures of the constituents are already known, as indexing a data set is typically easier than solving the structure *ab initio*.

Recent trends in structural biology emphasize the importance of collecting multiple sets of redundant data, especially for systems with low signal-to-noise ratios. This has significant implications for MicroED. Firstly, for complex systems where only a fraction of a sample diffracts to the desired resolution, a major part of the data collection process is dedicated to finding the best crystals. Secondly, redundant data can improve the signal-to-noise ratio and reduce the impact of systematic artifacts during data collection[28]. Finally, merging datasets to increase completeness is often necessary, but the outcome of this approach depends on the isomorphism of the data sets[29].In particular, since phasing using *ab initio* methods depends entirely of the structure factors, reliable estimates are required. A large pool of data sets facilitates the prospect of finding isomorphous data sets that could then be merged together to produce a complete data set for structure determination. This approach was demonstrated here with Mixture C where despite the use of a short rotation range, all structures in mixture C were successfully solved as a step toward future high throughput MicroED structure determination.

This study presents an autonomous approach to MicroED using commonly used and freely available software. The data collection process requires no human intervention after initial setup, making it suitable to run during less busy microscope shifts. This approach is built on the widely distributed CryoEM data collection software, SerialEM, which is already installed at many CryoEM labs. The hope is that by using commonly available software, this will inspire more laboratories to implement a higher level of automation in their MicroED data collection processes.

Methods

Materials

All the compounds were commercially purchased and used as received without further recrystallization. D-Glucose, D-Sucrose were purchased from Acros Organics. L-Valine was purchased from Alfa Aesar. Sodium bicarbonate, Magnesium sulfate heptahydrate, Sodium citrate dihydrate, Calcium acetate monohydrate, D-Galactose, L-Ascorbic acid were purchased from Fisher Chemical. Sodium sulfate, Potassium sulfate, Sodium chloride, Calcium gluconate, Magnesium acetate tetrahydrate, D-Maltose monohydrate, D-Trehalose dihydrate, L-Alanine, L-Arabinose, L-Aspartic acid, L-Cysteine, L-Glutamic acid, L-Glutamine, L-Rhamnose monohydrate, L-Serine, L-Threonine, L-Tyrosine were purchased from Sigma-Aldrich. D-Xylose was purchased from Tokyo Chemical Industry (TCI).

Sample Preparation

Three mixtures of compounds were carefully weighed by a Mettler Toledo (XPR225DR) analytical balance and mixed in a 20 mL scintillation vial. Mixture A was prepared as follows: Sodium bicarbonate (16.32 mg), Sodium sulfate (29.31 mg), Potassium sulfate (19.49 mg), Magnesium sulfate heptahydrate (9.14 mg), Sodium chloride (6.07 mg), Calcium gluconate (5.14 mg), Sodium citrate dihydrate (13.92 mg), Calcium acetate monohydrate (3.66 mg), Magnesium acetate tetrahydrate (8.93 mg). Mixture B was prepared as follows: D-Glucose (5.31 mg), D-Sucrose (3.88 mg), D-Maltose monohydrate (14.7 mg), L-Arabinose (5.42 mg), L-Ascorbic acid (17.55 mg), D-Galactose (7.3 mg), D-Trehalose dihydrate (28.05 mg), D-Xylose (14.73 mg), L-Rhamnose monohydrate (32.73 mg). Mixture C was prepared as follows: L-Glutamic acid (22.72 mg), L-Alanine (6.45 mg), L-Tyrosine (10.47 mg), L-Serine (6.62 mg), L-Valine (13.04 mg), L-Cysteine (24.95 mg), L-Threonine (5.5 mg), L-Aspartic acid (14.11 mg), L-Glutamine (14.13 mg) (See Tables S1 and S3 in Supplementary Information). The volume percent of each compound was calculated to generate a wide ratio range, from 3.0 to 27.5%. The compounds in the respective mixture were mixed together

before being ground separately by an agate mortar and pestle set (internal diameter 50 mm) three times at room temperature to yield a fine powder without any visible crystalline solids left. The total weight of each mixture was more than 100 mg to ensure a thorough interaction with the agate mortar and pestle during the grinding process.

Grid Preparation

The carbon-coated copper grids (400-mesh, 3.05 mm O.D., Ted Pella Inc.) were pretreated with glow-discharge plasma at 15 mA on the negative mode using PELCO easiGlow (Ted Pella Inc.), with no glow discharge for mixture A grid, 60s glow discharge time for mixture B grid, and 30s for mixture C grid. Around 1 mg powder from each set was transferred to a 10 mL scintillation vial and separately mixed with the grid. After a gentle shaking of the vial, the grids were taken out and clipped at room temperature.

Automatic MicroED Data Collection with SerialEM.

The clipped grids were loaded in an aligned Thermo Fisher Talos Arctica Cryo-TEM (200 kV, ~0.0251 Å) at 100 K, equipped with a Falcon III direct electron detector (4096 × 4096 pixels).[21] Intensity of 45.2% was found to be the condition for parallel beam during diffraction using the contrast of the objective aperture.[30] For diffraction movies the data was collected with an 829 mm diffraction length and using the smallest C2 aperture of 20 µm without the selected area aperture. The resulting beam size was approximately 1.5 µm. MicroED data was automatically collected using the SerialEM software in microprobe mode.

An atlas over the entire grid was collected as a low-magnification montage of 8×8 tiles. In the next step, typically more than 150 grid squares of interest were selected, and points were saved using “Add Points” in the SerialEM navigator window. After aligning the atlas with the magnification used in the medium magnification montages, one medium magnification montage of 3×3 pieces was collected at all saved points using the View mode. The “Rough eucentricity” and “Fine eucentricity” functions in SerialEM were used to automatically assign the eucentric height to each grid square that was stored with the corresponding maps. A point was added for each crystal of interest in the medium montage maps and saved in the navigator window. All points in the generated list were set up for data collection using the command “Acquire at items” function in the Navigator menu of SerialEM. Crystals ranging from 0.2 to 1.5 µm (size of the parallel beam) were picked for the data collection. Diffraction data using continuous rotation of the stage at 2°/s covering a total rotation range from -25° to +25° was automatically collected for each crystal using a SerialEM macro script. The script also used image mode to save an image of the content in the beam using the Search preset and was used to estimate crystal area and to visualize the crystal position during the MicroED data acquisition (Figure 1). During the rotation, the camera integrated frames continuously at a rate of ~0.5s per frame, a total of 24.95s exposure time for 50 frames.

Automatic Processing

An in-house developed python script automatically processed the MicroED data via three steps using available software: (1) the raw MicroED data in MRC format were automatically converted to SMV format using mrc2smv software (https://cryoem.ucla.edu/microed); (2) the converted data were processed in XDS and (3) data sets were merged using XSCALE. XDS used a few typical settings in input: the detector distance is not refined together with the unit cell refinement due to the flatness of the Ewald’s sphere, DELPHI was set to 30°, maximum errors was set to 10 and 3 for spot and spindle respectively, and occasionally we used

MINIMUM_FRACTION_OF_INDEXED_SPOTS equals 0.1 to include weak data. (See "XDS.INP template", Supplementary Information). The script evaluates different settings in XDS input for STRONG_PIXEL, SIGNAL_PIXEL, MINIMUM_NUMBER_OF_PIXELS_IN_A_SPOT, OFFSET, DATA_RANGE, SPOT_RANGE. The best merged data was found by evaluating all combinations of data sets to a certain maximal number of data sets included and the solutions were scored using statistics from XSCALE. The structures were solved using SHELXT[31] and refined with SHELXL[32].

Compositional Analysis and Crystal Area Estimation

The resulting unit cell parameters were compared to the literature-reported unit cells in Cambridge Structure Database (CSD) to assign the composition compound and crystal phase to the data. For the assignments of the proper crystal structure to an indexed data set, a maximum 1Å length tolerance and 10° angle tolerance were used (see Figures S2 and S3; Table S3, Supplementary Information). These limits ensured that crystals were appropriately identified by the unit cells within reasonable error. The pattern of the unit cell parameters was unique for every crystal form and with no overlap between any two unit cells and could therefore be used to unambiguously assign the proper crystal structure to any indexed data set.

The search images as saved in MRC format were converted to TIFF format using mrc2tif software.[33] Then the converted images were imported into ImageJ software[34] with the Threshold values to be adjusted until the whole crystals were fully detected (colored in red, Figure 1). In ImageJ, the "Analyze Particles" function was used to automatically integrate, the crystal and each outline of integration (colored in black, Figure 1) was manually inspected to ensure that only the crystal area was integrated. The crystal area was calculatedSupplementary and summed for each composition (see above), as S_{Comp} and S_{Total} for ratio analysis.

The ratio of a compound in the mixture was calculated by the equation (1) as shown below:

$$Ratio_{Est} = \frac{S_{Comp}}{S_{Total}} \cdots\cdots (1)$$

The ratio of the total area of one component (S_{Comp}) to the total area of all crystals (S_{Total}) for successfully indexed data sets only. The input ratio and observed ratio were plotted by Graphpad Prism 8.0.1 for Windows (GraphPad Software, San Diego, California USA, www.graphpad.com) as shown in Figure 3.

Notes

The authors declare no competing financial interest.

Supplementary information

Text, experimental details and additional data, including Tables

S1-S6 and Figures S1−S3 (***xxx***)

Author Information

Corresponding Authors

*tgonen@g.ucla.edu

ORCID

Johan Unge: 0000-0002-1077-8137

Jieye Lin: 0009-0003-9008-5115

Sara J Weaver: 0000-0001-7753-6215

Ampon Sae Her: 0000-0001-9594-8678

Tamir Gonen: 0000-0002-9254-4069

Author Contribution

J.U. and J.L contributed equally. J.U. designed the experiments, developed the sample preparation techniques and the data collection workflow, wrote the Python script for automatic processing, participated in data analysis, managed the activities and assisted in manuscript preparation. J.L. performed the sample preparation, performed the data collection, analyzed the data and prepared the tables, performed the refinement and structure determination, prepared the figures and assisted in manuscript preparation. S.W. prepared the initial setup of the microscope workflow and data processing. A.S. provided experience in sample preparation, data analysis and microscope setup. T.G. conceived of the project, designed experiments, provided expertise, assisted with manuscript preparation and supervised the project.

Acknowledgments

The authors thank Johan Hattne for the setup as well as providing the support for the computer cluster and its SLURM interface used for the processing in this work. This study was supported by the National Institutes of Health P41GM136508.Portions of this research or manuscript completion were developed with funding from the Department of Defense grants MCDC-2202-002 and HDTRA1-21-1-0004. Effort sponsored by the U.S. Government under Other Transaction number W15QKN-16-9-1002 between the MCDC, and the Government. The US Government is authorized to reproduce and distribute reprints for Governmental purposes, notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the U.S. Government. The PAH shall flow down these requirements to its sub awardees, at all tiers. The Gonen laboratory is supported by funds from the Howard Hughes Medical Institute.

Reference

1 Shi, D., Nannenga, B. L., Iadanza, M. G. & Gonen, T. Three-dimensional electron crystallography of protein microcrystals. *elife* **2**, e01345 (2013).

2 Nannenga, B. L., Shi, D., Leslie, A. G. W. & Gonen, T. High-resolution structure determination by continuous-rotation data collection in MicroED. *Nat. Methods.* 11, 927-930 (2014).

3 Martynowycz, M. W., Clabbers, M. T. B., Unge, J., Hattne, J. & Gonen, T. Benchmarking the ideal sample thickness in cryo-EM. *Proc. Natl. Acad. Sci. U.S.A.* 118, e2108884118 (2021)

4 Chapman, H. N. Structure determination using X-ray free-electron laser pulses. *Protein Crystallography: Methods and Protocols*, 295-324 (2017).

5 Diederichs, K. & Wang, M. Serial synchrotron X-ray crystallography (SSX). *Protein Crystallography: Methods and Protocols*, 239-272 (2017).

6 Henderson, R. The potential and limitations of neutrons, electrons and X-rays for atomic resolution microscopy of unstained biological molecules. *Q. Rev. Biophys.* 28, 171-193 (1995).

7 Kumar, K. J. & Vijayan, V. An overview of liquid chromatography-mass spectroscopy instrumentation. P*harmaceutical methods* , 5, *47* (2014).

8 Rathod, R. H., Chaudhari, S. R., Patil, A. S. & Shirkhedkar, A. A. Ultra-high performance liquid chromatography-MS/MS (UHPLC-MS/MS) in practice: Analysis of drugs and pharmaceutical formulations. *Future J. Pharm. Sci.* 5, 1-26 (2019).

9 Li, X. & Hu, K. Quantitative NMR studies of multiple compound mixtures. *Annu. Rep. NMR Spectrosc.* 90, 85-143 (2017).

10 Abelian, A., Dybek,M.,Wallach, J, Gaye, B. & Adejare, A. Theory of isomorphous replacement for protein crystals. *Remington (Twentythree Edition), The Science and Practice of Pharmac*y 105-128 (2021)

11 Wang, B., Zou, X. & Smeets, S. Automated serial rotation electron diffraction combined with cluster analysis: an efficient multi-crystal workflow for structure determination. *IUCrJ* 6, 854-867 (2019).

12 Smeets, S., Zou, X. & Wan, W. Serial electron crystallography for structure determination and phase analysis of nanocrystalline materials. *J. Appl. Crystallogr.* 51, 1262-1273 (2018).

13 Jones, C. G. *et al*. The CryoEM method MicroED as a powerful tool for small molecule structure determination. *ACS Cent. Sci.* 4, 1587-1592 (2018).

14 Ge, M. *et al*. High-Throughput Electron Diffraction Reveals a Hidden Novel Metal–Organic Framework for Electrocatalysis. *Angew. Chem., Int. Ed.* **133**, 11492-11498 (2021).

15 Broadhurst, E. T., Xu, H., Parsons, S. & Nudelman, F. Revealing the early stages of carbamazepine crystallization by cryoTEM and 3D electron diffraction. *IUCrJ* **8**, 860-866 (2021).

16 Luo, Y. *et al*. High-throughput phase elucidation of polycrystalline materials using serial rotation electron diffraction. *Nat. Chem.* **15**, 483-490 (2023).

17 Sasaki, T., Nakane, T., Kawamoto, A., Nishizawa, T. & Kurisu, G. Microcrystal Electron Diffraction (MicroED) Structure Determination of a Mechanochemically Synthesized Co-crystal not Affordable from Solution Crystallization. *CrystEngComm* (2023).

18 Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36-51 (2005).

19 Schorb, M., Haberbosch, I., Hagen, W. J. H., Schwab, Y. & Mastronarde, D. N. Software tools for automated transmission electron microscopy. *Nat. Methods.* **16**, 471-477 (2019).

20 de la Cruz, M. J., Martynowycz, M. W., Hattne, J. & Gonen, T. MicroED data collection with SerialEM. *Ultramicroscopy* **201**, 77-80 (2019).

21 Hattne, J., Martynowycz, M. W., Penczek, P. A. & Gonen, T. MicroED with the Falcon III direct electron detector. *IUCrJ* **6**, 921-926 (2019).

22 Kabsch, W. xds. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125-132 (2010).

23 Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 133-144 (2010).
24 Crick, F. H. C. & Magdoff, B.S. The Theory of Method of Isomorphous Replacement for Protein Crystals. *Acta Crystallogr.,* **9**, 901 (1956),
25 Powell, S. M., Novikova, I.V., Kim, D. N. & Evans, J. E. AutoMicroED: A semi-automated MicroED processing pipeline. *doi: https://doi.org/10.1101/2021.12.13.472146*
26 Brázda,P., Klementová, M.,Krysiak, Y. & Palatinus, L., Accurate lattice parameters from 3D electron diffraction data. I. Optical distortions*, IUCrJ* 9, 735-755, (2022)
27 Bücker, R., Hogan-Lamarre, P., Mehrabi, P. *et al.* Serial protein crystallography in an electron microscope. *Nat Commun* **11**, 996 (2020).
28 Karplus PA, Diederichs K. Assessing and maximizing data quality in macromolecular crystallography. *Curr Opin Struct Biol.* 2015 Oct;**34**:60-8.
29 Foadi J, Aller P, Alguel Y, Cameron A, Axford D, Owen RL, Armour W, Waterman DG, Iwata S, Evans G. Clustering procedures for the optimal selection of data sets from multiple crystals in macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr.* 69(Pt 8):1617-32. (2013)
30 Herzik, M. A. Setting up parallel illumination on the Talos Arctica for high-resolution data collection. *cryoEM: Methods and Protocols*, 125-144 (2021)
31 Sheldrick, G. M. SHELXT–Integrated space-group and crystal-structure determination. *Acta Crystallogr. A: Found. Adv.* **71**, 3-8 (2015).
32 Sheldrick, G. M. Crystal structure refinement with SHELXL. *Acta crystallogr. C Struct. Chem.* **71**, 3-8 (2015).
33 Kremer, J. R., Mastronarde, D. N. & McIntosh, J. R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
34 Schneider, C. A., Rasband, W. S. & Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods.* **9**, 671-675 (2012).
35 Dolomanov, O. V., Bourhis, L. J., Gildea, R. J., Howard, J. A. K. & Puschmann, H. OLEX2: a complete structure solution, refinement and analysis program. *J. Appl. Crystallogr.* **42**, 339-341 (2009).

Figures

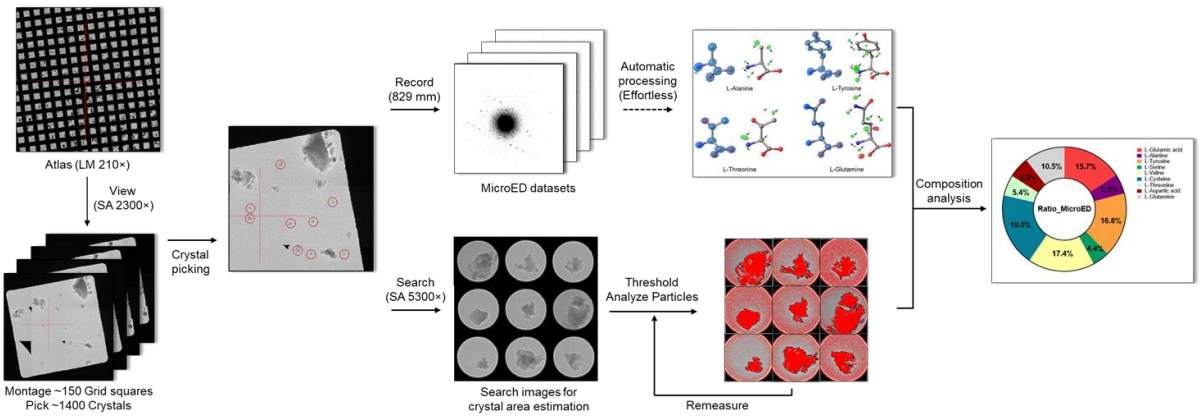

Figure 1. Workflow.

The workflow for autonomous MicroED data collection and processing involves several steps. First, a low-magnification atlas is used to screen the grid containing a mixture of compounds, and suitable grid squares are selected for medium-magnification montages. Crystals are then chosen and added to the list for data collection. Next, continuous rotation MicroED movie and an image of the crystal are collected autonomously using SerialEM.[18-20] The data is then processed automatically in real time. Crystal volume is estimated for compositional analysis. This workflow allows for efficient and autonomous MicroED data collection and processing of thousands of data sets per night, reducing the need for human intervention and saving researchers' time while producing high-quality data.

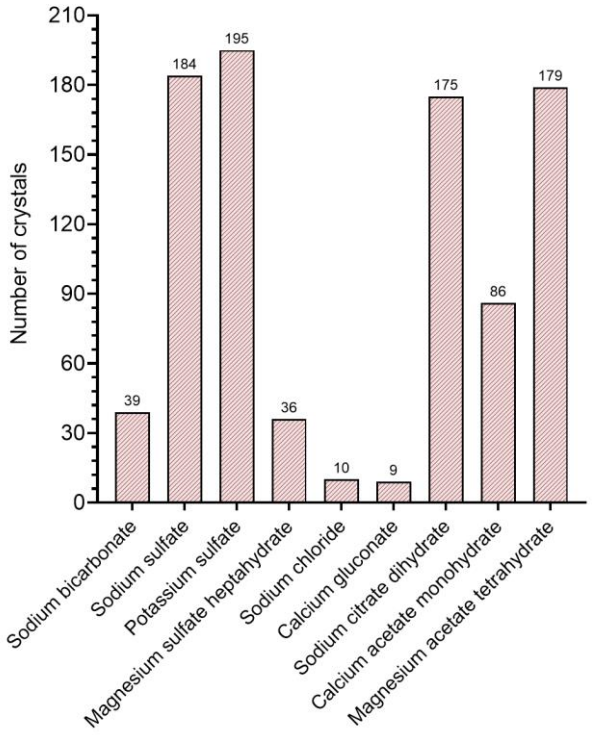

Figure 2. Compound identification.

The bars represent the number of crystals identified for each composition in mixture A. Despite the smaller number of data sets used, 913 in total, all compounds could be identified including the compounds with the lowest relative mass of 3% of the total weight of the composition.

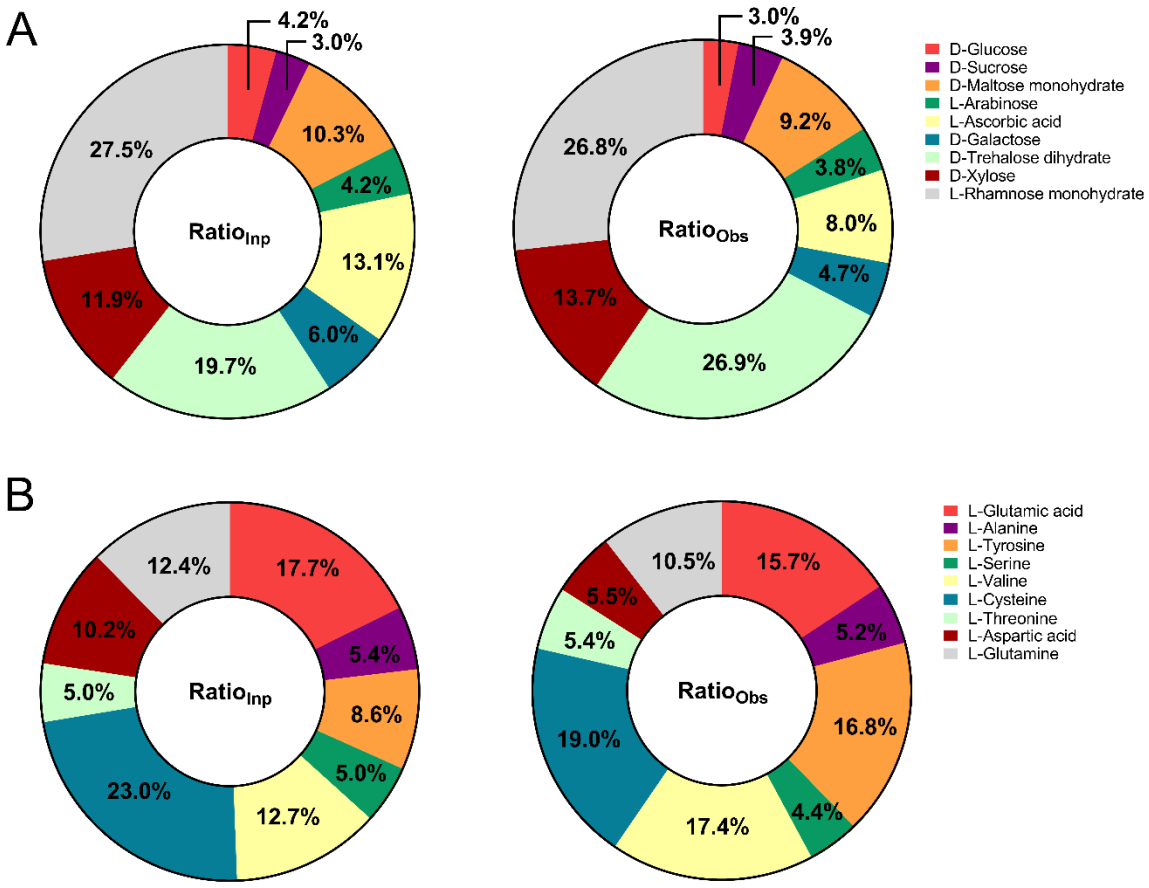

Figure 3. Compositional analysis

Compositional analysis of mixture B and C shows the relative volumes of the compounds in the actual mixture compared to the analysis based on autonomous MicroED data. In mixture B, the relative volumes of the compounds were found to be consistent between the original composition and the MicroED analysis, with only slight deviations for some compounds. The $Ratio_{Inp}$ represents the volume ratio (%) for each component in the original composition, while the $Ratio_{Obs}$ was calculated by dividing the ratio of the total area of one component to the total area of all crystals in the MicroED analysis. Overall, the results demonstrate the feasibility and accuracy of using MicroED data for composition analysis of complex mixtures.

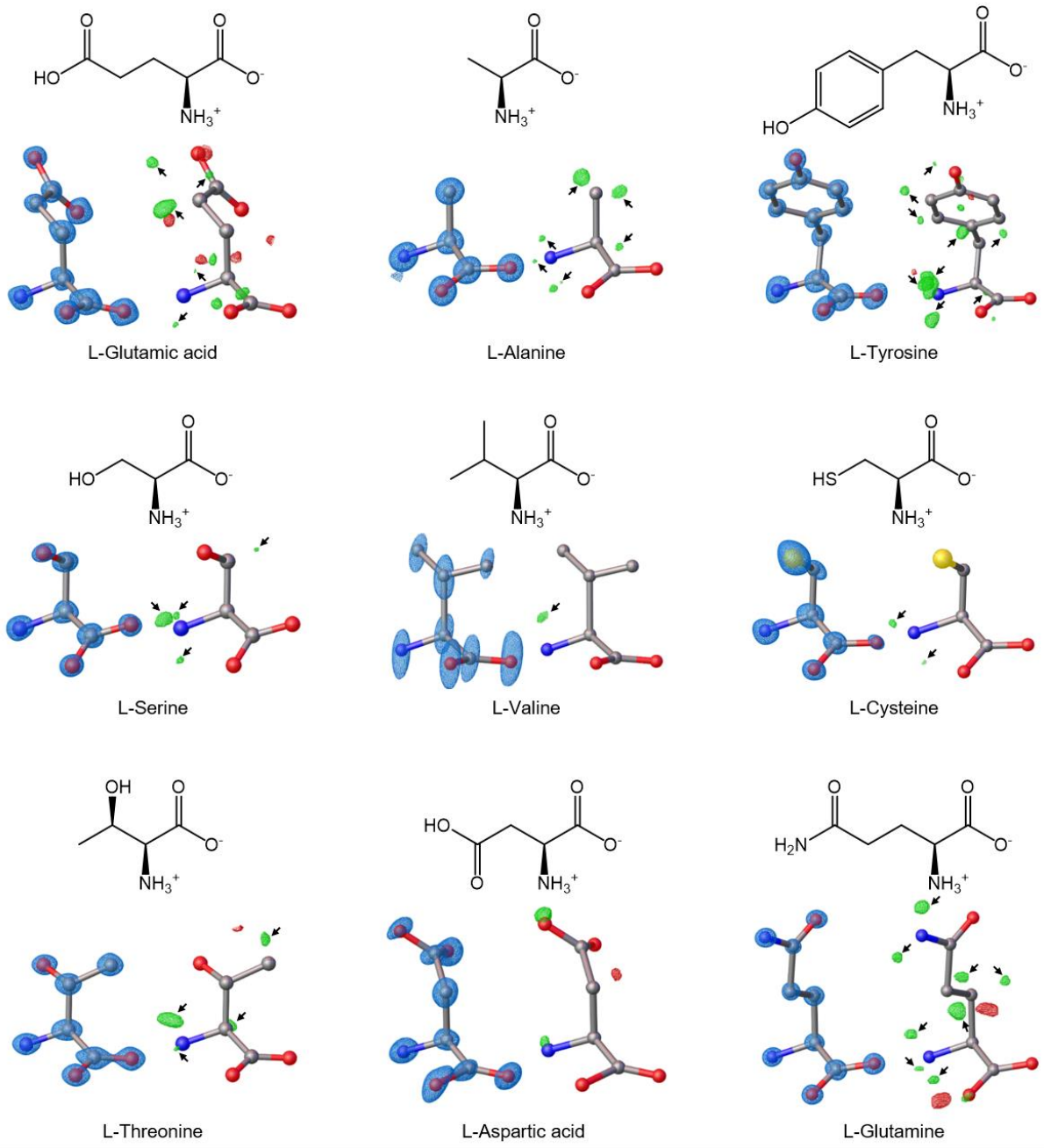

Figure 4. MicroED structures.

The MicroED structures of the nine amino acids in mixture C as solved by SHELXT[31] and refined using SHELXL[32] at 0.75 Å resolution. The blue meshes are $2F_o$-F_c electrostatic potential maps, green meshes are F_o-F_c electrostatic potential maps, the position of hydrogen atoms are marked with black arrows (presented by Olex2).[35] The electrostatic potential level for each compound in $2F_o$-F_c and F_o-F_c maps are as listed: L-Glutamic acid (0.85 and 0.21 e·Å^{-3}), L-Alanine (0.85 and 0.20 e·Å^{-3}), L-Tyrosine (0.85 and 0.15 e·Å^{-3}), L-Serine (0.85 and 0.20 e·Å^{-3}), L-Valine (0.46 and 0.09 e·Å^{-3}), L-Cysteine (0.85 and 0.31 e·Å^{-3}), L-Threonine (0.85 and 0.16 e·Å^{-3}), L-Aspartic acid (0.38 and 0.11 e·Å^{-3}), L-Glutamine (0.85 and 0.11 e·Å^{-3}).