Title: The CryoEM method MicroED as a powerful tool for small molecule structure determination

Authors: Christopher G. Jones^{1†}, Michael W. Martynowycz^{2†}, Johan Hattne², Tyler Fulton³, Brian M. Stoltz^{3\$}, Jose A. Rodriguez^{1\$}, Hosea M. Nelson^{1\$}, Tamir Gonen^{2\$}

5 Affiliations:

¹Department of Chemistry and Biochemistry, University of California, Los Angeles, 90095, USA

²Howard Hughes Medical Institute, Departments of Biological Chemistry and Physiology, University of California, Los Angeles, 90095, USA

³The Warren and Katharine Schlinger Laboratory of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, 91125, USA

^{\$}Correspondence to: <u>stoltz@caltech.edu</u>, jrodriguez@chem.ucla.edu, <u>hosea@chem.ucla.edu</u>, <u>tgonen@ucla.edu</u>

15

20

25

[†]These authors contributed equally.

Abstract: In the many scientific endeavors that are driven by organic chemistry, unambiguous identification of small molecules is of paramount importance. Over the past 50 years, NMR and other powerful spectroscopic techniques have been developed to address this challenge. While almost all of these techniques rely on inference of connectivity, the unambiguous determination of a small molecule's structure requires X-ray and/or neutron diffraction studies. In practice, however, x-ray crystallography is rarely applied in routine organic chemistry due to intrinsic limitations of both the analytes and the technique. Here we report the use of the CryoEM method MicroED to provide routine and unambiguous structural determination of small organic molecules. From simple powders, with minimal sample preparation, we could collect high quality MicroED data from nanocrystals (~100x100x100 nm, ~10⁻¹⁵ g) resulting in atomic resolution (<1 Å) crystal structures in minutes.

Main Text: The history of organic chemistry closely parallels the development of new methods for structural characterization. The earliest studies were driven by melting point determination and over the past 175 years more complex methods for interrogation of structure have been developed. Techniques such as polarimetry (1), UV-vis (2), and infrared spectroscopy (3), coupled with EPR (4), VCD (5), CD (6), and mass spectrometry(7) have been commonly employed over the years, dramatically expanding our ability to assign structures. In the past 50 years, however, the explosion of NMR spectroscopy (8) and the accompanying abundance of individual NMR experiments have produced a wealth of structural information. Indeed, NMR is a mainstay in chemistry and the most predominant method employed in both routine synthetic chemistry experiments and in advanced structural elucidation of complex small molecules. In the current state of the art, only single crystal X-ray diffraction holds a higher place in terms of precision, producing unequivocal structural information about the position, orientation, connectivity, and placement of individual atoms and bonds within a given molecule.

For decades, small molecule x-ray analysis has been the definitive tool for structural analysis (9). This technique, however, is not without limitations. The process is considered by many an *art*, where the production of high quality crystals suitable for x-ray diffraction requires uncodified "tricks of the trade" and a certain amount of luck! Additionally, even once a substance has been successfully crystallized, there is no guarantee that the particular crystal form will be amenable to x-ray diffraction. Since crystal growth is slow and arduous, X-ray diffraction has not been useful for on-the-fly structure determination of chemical structure. For this reason, X-ray diffraction is not routinely used by most practicing organic chemists, despite the fact the structural data provided is superior to any other method thus far.

Herein, we employ the recently developed cryoEM method MicroED (10) to address the need for fast and reliable structure determination in organic chemistry. We demonstrate that with minimal sample preparation and experiment time, simple powders and amorphous materials (in some cases, solids isolated *via* silica gel chromatography and rotary evaporation) could be directly used in MicroED studies, rapidly leading to high quality molecular structures often at atomic resolutions (better than 1Å). MicroED has the potential to dramatically accelerate and impact the fields of synthetic chemistry, natural product chemistry, drug discovery, among many others, by rapidly delivering high quality atomic resolution structures of complex, small molecules.

We initially tested the applicability of MicroED to a model system, progesterone (1) (Fig. 1). The sample was obtained as a powder from chemical supplier Preparations Laboratories Inc. Small quantities of this amorphous solid were placed in between two glass cover slides and crushed. The fine powder was deposited on a holey carbon copper grid, cooled to liquid nitrogen temperatures, and transferred to a cryo electron microscope operating at an acceleration voltage of 200kV (Thermo Fisher Talos Arctica). An overview of the preparation is shown in Figure 1. Thousands of nano crystals were easily discernable and were investigated for diffraction. Typically, for samples such as these, the vast majority of nano crystals diffracted to ~ 1 Å





15

5

10

resolution or better (Fig. 1). 140 degrees of diffraction data could be collected from a single nano crystal by continuous rotation (11) as the improved autoloader and piezo stage of the Talos Arctica allowed us to travel through the 0 degree point without introducing errors in crystal position. Typically, the stage was rotated at 0.5 degrees per second and an entire data set collected in less than 3 minutes as a movie (video 1) using a bottom mount CetaD CMOS detector. This detector was fitted with a thick scintillator for diffraction studies. Software written to convert the movie frames into SMV format allowed for processing in XDS(12). From the collected data, the structure of steroid 1 was determined to 1Å resolution from a single nano crystal. The entire process from powder to structure took less than 30 minutes.

Encouraged by these results, we wanted to explore the scope of this structural 10 determination method by investigating a wide range of small molecules (Fig. 2a). Up to 12 different samples could be loaded onto the Talos Arctica autoloader allowing us to investigate up to 12 different samples rapidly. Notably, over-the-counter medications, CVS[®] branded acetaminophen and Kroger[®] branded ibuprofen tablets were obtained as tablets and crushed as described above allowing us to obtain atomic resolution structures of acetaminophen (2) and 15 ibuprofen (3) in a rapid fashion. Importantly, we were able to determine structures of these active pharmaceutical ingrediaents despite the heterogeneity of the tablets resulting from the presence of binders and other formulation agents. The structures of the sodium channel blocker carbamazepine (4) and the macrocyclic polypetide antiobiotic thiostrepton (5) were also determined from seemingly amorphous powders, which were used as recieved from Sigma-20 Aldrich. We went on to study several natural products obtained from commercial sources. Used as recieved, without any crystallization, we were able to obtain atomic resolution structures of biotin (6), ethisterone (7), cinchonine (8), and brucine (9). Of the eleven different commercial bioactives examined, all eleven yielded MicroED data and ten were amenable to rapid structure determination by direct methods (13) while one was determined by molecular replacement (14). 25 Importantly, all were obtained without any crystallization attempts. While these pharmaceutical and commercial natural products were likely recrystallized for purification purposes by the manufacturer, the powders examined by MicroED were nano crystals a billionth the size (~100x100x100 nm) of crystals typically needed for X-ray crystallography. This was powder to structure. 30

Next we decided to investigate compounds that were never crystallized but instead were purified by flash column chromatography. Since, silica gel chromatography is the most common method of purification in early stage research for complex molecules in drug discovery, natural product isolation efforts, and in general synthesis efforts, we were interested to determine whether solid samples prepared in this way would be amenable to analysis by MicroED. Four compounds, purified by chromatographic methods, were collected from our laboratories and samples of these seemingly amorphous solids were analyzed. Here, two of four compounds diffracted and we were able to resolve both strucutres at atomic resolution (10 and 11, Figs. 2A and B). While the success rate for these compounds was 50%, it is worthy to note that no crystallization procedures were employed in the isolation of these materials. Notably, (-)limaspermidine (10), an alkaloid natural product synthesized by our labs (15), was resolved from a residue of only milligram quantities of material following flash chromotography and rotary evaporation from a scintillation vial (Fig. 2B). Electrons interact with matter more strongly than X-rays and they are affected by the charge (16). While it is extremely challenging to observe protons in X-ray structures, it is relatively common in MicroED data (17-20). Several protons were observed in all structures. For example, the density maps obtained for limaspermidine (10)

5

35



Fig. 2. Different types of small molecules solved by MicroED. A) Several Pharmaceutical, vitamins, commercial natural products and synthetic samples solveds by MicroED, and B) example of amorphous film utilized in this study leading to 1Å resolution MicroED data. C) many protons observed in compounds. Green spheres are Fo-Fc maps showing positive density belonging to hydrogens in the molecules.

and carbamazepine (4) after refinement allowed us to identify protons associated with almost all atoms in these molecules (Fig. 2C).

Astounded by the ease with which such high quality data was obtained and the apparent generality of MicroED to small molecules, we undertook studies to examine heterogeneous samples (mixtures of compounds). Here, single crystal x-ray diffraction precludes the study of mixtures and NMR is poorly suited for this task. Mixtures of four compounds (4,6, 8 and 9, *cf*.

Fig. 3) were crushed together and deposited on a holey carbon grid. Several crystal forms belonging to the different materials in the mixture, were visually identified on the grids (Fig. 3). MicroED data was collected from several nano crystals and the identity of each species identified within minutes based on the diffraction and unit cell parameters. Atomic resolution structures were determined for all small molecules in the mixture rapidly (Fig. 3).



Fig. 3. Identification of compounds from heterogeneous mixtures. EM grid prepared as above with biotin, brucine, carbamazepine, and cinchonine powders mixed together. All four compounds identified by unit cell parameters using MicroED data from within the same grid square. All structures were solved to ~1Å resolution. Grid holes are $2\mu m$ in diameter.

The results described here introduce a new characterization tool into the organic chemists toolbox. While MicroED was developed for structure determination of biological material such as proteins in a frozen hydrated state (21, 22), we demonstrate that cross pollination of macromolecular structural methods of cryoEM are powerful tools for chemical synthesis and drug characterization and discovery. MicroED has allowed for the structural charaterizarion of several proteins (21–23) from small crystals and crystals that are unsuitable for X-ray crystallography because of their size or pathologies (21), but the method has largely gone unnoticed in the small molecule communities. Based on our findings, we anticipate that MicroED will be enthusiastically received by many types of small molecule chemists. We have shown that a variety of amorphous solid materials can lead to rapid atomic resolution structure determination by MicroED with little or no additional sample preparation or crystallization. The fact that a solid film in a flask, following solvent removal from a flash chromatography purification, can lead to an atomic resolution molecular structure is evidence that MicroED will likely have a profound effect on the structural characterization work-flow of organic chemists. Although the past 50 years have seen huge advances in the state-of-the-art, no completely new

techniques have been introduced that alter the routine structural interrogation of organic substances. NMR (8), IR (3), UV-Vis (2), and X-ray diffraction (9) have been routinely in place since the 1960's and are still utilized today as the most common methods for structure determination in chemistry. We believe that MicroED (24, 25) is potentially the next big advance in the field and are enthusiastic about the prospects of expanding its utility as a routine method for chemists.

References and Notes:

5

10

15

20

25

30

- 1. P. Schreier, A. Bernreuther, M. Huffer, *Analysis of chiral organic molecules: methodology and applications*. (Walter de Gruyter, 2011).
- 2. A. I. Scott, Interpretation of the Ultraviolet Spectra of Natural Products: International Series of Monographs on Organic Chemistry. (Elsevier, 2013), vol. 7.
 - 3. J. Coates, Interpretation of infrared spectra, a practical approach. *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation* **12**, 10815–10837 (2000).
 - 4. D. A. Dougherty, Spin control in organic molecules. Acc. Chem. Res. 24, 88–94 (1991).
- 5. P. J. Stephens, F. J. Devlin, J. J. Pan, The determination of the absolute configurations of chiral molecules using vibrational circular dichroism (VCD) spectroscopy. *Chirality* **20**, 643–663 (2008).
 - 6. N. Berova, L. D. Bari, G. Pescitelli, Application of electronic circular dichroism in configurational and conformational analysis of organic compounds. *Chem. Soc. Rev.* **36**, 914–931 (2007).
 - 7. E. Hoffmann, Mass spectrometry. *Kirk Othmer Encyclopedia of Chemical Technology* (2000).
 - 8. H. Günther, *NMR spectroscopy: basic principles, concepts and applications in chemistry.* (John Wiley & Sons, 2013).
- 9. J. D. Dunitz, *X-ray Analysis and the Structure of Organic Molecules*. (Verlag Helvetica Chimica Acta, 1995).
 - 10. D. Shi, B. L. Nannenga, M. G. Iadanza, T. Gonen, Three-dimensional electron crystallography of protein microcrystals. *Elife*. **2013**, e01345 (2013).
 - B. L. Nannenga, D. Shi, A. G. W. Leslie, T. Gonen, High-resolution structure determination by continuous-rotation data collection in MicroED. *Nat. Methods.* 11, 927– 930 (2014).
 - 12. D. Shi *et al.*, The collection of MicroED data for macromolecular crystallography. *Nat. Protoc.* **11**, 895–904 (2016).
 - 13. G. M. Sheldrick, SHELXT Integrated space-group and crystal-structure determination. *Acta Crystallogr. Sect. A Found. Crystallogr.* **71**, 3–8 (2015).

35

- 14. A. J. McCoy *et al.*, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- 15. B. P. Pritchett, E. J. Donckele, B. M. Stoltz, Enantioselective Catalysis Coupled with Stereodivergent Cyclization Strategies Enables Rapid Syntheses of (+)-Limaspermidine and (+)-Kopsihainanine A. *Angew. Chem. Int. Ed.* **56**, 12624–12627 (2017).
- R. Henderson, 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).
- 17. S. Vergara *et al.*, MicroED Structure of Au146(p-MBA)57at Subatomic Resolution Reveals a Twinned FCC Cluster. *J. Phys. Chem. Lett.* **8**, 5523–5530 (2017).
- 18. J. A. Rodriguez *et al.*, Structure of the toxic core of α-synuclein from invisible crystals. *Nature*. **525**, 486–490 (2015).
- 19. M. R. Sawaya *et al.*, Ab initio structure determination from prion nanocrystals at atomic resolution by MicroED. *Proc. Natl. Acad. Sci.* **113**, 11232–11236 (2016).
- 20. J. Hattne *et al.*, Analysis of Global and Site-Specific Radiation Damage in Cryo-EM. *Structure* (2018), doi:10.1016/j.str.2018.03.021.
 - 21. M. J. de la Cruz *et al.*, Atomic-resolution structures from fragmented protein crystals with the cryoEM method MicroED. *Nat. Methods.* **14**, 399–402 (2017).
 - 22. B. L. Nannenga, D. Shi, J. Hattne, F. E. Reyes, T. Gonen, Structure of catalase determined by MicroED. *Elife*. **3**, e03600 (2014).
 - 23. M. Gallagher-Jones *et al.*, Sub-ångström cryo-EM structure of a prion protofibril reveals a polar clasp. *Nat. Struct. Mol. Biol.* (2018), doi:10.1038/s41594-017-0018-0.
 - 24. M. W. Martynowycz, T. Gonen, From electron crystallography of 2D crystals to MicroED of 3D crystals. *Curr. Opin. Colloid Interface Sci.* (2018), doi:10.1016/j.cocis.2018.01.010.
- 5 25. B. L. Nannenga, T. Gonen, Protein structure determination by MicroED. *Curr. Opin. Struct. Biol.* **27**, 24–31 (2014).

Acknowledgments: We thank Professors Doug Rees and Bil Clemmons for useful discussions. We thank Byungkuk Yoo and Michael Takase. for technical assistance with data analysis.
Funding: C. G. J. would like to acknowledge the National Science Foundation for a predoctoral fellowship. B. M. S. acknowledges the NIH-NIGMS for generous funding (R01GM080269). J. A. R. is supported by DOE grant DE-FC02-02ER63421, NIH-NIGMS grant R35GM128867, and as a Beckman Young Investigator, a Searle Scholar and a Pew Scholar. H. M. N. thanks The Packard Foundation, The Sloan Foundation, Pew Charitable Trusts, and the NIH-NIGMS (R35 GM128936) for generous funding. The Gonen laboratory is supported by funds from the Howard Hughes Medical Institute.; Author contributions: C.G. J. performed experiments, developed the

10

5

20

15

25

sample preparation techniques, refined structural data, prepared figures and assisted with manuscript preparation. M. W. M performed experiments, developed the sample preparation techniques, collected data, refined structural data, prepared figures and assisted with manuscript preparation. T. F. performed experiments. J. H. wrote the software for image conversion, participated in data analysis, refinement and structure determination. J. A. R. performed experiments, developed the sample preparation techniques, and assisted with manuscript preparation. B. M. S. conceived of the project, designed experiments, and assisted with manuscript preparation. H. M. N. conceived of the project, designed experiments, performed experiments, developed the sample preparation techniques, and assisted with manuscript and figure preparation. T. G. performed experiments, collected data, developed the sample prep techniques, developed microscope data collection parameters, provided the microscope and expertise, and assisted in manuscript and figure preparation. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** MicroED density maps have been deposited to the EMDB and CCDC.

15

5

Supplementary Materials for

The CryoEM method MicroED as a powerful tool for small molecule structure determination

Christopher Jones, Mike Martynowycz, Johan Hattne, Tyler Fulton, Brian M. Stoltz*, Jose A. Rodriguez*, Hosea M. Nelson*, Tamir Gonen*

*Corresponding author. Email: , <u>stoltz@caltech.edu</u>, <u>jrodriguez@chem.ucla.edu</u>, <u>hosea@chem.ucla.edu</u>, <u>tgonen@g.ucla.edu</u>

Table of Contents

1. Materials and Methods	1
1.1 Sample Preparation	1
1.2 Instrument Parameters, Data Collection, and Analysis	2
2.Compound Data and Statistics	3

1. Materials and methods

All commercial samples were used as received with no additional crystallization or chemical modification. Ethisterone, cinchonine, carbamazepine, and biotin were purchased from Sigma-Aldrich. Brucine was purchased from the The Matheson Company, Inc. Progesterone was purchased from Preparations Laboratories Inc. Thiostrepton was purchased from EMD Millipore. CVS[®] brand acetaminophen and Kroger[®] brand ibuprofen were used as over-the-counter medications. (-)-Limaspermadine and HML-I-029 were synthesized according to previously reported literature procedures (*1*,*2*).

1.1 Sample Preparation

To prepare commercial compounds for MicroED, approximately 1 mg of product as received was placed between two microscope slides and ground to a fine powder. The ground powder was placed into an Eppendorf tube along with a prepared TEM grid and shaken. The loaded TEM grid was then removed from the Eppendorf tube and gently tapped against a filter paper to

remove excess powder. Non-commercial samples of HML-I-029 and (+)-limaspermadine were concentrated under vacuum to yield a dry film and solid powder respectively. Sample grids of HML-I-029 were prepared by adding a TEM grid directly to a 20 mL scintillation vial with gentle shaking. (+)-Limaspermadine grids were prepared by scraping the residue off the side of a 20 mL scintillation vial over a TEM grid. Once sample grids were prepared, they were subsequently plunged into liquid nitrogen, placed into the sample cartridge, and loaded into the microscope for analysis. Heterogenous sample mixtures were prepared by adding ~1 mg of biotin, carabamazepine, cinchonine, and brucine to a glass cover slide and grinding to a fine powder. The heterogenous powder was then added to an Eppendorf tube and the grid was prepared in the same manner as the homogeneous samples.

1.2 Instrument Parameters, Data Collection, and Analysis

The holey carbon copper grid, cooled to liquid nitrogen temperatures, and transferred to a cryo electron microscope operating at an acceleration voltage of 200kV (Thermo Fisher Talos Arctica). An overview of the preparation is shown in Figure 1. 140 degrees of diffraction data could be collected from a single nano crystal by continuous rotation as the improved autoloader and piezo stage of the Talos Arctica allowed us to travel through the 0 degree point without introducing errors in crystal position. Typically, the stage was rotated at 0.5 degrees per second and an entire data set collected in less than 3 minutes as a movie using a bottom mount CetaD CMOS detector. This detector was fitted with a thick scintillator for diffraction studies. Software written to convert the movie frames into SMV format allowed for processing in XDS.





stoichiometric formula	C23 N2 O4
temperature (K)	100
Space group	P 21
Unit cell a, b, c (Å)	15.340(3), 7.540(2), 20.010(4)
angles α, β, ɣ (°)	90.00(3), 112.49(3), 90.00(3)
Reflections (#)	12427 (814)
Unique reflections (#)	5858 (416)
R obs	18.2 (56.1)
R meas	24.2 (74.9)
CC 1/2	95.1 (25.9)
Resolution (Å)	0.9
Completeness (%)	95.3 (96.1)
Total exposure (e ⁻ Å ⁻²)	~3
R	0.2244
wR2	0.4468
GooF	1.711

carba	mazepine	0.9 Å	
stoichiometric formula	C15 N2 O		
temperature (K)	100		
Space group	P 21/n		
Unit cell a, b, c (Å)	7.460(2), 11.040(2), 13.760(3)		
angles α , β , γ (°)	90.00(3), 92.61(3), 90.00(3)		
Reflections (#)	4682 (678)		
Unique reflections (#)	1044 (146)		
R obs	17.3 (22.1)		
R meas	19.5 (24.7)		
CC 1/2	97.3 (93.8)		
Resolution (Å)	1	U NH	
Completeness (%)	88.3 (84.9)		
Total exposure (e ⁻ Å ⁻²)	~3		
R	0.1931	57	
wR2	0.3902		
GooF	2.398		
		A A A A A A A A A A A A A A A A A A A	

cinc	honine
toichiometric formula	C19 N2 O1
temperature (K)	100
Space group	P 21/n
Unit cell a, b, c (Å)	10.710(2), 7.060(2), 11.150(2)
angles α, β, ɣ (°)	90.00(3), 109.66(3), 90.00(3)
Reflections (#)	1933 (399)
Unique reflections (#)	1289 (262)
R obs	11.0 (14.8)
R meas	15.6 (21.0)
CC 1/2	95.0 (89.2)
Resolution (Å)	1
Completeness (%)	77.4 (78.9)
Total exposure (e ⁻ Å ⁻²)	~3
R	0.1793
wR2	0.3907
GooF	1.831





ibu	profen
stoichiometric formula	C13 O2
temperature (K)	100
Space group	P 21/c
Unit cell a, b, c (Å)	14.65(3),7.88(2),10.73(2)
angles α, β, γ (°)	90.00(3),99.7(3) ,90.00(3)
Reflections (#)	1452 (402)
Unique reflections (#)	506 (138)
R obs	14.7 (20.8)
R meas	17.8 (25.2)
CC 1/2	97.8 (89.9)
Resolution (Å)	1.1
Completeness (%)	54.3 (53.1)
Total exposure (e⁻ Å⁻²)	~3
R	0.2559
wR2	0.5282
GooF	2.686

limasp	permidine
stoichiometric formula	C19 N2 O1
temperature (K)	100
Space group	P 21 21 21
Unit cell a, b, c (Å)	7.620(2), 13.880(3), 15.200(3)
angles α, β, γ (°)	90.00(3),90.00(3),90.00(3)
Reflections (#)	8252 (387)
Unique reflections (#)	3430 (185)
R obs	16.7 (68.2)
R meas	21.6 (88.6)
CC 1/2	97.0 (34.2)
Resolution (Å)	0.77
Completeness (%)	93.0 (69.3)
Total exposure (e ⁻ Å ⁻²)	~3
R	0.2422
wR2	0.4309
GooF	1.541

proge	esterone
stoichiometric formula	C21 O2
temperature (K)	100
Space group	P 21 21 21
Unit cell a, b, c (Å)	10.270(2), 12.680(3), 13.750(3)
angles α , β , γ (°)	90.00(3),90.00(3),90.00(3)
Reflections (#)	3246 (611)
Unique reflections (#)	1374 (262)
R obs	24.1 (36.7)
R meas	29.8 (45.3)
CC 1/2	91.7 (48.9)
Resolution (Å)	1
Completeness (%)	72.9 (74.4)
Total exposure (e ⁻ Å ⁻²)	~3
R	0.2045
wR2	0.4326
GooF	1.439

References

- 1. Y. Liu, S.-J. Han, W.-B. Liu, B. M. Stoltz, Catalytic Enantioselective Construction of Quaternary Stereocenters: Assembly of Key Building Blocks for the Synthesis of Biologically Active Molecules. *Accounts Chem. Res.* **48**, 740-751 (2015).
- 2. B. P. Pritchett, E. J. Donckele, B. M. Stoltz, Enantioselective Catalysis Coupled with Stereodivergent Cyclization Strategies Enables Rapid Syntheses of (+)-Limaspermidine and (+)-Kopsihainanine A. *Angew. Chem. Int. Ed.* 56, 12624-12627 (2017).