Recent Developments Toward Integrated Metabolomics Technologies (UHPLC-MS-SPE-NMR and MicroED) for Higher-Throughput Confident Metabolite Identifications

Rajarshi Ghosh[1,2,3,4], Guanhong Bu[5,6], Brent L. Nannenga[5,6] and Lloyd W. Sumner[1,2,3,4]*

[1]Division of Biochemistry, University of Missouri, Columbia, MO, United States, [2]MU Metabolomics Center, University of Missouri, Columbia, MO, United States, [3]Christopher S. Bond Life Sciences Center, University of Missouri, Columbia, SC, United States, [4]Interdisciplinary Plant Group, University of Missouri, Columbia, SC, United States, [5]Chemical Engineering, School for Engineering of Matter, Transport & Energy, Arizona State University, Tempe, AZ, United States, [6]Center for Applied Structural Discovery, The Biodesign Institute, Arizona State University, Tempe, AZ, United States

Metabolomics has emerged as a powerful discipline to study complex biological systems from a small molecule perspective. The success of metabolomics hinges upon reliable annotations of spectral features obtained from MS and/or NMR. In spite of tremendous progress with regards to analytical instrumentation and computational tools, < 20% of spectral features are confidently identified in most untargeted metabolomics experiments. This article explores the integration of multiple analytical instruments such as UHPLC-MS/MS-SPE-NMR and the cryo-EM method MicroED to achieve large-scale and confident metabolite identifications in a higher-throughput manner. UHPLC-MS/MS-SPE allows for the simultaneous automated purification of metabolites followed by offline structure elucidation and structure validation by NMR and MicroED. Large-scale study of complex metabolomes such as that of the model plant legume Medicago truncatula can be achieved using an integrated UHPLC-MS/MS-SPE-NMR metabolomics platform. Additionally, recent developments in MicroED to study structures of small organic molecules have enabled faster, easier and precise structure determinations of metabolites. A MicroED small molecule structure elucidation workflow (e.g., crystal screening, sample preparation, data collection and data processing/structure determination) has been described. Ongoing MicroED methods development and its future scope related to structure elucidation of specialized metabolites and metabolomics are highlighted. The incorporation of MicroED with a UHPLC-MS/MS-SPE-NMR instrumental ensemble offers the potential to accelerate and achieve higher rates of metabolite identification.

Keywords: integrated metabolomics, UHPLC-MS-SPE, NMR, MicroED, metabolite identification

INTRODUCTION

Metabolomics is the large-scale profiling of metabolites and it has now been applied to a multitude of biological systems. Metabolomics has seen exponential growth over the past 2 decades mostly due to continual technology advancements in analytical instrumentation and computational tools. The discipline continues to play a key role in understanding metabolism, the elucidation of novel gene functions and the discovery of biomarkers (Sumner et al., 2015; Beger et al., 2020; Nakabayashi and Saito, 2020). In spite of the overwhelming success of metabolomics, the large-scale confident identification of metabolites continues to be a major challenge. In most published metabolomics studies, typically 20% or less of the spectral features obtained from mass spectrometry (MS) or nuclear magnetic resonance (NMR) are confidently identified (Blaženović et al., 2018). The primary approach to metabolite identification includes spectral matching with data from authentic standards hosted in mass spectral databases (e.g., HMDB, MassBank, NIST, METLIN etc. (Stein and Scott, 1994; Smith et al., 2005; Horai et al., 2010; Wishart et al., 2018)) and NMR databases (e.g., BMRB, COLMAR etc. (Ulrich et al., 2008; Bingol et al., 2015)). This approach offers a confident and fast way to annotate compounds (Blaženović et al., 2018). Currently, chemical structure databases such as PubChem contain hundreds of millions of compounds but only 1-2 million of these compounds are of biological relevance and only a fraction of these compounds are available as authentic standards or included in the spectral databases. Hence an overwhelming number of metabolites are not represented in metabolomics spectral databases resulting in poor rates of metabolite identification. Additional *in silico* computational tools (e.g., MetFragCL, CFM-ID, MAGMa+, MS-FINDER etc. (Allen et al., 2014; Ruttkies et al., 2016; Tsugawa et al., 2016; Verdegem et al., 2016; Djoumbou-Feunang et al., 2019)) can exploit the content of structural databases for alternative metabolite identification capabilities (Blaženović et al., 2017; Wolfender et al., 2019) and *in silico* predicted spectral databases have been developed from the approximately 100 million known compounds in PubChem and ChemSpider (Milman and Zhurkovich, 2017). The use of molecular networking (e.g., GNPS (Wang et al., 2016)) to cluster compounds based on mass spectral fragmentation similarity followed by putative annotation has also gained popularity in recent years. Unfortunately, it is difficult to unambiguously identify compounds based solely on molecular networking data. Overall, the growth and expansion of authentic spectral databases continue to be slow and there is an urgent need to explore and advance other strategies to achieve higher-throughput confident metabolite identifications. This article discusses recent developments involving the integration of multiple analytical platforms such as ultra-high-performance liquid chromatography (UHPLC)-tandem mass spectrometry (MS/MS)-solid phase extraction (SPE) and nuclear magnetic resonance (NMR) to facilitate higher throughput empirical metabolite identifications. We also explore the potential scope and application of microcrystal electron diffraction (MicroED), a cryogenic electron microscopy (cryoEM) method, to the elucidation of small molecule structures. We believe that the incorporation of MicroED into the UHPLC-MS/MS-SPE workflow might lead to faster and higher-throughput identifications of biologically important unidentified metabolites. Development of these integrated metabolomics platforms can especially benefit the plant and microbial natural products community where large-scale identification of specialized metabolites is a major challenge due to the vast chemical diversity of known as well as truly novel compounds.

INTEGRATED ANALYTICAL INSTRUMENTS FOR CONFIDENT METABOLITE IDENTIFICATIONS

Most metabolomics studies over the last decade rely on either MS or NMR with MS being the most widely used technique (Letertre et al., 2021). MS coupled with liquid chromatography (LC) or gas chromatography (GC) has dominated the metabolomics arena due to its increased specificity, sensitivity and relative depth of coverage. Currently, LC-MS offers the greatest depth of coverage and dynamic range for metabolite analyses. Cumulatively, mass spectral database resources are substantially larger compared to NMR databases enabling significantly greater number of metabolite identifications in biological samples. However, NMR offers noninvasive, nondestructive and broad quantification of analytes using a single internal or external standard (Letertre et al., 2021). In addition, the historically lower relative sensitivity of NMR compared to MS has been somewhat mitigated with the advent of newer instruments with higher magnetic field strengths and NMR probes with cryogenically cooled receiver coils. For example, the Bruker 1.7 mm TCI MicroCryoProbe is 14-fold more sensitive than the conventional 5 mm room temperature probes. The increased sensitivity of modern cryoprobes currently enables the detection and identification of nanomole concentrations of metabolites (Molinski, 2010; Bhatia et al., 2019). Although both MS and NMR have their pros and cons, it is clear that none of the current analytical methods offers full coverage of the metabolome. Thus, the combined use of multiple analytical techniques and an integrated metabolomics platform are sensible options to maximize metabolome coverage and facilitate confident metabolite identifications based upon multiple orthogonal datatypes.

According to the Metabolomics Standards Initiative (MSI) guidelines, the confident identity of a metabolite is based upon a minimum of two orthogonal parameters such as retention time, accurate mass, MS/MS fragmentation pattern, collision cross section, NMR chemical shifts etc. relative to an authentic standard (Sumner et al., 2007; Schymanski et al., 2014). The integration of multiple analytical data not only increase the chances of identifying a metabolite, but also increases the level of confidence in the identification, thus potentially reducing false positives in the identification process. A compound identified by accurate mass, MS/MS and further confirmed by NMR has a much higher confidence level compared to a compound

putatively or tentatively identified by accurate mass and MS/MS. This strategy of utilizing multiple analytical data from complimentary instruments has been successfully used to confidently identify metabolites in several plants such as tomato, tea and barrel medic (van der Hooft et al., 2011; van der Hooft et al., 2012; Qiu et al., 2016).
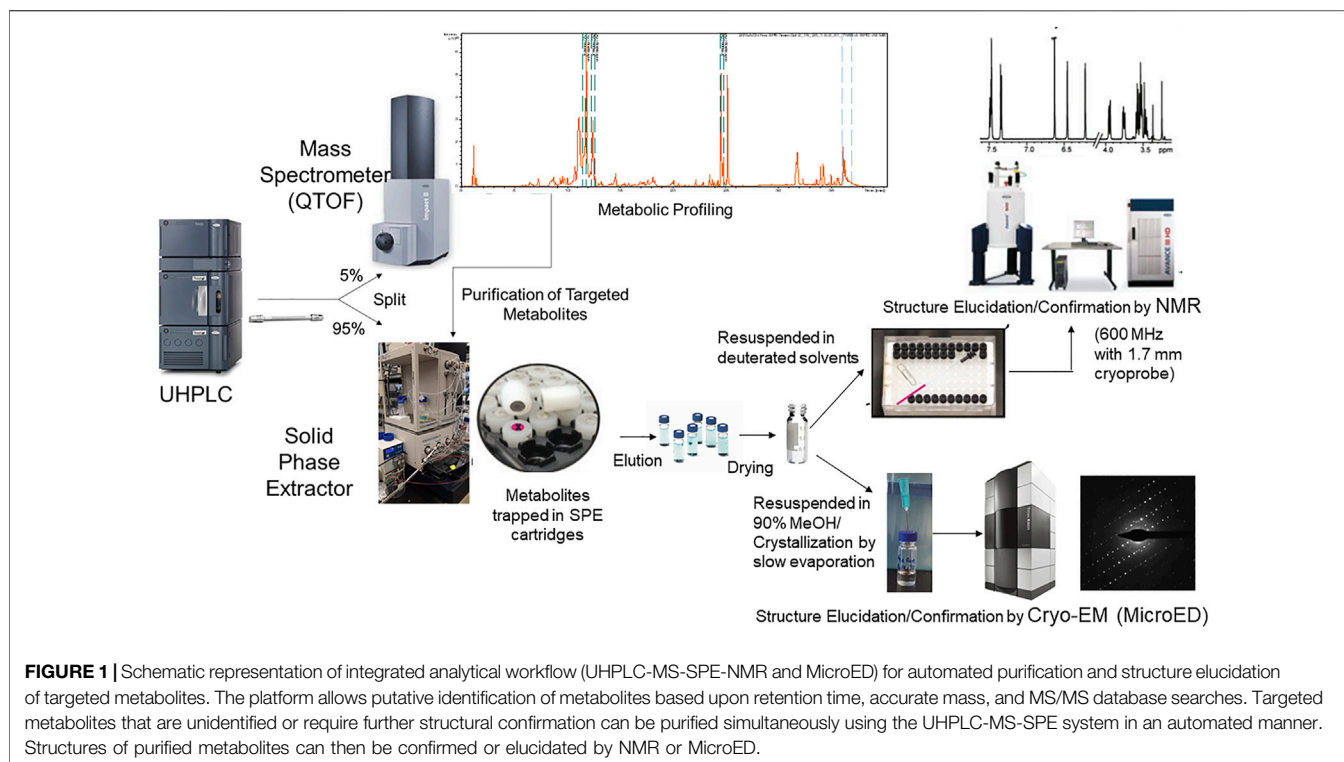
Historically, organic chemistry journals require purity, accurate mass/molecular formula and NMR connectivity evidence for the identification of organic molecules. Further, the natural products community utilizes high resolution MS and NMR together for structure elucidation purposes. The development of the hyphenated HPLC-UV-SPE-NMR (Clarkson et al., 2005), HPLC-MS/MS-SPE-NMR (Exarchou et al., 2003; van der Hooft et al., 2011) and UHPLC-MS/MS-SPE-NMR (Sumner et al., 2015) automated purification platforms coupled with structure elucidation/validation by NMR accelerated the traditionally lengthy and laborious natural product discovery process (Letertre et al., 2021). The integration of multiple instruments greatly benefited the metabolomics community as the platform allows automated purification of targeted metabolites from complex biological samples followed by confident identification using a combination of MS/MS, 1D and 2D NMR experiments. Simultaneous purification of multiple targeted metabolites can be achieved by splitting the UHPLC eluent partly towards a MS detector (5%) and partly towards a solid phase extraction (SPE) cartridge (95%) (Bhatia et al., 2019). Repeated injections of the sample lead to the purification and concentration of microgram amounts of desired metabolites within the SPE cartridges that are subsequently eluted using minimal amounts of deuterated solvents for NMR analyses (Bhatia et al., 2019). Important biological spectral features that remain unidentified in a conventional MS or NMR-based metabolomics database search can be identified by analyzing the accurate mass, MS/MS fragmentation patterns and the NMR spectra in a complementary fashion. Additionally, compounds identified through database search can be validated by NMR for a more confident identification. Less than 10 μg of SPE-purified metabolite is reported to be sufficient to generate 1H NMR spectrum suitable for structure validation (van der Hooft et al., 2011). An application of the UHPLC-MS/MS-SPE-NMR integrated platform was demonstrated in the plant metabolomics community by studying the metabolome of the model legume barrel medic (*Medicago truncatula*). Previously, nontargeted metabolic profiling of *M. truncatula* aerial and root tissues by either MS/MS or NMR had resulted in the tentative identification of approximately 40 flavonoids and 80 triterpenoid saponins (Huhman and Sumner, 2002; Kapusta et al., 2005a; Kapusta et al., 2005b; Huhman et al., 2005; Farag et al., 2007; Pollier et al., 2011; Qiu et al., 2016). UHPLC-MS/MS-SPE-NMR confirmed the identity of 88 compounds that are publicly available as part of the Bruker Sumner MetaboBASE® Plant Library (https://www.bruker.com/en/products-and-solutions/mass-spectrometry/ms-software/metabolomics-spectral-libraries.html). The confirmed identity of these compounds was aimed towards elucidation of several gene functions and characterization of triterpene saponin and flavonoid biosynthetic pathways. Overall,

UHPLC-MS/MS-SPE-NMR has been proven to be a powerful, cost-effective, and less labor-intensive platform to identify novel metabolites in the model legume.

Absolute structure elucidation or even structure confirmation by NMR can be a lengthy and complex process and its success is often dependent on the NMR interpretation skills of the researcher. Although, substantial progress has been made to simplify compound identification by NMR, it can still be a daunting task for non-experts. In the quest for simpler, faster and higher-throughput confident compound identifications in metabolomics, new technologies such as the cryo-EM method MicroED are being explored. MicroED was recently reported to enable precise stereochemical structure elucidation of milligram quantities of metabolites in a few hours (Jones et al., 2018) and we believe this technology can be pushed further to the structure elucidation of high nanogram to low microgram levels. The addition of MicroED into the existing UHPLC-MS/MS-SPE-NMR platform is exciting as it can potentially pave the way for faster and more straight forward identification of metabolites with less interpretative skills. A schematic representation of integrated analytical instruments (LC-MS-SPE-NMR and MicroED) for automated purification and structure elucidation of targeted metabolites is depicted in **Figure 1**.

MICROED IN SMALL MOLECULE STRUCTURE ANALYSES

In recent years, diffraction-based EM techniques such as MicroED have emerged as an effective tool to elucidate structures of frozen, beam-sensitive samples. MicroED utilizes a cryo-TEM (transmission electron microscope) to collect diffraction data from microcrystalline samples in order to determine their 3D structures (Nannenga and Gonen, 2019; Nannenga, 2020). As electrons interact strongly with matter, MicroED is highly successful for structure determination from crystals several orders of magnitude smaller than those used for single crystal X-ray diffraction. The MicroED method was initially developed for the determination of high-resolution protein structures (Shi et al., 2013; Nannenga BL. et al., 2014; Nannenga B. L. et al., 2014), later used for to the study of peptide structures as well (Rodriguez et al., 2015; Sawaya et al., 2016; Gallagher-Jones et al., 2018; Warmack et al., 2019), and in recent years, MicroED has been extended to the structural study of small organic molecules. The use of MicroED for the high-resolution structure determination of small molecules offers several advantages over other methods of structural analysis. Many structures of organic molecules and materials studied by MicroED have been determined directly from electron diffraction data collected from small amounts of synthesized material, thereby circumventing the need for large amounts of material and time-consuming recrystallization and optimization (Gruene et al., 2018; Jones et al., 2018; Banihashemi et al., 2020; Levine et al., 2020; Gleason et al., 2021). Additionally, MicroED is capable of detecting and solving the structure of several polymorphs or alternative compounds from a single sample preparation (Jones et al., 2018). Because secondary metabolites

FIGURE 1 | Schematic representation of integrated analytical workflow (UHPLC-MS-SPE-NMR and MicroED) for automated purification and structure elucidation of targeted metabolites. The platform allows putative identification of metabolites based upon retention time, accurate mass, and MS/MS database searches. Targeted metabolites that are unidentified or require further structural confirmation can be purified simultaneously using the UHPLC-MS-SPE system in an automated manner. Structures of purified metabolites can then be confirmed or elucidated by NMR or MicroED.
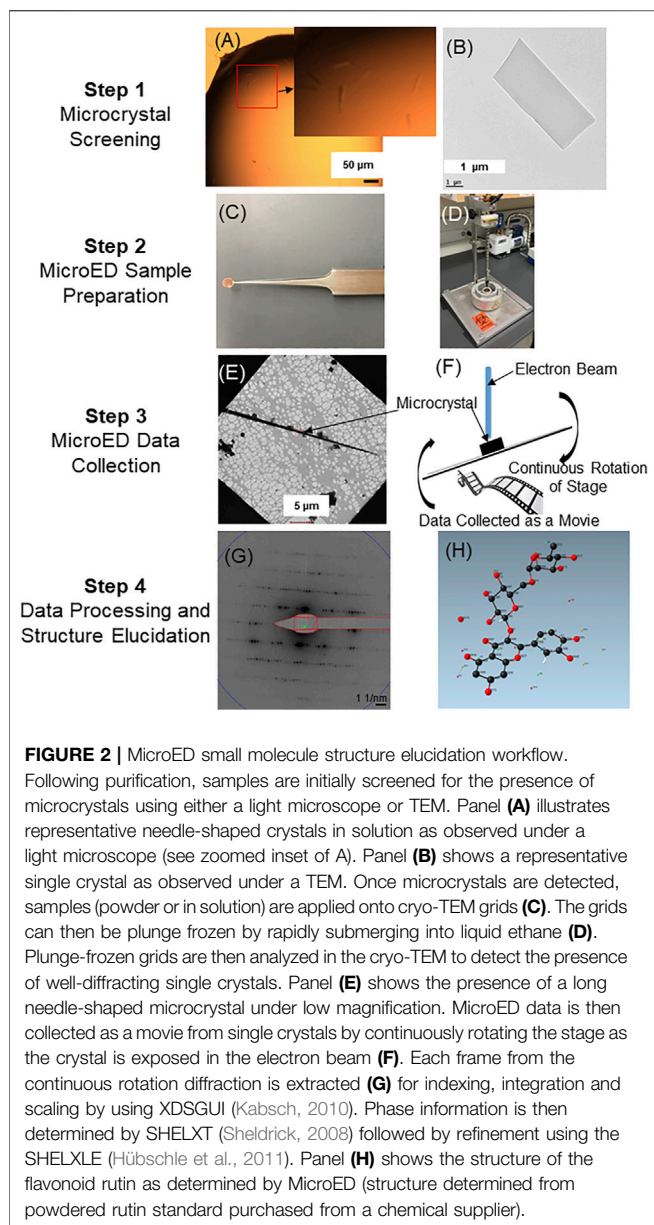
are often purified in low amounts from their native sources, it can be very difficult to grow large crystals suitable for X-ray diffraction experiments. Therefore, the use of MicroED for the structural studies of natural products is very attractive (Danelius et al., 2021). Previous work has demonstrated the ability of MicroED to determine atomic resolution structures of specialized metabolites, including biotin, niacin, brucine, and thiostrepton, directly from powders obtained following purification or procurement from chemical suppliers (van Genderen et al., 2016; Jones et al., 2018; Zhou et al., 2019). Additionally, MicroED is likely to solve the structures from powder mixtures on the same grid, which was demonstrated using mixtures of biotin, carbamazepine, cinchonine, and brucine powders (Jones et al., 2018). These initial studies indicate that MicroED can be a powerful tool in small molecule structure elucidation and can potentially accelerate compound identification in the natural products and metabolomics community.

MICROED STRUCTURE ANALYSIS WORKFLOW

The MicroED sample preparation, data collection, and data processing pipeline for proteins as well as small molecules has been described previously (Shi et al., 2016; Bu and Nannenga, 2021; de la Cruz, 2021; Martynowycz and Gonen, 2021; Zee et al., 2021). A MicroED structure elucidation workflow for natural products and specialized targeted metabolites has been optimized from previously published methods (**Figure 2**). Samples can be

initially screened for the presence of microcrystals using either an optical light microscope or a TEM. Once the presence of microcrystals in the samples are confirmed, they are deposited onto TEM grids by adding powdered samples directly to the grid, creating a crystalline suspension and pipetting the suspension on to the grid followed by blotting or drying, or by solubilizing the sample in a compatible solvent and adding this on the grid followed by drying of the solvent and sample crystallization directly on the grid. Following the deposition of the sample on the grid, the grids are then either frozen in liquid ethane, liquid nitrogen, or loaded into the cryo-TEM at room temperature and allowed to freeze within the cryo-TEM. The grid is then searched at low magnification to identify the presence of crystals on the grid. When a promising crystal has been identified, the crystal is brought to the center of the field of view, which is important as the diffraction mode of the microscope should be aligned such that the beam used for exposure is sampling the same area. After the crystal has been aligned, an initial diffraction pattern is collected to assess the diffracting power of the crystal. If the crystal shows high-quality diffraction (e.g. high-resolution, sharp spots) then a complete dataset is collected. To collect a MicroED dataset, the crystal is first tilted to the maximal angle where no other crystals or grid bars enter the path of the beam. The data collection is then performed by continuously rotating the crystal in the electron beam as the high speed detector records the diffraction data (Nannenga B. L. et al., 2014). It is critical that the stage be at eucentric height so that the crystal does not rotate out of the beam as the stage rotates during data collection. The resulting diffraction movies containing frames of continuous rotation diffraction patterns are converted to Super Marty View (SMV) format to extract each frame (Hattne et al., 2015). The frames are indexed, integrated and scaled using standard

FIGURE 2 | MicroED small molecule structure elucidation workflow. Following purification, samples are initially screened for the presence of microcrystals using either a light microscope or TEM. Panel **(A)** illustrates representative needle-shaped crystals in solution as observed under a light microscope (see zoomed inset of A). Panel **(B)** shows a representative single crystal as observed under a TEM. Once microcrystals are detected, samples (powder or in solution) are applied onto cryo-TEM grids **(C)**. The grids can then be plunge frozen by rapidly submerging into liquid ethane **(D)**. Plunge-frozen grids are then analyzed in the cryo-TEM to detect the presence of well-diffracting single crystals. Panel **(E)** shows the presence of a long needle-shaped microcrystal under low magnification. MicroED data is then collected as a movie from single crystals by continuously rotating the stage as the crystal is exposed in the electron beam **(F)**. Each frame from the continuous rotation diffraction is extracted **(G)** for indexing, integration and scaling by using XDSGUI (Kabsch, 2010). Phase information is then determined by SHELXT (Sheldrick, 2008) followed by refinement using the SHELXLE (Hübschle et al., 2011). Panel **(H)** shows the structure of the flavonoid rutin as determined by MicroED (structure determined from powdered rutin standard purchased from a chemical supplier).

diffraction processing programs developed for X-ray crystallography (e.g. XDS (Kabsch, 2010), DIALS (Clabbers et al., 2018)). Datasets, typically consisting of data from several crystals merged together are then phased by direct methods followed by structure refinement with electron scattering factors, using programs commonly used for small molecule structure determination (e.g., SHELX (Sheldrick, 2008)). The structure of the flavonoid rutin as solved by the MicroED structure elucidation workflow is shown in **Figure 2**.

INTEGRATION OF UHPLC-MS-SPE AND MICROED

The incorporation of MicroED into the UHPLC-MS-SPE workflow was envisioned to potentially achieve faster, higher throughput and

confident metabolite identifications based upon multiple orthogonal data. Similar to UHLC-MS-SPE-NMR, several biologically relevant unidentified or putatively identified metabolites can be targeted and purified simultaneously and further analyzed by MicroED. Preliminary experiments suggest that it is possible to generate sufficient microcrystals of multiple purified flavonoids and triterpenoid saponins from 20 UHPLC-MS-SPE injections of a plant extract (10 mg lyophilized powdered sample extracted with 1 ml 80% methanol). Structure analysis by MicroED requires substantially lesser number of UHPLC-MS-SPE injections compared to NMR where 40–60 plant extract injections are often required to achieve adequate purified material for structure validation by 1D NMR and absolute structure elucidation by 2D NMR. A previous study by Jones et al., 2018 (Jones et al., 2018) reported a 50% success rate of obtaining high quality diffraction data from flash column purified small molecules without additional crystallization. Our initial experiments suggest that slow evaporation of SPE-purified metabolite solutions in 90% methanol at 4°C for 7–10 days has an 80% success rate of producing well-diffracting microcrystals. Ten putatively identified flavonoids and triterpenoid saponins were used to evaluate the crystallization efficiency of the slow evaporation method. Microcrsytals and the corresponding diffraction pattern of SPE-purified 6-malonyl ononin is shown in **Supplementary Figure S1**. Due to the minute amounts (nanogram-low microgram) of purified metabolites and the resultant low crystal density in our study, crystal screening and identification of high-quality single crystals were often more time-consuming. Automated cryo-EM crystal screening software can potentially accelerate this process in future. It must be noted that the success of MicroED also depends on the quality of microcrystals and further optimization of crystallization conditions is crucial. In our initial studies, we have encountered multiple crystals that are sometimes stacked together leading to low resolution diffraction spots unsuitable for structure determination. Larger and improperly shaped crystals may need to be further broken down by vortexing, sonicating or pipetting. Other alternative and more sophisticated methods for preparation of microcrystals include the use of focused ion beam (FIB) milling under cryogenic conditions (Nannenga and Gonen, 2019). Once suitable microcrystals producing high quality diffraction patterns are found, MicroED data can be acquired in minutes and compounds can be identified in less than an hour. Presence of minor impurities from co-eluting compounds in the crystalline solution do not seem to interfere greatly with structure elucidation of the targeted metabolites. However, improper sample handling can lead to sample degradation and elucidation of only partial structures. This was evident in case of malonylated isoflavonoids where degradation of the highly labile malonyl group was noticed. Overall, in its current form the authors believe that MicroED is best utilized as an orthogonal structure elucidation tool in combination with other orthogonal datasets for higher confident structure determinations. MicroED is not perceived to be a totally independent structural elucidation tool as it still faces challenges in determining bond lengths accurately enough to resolve single from double bonds or alcohols from ketones. As the method continues to develop, it has the potential to be used for absolute structure elucidation of unknown compounds with precise stereochemistry.

For MicroED, there are several areas of active methods development that will continue to push the technique forward. Areas of method development relevant to small molecule structure determination include enhanced data collection procedures, the modeling of charge and chemical bonding within the crystal structures (Yonekura and Maki-Yonekura, 2016; Yonekura et al., 2018), and procedures for using electron diffraction to determine the absolute stereochemistry of a molecule (Brázda et al., 2019). Also, sample preparation is critically important for MicroED and represents an important area of continued development. It is yet to be determined if a simple slow evaporation crystallization strategy is suitable to generate microcrystals for different classes of specialized natural products. Our ongoing efforts are focused on improving the UHPLC-MS-SPE-MicroED workflow for faster, easier and confident metabolite identifications. Several flavonoids, sapogenins and triterpenoid saponins have been SPE-purified and currently being analyzed by MicroED.

CONCLUSION

Integrated analytical platforms such as UHPLC-MS-SPE coupled with NMR and MicroED offer alternative and effective approaches toward identification of metabolites that often remain unidentified in conventional metabolomics database search workflows. Introduction of MicroED into the LC-MS-SPE purification and structure elucidation pipeline is especially exciting as it opens up new avenues for faster, easier and higher-throughput identification of metabolites. The most attractive aspect of MicroED is its wide accessibility as electron microscopes and detectors needed for it are available in most modern EM cores. However, it must be noted that wide-spread application of MicroED in small molecule research is relatively recent and further development of automated software packages for crystal screening, data collection and structure determination are still needed for routine analyses. As MicroED becomes more targeted towards small molecule research, we expect to see increased applications of the method in metabolomics workflows.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

LS conceptualized the integrated UHPLC-MS-SPE-NMR/MicroED approaches. RG and GB conducted the experiments. Brent Nannenga supervised the MicroED experiments. RG drafted the initial manuscript. All authors provided editorial feedback, revisions and approved the manuscript.

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2021.720955/full#supplementary-material

REFERENCES

Allen, F., Pon, A., Wilson, M., Greiner, R., and Wishart, D. (2014). CFM-ID: a Web Server for Annotation, Spectrum Prediction and Metabolite Identification from Tandem Mass Spectra. *Nucleic Acids Res.* 42, W94–W99. doi:10.1093/nar/gku436

Banihashemi, F., Bu, G., Thaker, A., Williams, D., Lin, J. Y. S., and Nannenga, B. L. (2020). Beam-sensitive Metal-Organic Framework Structure Determination by Microcrystal Electron Diffraction. *Ultramicroscopy* 216, 113048. doi:10.1016/j.ultramic.2020.113048

Beger, R. D., Schmidt, M. A., and Kaddurah-Daouk, R. (2020). Current Concepts in Pharmacometabolomics, Biomarker Discovery, and Precision Medicine. *Metabolites* 10, 129. doi:10.3390/metabo10040129

Bhatia, A., Sarma, S. J., Lei, Z., and Sumner, L. W. (2019). UHPLC-QTOF-MS/MS-SPE-NMR: a Solution to the Metabolomics Grand challenge of Higher-Throughput, Confident Metabolite Identifications. *NMR-Based Metabolomics*. 2037, 113–133. doi:10.1007/978-1-4939-9690-2_7

Bingol, K., Li, D.-W., Bruschweiler-Li, L., Cabrera, O. A., Megraw, T., Zhang, F., et al. (2015). Unified and Isomer-specific NMR Metabolomics Database for the Accurate Analysis of 13C-1H HSQC Spectra. *ACS Chem. Biol.* 10, 452–459. doi:10.1021/cb5006382

Blaženović, I., Kind, T., Ji, J., and Fiehn, O. (2018). Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. *Metabolites* 8, 31. doi:10.3390/metabo8020031

Blaženović, I., Kind, T., Torbašinović, H., Obrenović, S., Mehta, S. S., Tsugawa, H., et al. (2017). Comprehensive Comparison of In Silico MS/MS Fragmentation Tools of the CASMI Contest: Database Boosting Is Needed to Achieve 93% Accuracy. *J. cheminformatics* 9, 1–12. doi:10.1186/s13321-017-0219-x

Brázda, P., Palatinus, L., and Babor, M. (2019). Electron Diffraction Determines Molecular Absolute Configuration in a Pharmaceutical Nanocrystal. *Science* 364, 667–669. doi:10.1126/science.aaw2560

Bu, G., and Nannenga, B. L. (2021). MicroED and Data Collection for Protein. *Methods Mol. Biol.* 2215, 287–297. doi:10.1007/978-1-0716-0966-8_13

Clabbers, M. T. B., Gruene, T., Parkhurst, J. M., Abrahams, J. P., and Waterman, D. G. (2018). Electron Diffraction Data Processing with DIALS. *Acta Cryst. Sect D Struct. Biol.* 74, 506–518. doi:10.1107/s2059798318007726

Clarkson, C., Dan Stærk, D., Hansen, S. H., and Jaroszewski, J. W. (2005). Hyphenation of Solid-phase Extraction with Liquid Chromatography and

Nuclear Magnetic Resonance: Application of HPLC-DAD-SPE-NMR to Identification of Constituents of Kanahia Laniflora. *Anal. Chem.* 77, 3547–3553. doi:10.1021/ac050212k

Danelius, E., Halaby, S., Van Der Donk, W. A., and Gonen, T. (2021). MicroED in Natural Product and Small Molecule research. *Natural Product Reports* 38 (3), 423–431. doi:10.1039/d0np00035c

De La Cruz, M. J. (2021). "Automation of Continuous-Rotation Data Collection for," in *cryoEM: Methods and Protocols*. Editors T. Gonen and B. L. Nannenga (New York, NY: Springer US), 321–327. doi:10.1007/978-1-0716-0966-8_16

Djoumbou-Feunang, Y., Pon, A., Karu, N., Zheng, J., Li, C., Arndt, D., et al. (2019). CFM-ID 3.0: Significantly Improved ESI-MS/MS Prediction and Compound Identification. *Metabolites* 9, 23. doi:10.3390/metabo9040072

Exarchou, V., Godejohann, M., Van Beek, T. A., Gerothanassis, I. P., and Vervoort, J. (2003). LC-UV-solid-phase Extraction-NMR-MS Combined with a Cryogenic Flow Probe and its Application to the Identification of Compounds Present in Greek Oregano. *Anal. Chem.* 75, 6288–6294. doi:10.1021/ac0347819

Farag, M. A., Huhman, D. V., Lei, Z., and Sumner, L. W. (2007). Metabolic Profiling and Systematic Identification of Flavonoids and Isoflavonoids in Roots and Cell Suspension Cultures of Medicago Truncatula Using HPLC-UV-ESI-MS and GC-MS. *Phytochemistry* 68, 342–354. doi:10.1016/j.phytochem.2006.10.023

Gallagher-Jones, M., Glynn, C., Boyer, D. R., Martynowycz, M. W., Hernandez, E., Miao, J., et al. (2018). Sub-ångström Cryo-EM Structure of a Prion Protofibril Reveals a Polar Clasp. *Nat. Struct. Mol. Biol.* 25, 131–134. doi:10.1038/s41594-017-0018-0

Gleason, P. R., Nannenga, B. L., and Mills, J. H. (2020). Rapid Structural Analysis of a Synthetic Non-canonical Amino Acid by Microcrystal Electron Diffraction. *Front. Mol. Biosci.* 7, 609999. doi:10.3389/fmolb.2020.609999

Gruene, T., Wennmacher, J. T. C., Zaubitzer, C., Holstein, J. J., Heidler, J., Fecteau-Lefebvre, A., et al. (2018). Rapid Structure Determination of Microcrystalline Molecular Compounds Using Electron Diffraction. *Angew. Chem. Int. Ed.* 57, 16313–16317. doi:10.1002/anie.201811318

Hattne, J., Reyes, F. E., Nannenga, B. L., Shi, D., De La Cruz, M. J., Leslie, A. G. W., et al. (2015). MicroED Data Collection and Processing. *Acta Cryst. Sect A.* 71, 353–360. doi:10.1107/s2053273315010669

Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., et al. (2010). MassBank: a Public Repository for Sharing Mass Spectral Data for Life Sciences. *J. Mass. Spectrom.* 45, 703–714. doi:10.1002/jms.1777

Hübschle, C. B., Sheldrick, G. M., and Dittrich, B. (2011). ShelXle: a Qt Graphical User Interface forSHELXL. *J. Appl. Cryst.* 44, 1281–1284. doi:10.1107/s0021889811043202

Huhman, D. V., Berhow, M. A., and Sumner, L. W. (2005). Quantification of Saponins in Aerial and Subterranean Tissues ofMedicago Truncatula. *J. Agric. Food Chem.* 53, 1914–1920. doi:10.1021/jf0482663

Huhman, D. V., and Sumner, L. W. (2002). Metabolic Profiling of Saponins in Medicago Sativa and Medicago Truncatula Using HPLC Coupled to an Electrospray Ion-Trap Mass Spectrometer. *Phytochemistry* 59, 347–360. doi:10.1016/s0031-9422(01)00432-0

Jones, C. G., Martynowycz, M. W., Hattne, J., Fulton, T. J., Stoltz, B. M., Rodriguez, J. A., et al. (2018). The CryoEM Method MicroED as a Powerful Tool for Small Molecule Structure Determination. *ACS Cent. Sci.* 4, 1587–1592. doi:10.1021/acscentsci.8b00760

Kabsch, W. (2010). Xds. *Acta Crystallogr. D Biol. Cryst.* 66, 125–132. doi:10.1107/s0907444909047337

Kapusta, I., Janda, B., Stochmal, A., and Oleszek, W. (2005a). Determination of Saponins in Aerial Parts of Barrel Medic (Medicago Truncatula) by Liquid Chromatography−Electrospray Ionization/Mass Spectrometry. *J. Agric. Food Chem.* 53, 7654–7660. doi:10.1021/jf051256x

Kapusta, I., Stochmal, A., Perrone, A., Piacente, S., Pizza, C., and Oleszek, W. (2005b). Triterpene Saponins from Barrel Medic (Medicago Truncatula) Aerial Parts. *J. Agric. Food Chem.* 53, 2164–2170. doi:10.1021/jf048178i

Letertre, M. P. M., Dervilly, G., and Giraudeau, P. (2021). Combined Nuclear Magnetic Resonance Spectroscopy and Mass Spectrometry Approaches for Metabolomics. *Anal. Chem.* 93, 500–518. doi:10.1021/acs.analchem.0c04371

Levine, A. M., Bu, G., Biswas, S., Tsai, E. H. R., Braunschweig, A. B., and Nannenga, B. L. (2020). Crystal Structure and Orientation of Organic Semiconductor Thin Films by Microcrystal Electron Diffraction and Grazing-Incidence Wide-Angle X-ray Scattering. *Chem. Commun.* 56, 4204–4207. doi:10.1039/d0cc00119h

Martynowycz, M. W., and Gonen, T. (2021). *Microcrystal Electron Diffraction of Small Molecules*. New York City, US: JoVE, e62313. doi:10.3791/62313

Milman, B. L., and Zhurkovich, I. K. (2017). The Chemical Space for Non-target Analysis. *Trac Trends Anal. Chem.* 97, 179–187. doi:10.1016/j.trac.2017.09.013

Molinski, T. F. (2010). NMR of Natural Products at the 'nanomole-Scale'. *Nat. Prod. Rep.* 27, 321–329. doi:10.1039/b920545b

Nakabayashi, R., and Saito, K. (2020). Higher Dimensional Metabolomics Using Stable Isotope Labeling for Identifying the Missing Specialized Metabolism in Plants. *Curr. Opin. Plant Biol.* 55, 84–92. doi:10.1016/j.pbi.2020.02.009

Nannenga, B. L., Shi, D., Hattne, J., Reyes, F. E., and Gonen, T. (2014a). Structure of Catalase Determined by MicroED. *Elife* 3, e03600. doi:10.7554/eLife.03600

Nannenga, B. L., and Gonen, T. (2019). The Cryo-EM Method Microcrystal Electron Diffraction (MicroED). *Nat. Methods* 16, 369–379. doi:10.1038/s41592-019-0395-x

Nannenga, B. L. (2020). MicroED Methodology and Development. *Struct. Dyn.* 7, 014304. doi:10.1063/1.5128226

Nannenga, B. L., Shi, D., Leslie, A. G. W., and Gonen, T. (2014b). High-resolution Structure Determination by Continuous-Rotation Data Collection in MicroED. *Nat. Methods* 11, 927–930. doi:10.1038/nmeth.3043

Pollier, J., Morreel, K., Geelen, D., and Goossens, A. (2011). Metabolite Profiling of Triterpene Saponins in Medicago Truncatula Hairy Roots by Liquid Chromatography Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *J. Nat. Prod.* 74, 1462–1476. doi:10.1021/np200218r

Qiu, F., Fine, D. D., Wherritt, D. J., Lei, Z., and Sumner, L. W. (2016). PlantMAT: A Metabolomics Tool for Predicting the Specialized Metabolic Potential of a System and for Large-Scale Metabolite Identifications. *Anal. Chem.* 88, 11373–11383. doi:10.1021/acs.analchem.6b00906

Rodriguez, J. A., Ivanova, M. I., Sawaya, M. R., Cascio, D., Reyes, F. E., Shi, D., et al. (2015). Structure of the Toxic Core of Alpha-Synuclein from Invisible Crystals. *Nature.* 525, 486. doi:10.1038/nature15368

Ruttkies, C., Schymanski, E. L., Wolf, S., Hollender, J., and Neumann, S. (2016). MetFrag Relaunched: Incorporating Strategies beyond In Silico Fragmentation. *J. Cheminform* 8, 3. doi:10.1186/s13321-016-0115-9

Sawaya, M. R., Rodriguez, J., Cascio, D., Collazo, M. J., Shi, D., Reyes, F. E., et al. (2016). Ab Initio structure Determination from Prion Nanocrystals at Atomic Resolution by MicroED. *Proc. Natl. Acad. Sci. U S A.* doi:10.1073/pnas.1606287113

Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., et al. (2014). Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* 48, 2097–2098. doi:10.1021/es5002105

Sheldrick, G. M. (2008). A Short History ofSHELX. *Acta Cryst. Sect A.* 64, 112–122. doi:10.1107/s0108767307043930

Shi, D., Nannenga, B. L., Iadanza, M. G., and Gonen, T. (2013). Three-dimensional Electron Crystallography of Protein Microcrystals. *Elife* 2, e01345. doi:10.7554/eLife.01345

Shi, D., Nannenga, B. L., De La Cruz, M. J., Liu, J., Sawtelle, S., Calero, G., et al. (2016). The Collection of MicroED Data for Macromolecular Crystallography. *Nat. Protoc.* 11, 895–904. doi:10.1038/nprot.2016.046

Smith, C. A., Maille, G. O., Want, E. J., Qin, C., Trauger, S. A., Brandon, T. R., et al. (2005). Metlin. *Ther. Drug Monit.* 27, 747–751. doi:10.1097/01.ftd.0000179845.53213.39

Stein, S. E., and Scott, D. R. (1994). Optimization and Testing of Mass Spectral Library Search Algorithms for Compound Identification. *J. Am. Soc. Mass. Spectrom.* 5, 859–866. doi:10.1016/1044-0305(94)87009-8

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., et al. (2007). Proposed Minimum Reporting Standards for Chemical Analysis. *Metabolomics* 3, 211–221. doi:10.1007/s11306-007-0082-2

Sumner, L. W., Lei, Z., Nikolau, B. J., and Saito, K. (2015). Modern Plant Metabolomics: Advanced Natural Product Gene Discoveries, Improved Technologies, and Future Prospects. *Nat. Prod. Rep.* 32, 212–229. doi:10.1039/c4np00072b

Tsugawa, H., Kind, T., Nakabayashi, R., Yukihira, D., Tanaka, W., Cajka, T., et al. (2016). Hydrogen Rearrangement Rules: Computational MS/MS

Fragmentation and Structure Elucidation Using MS-FINDER Software. *Anal. Chem.* 88, 7946–7958. doi:10.1021/acs.analchem.6b00770

Ulrich, E. L., Akutsu, H., Doreleijers, J. F., Harano, Y., Ioannidis, Y. E., Lin, J., et al. (2008). BioMagResBank. *Nucleic Acids Res.* 36, D402–D408. doi:10.1093/nar/gkm957

Van Der Hooft, J. J. J., Akermi, M., Ünlü, F. Y., Mihaleva, V., Roldan, V. G., Bino, R. J., et al. (2012). Structural Annotation and Elucidation of Conjugated Phenolic Compounds in Black, Green, and White Tea Extracts. *J. Agric. Food Chem.* 60, 8841–8850. doi:10.1021/jf300297y

Van Der Hooft, J. J. J., Mihaleva, V., De Vos, R. C. H., Bino, R. J., and Vervoort, J. (2011). A Strategy for Fast Structural Elucidation of Metabolites in Small Volume Plant Extracts Using Automated MS-guided LC-MS-SPE-NMR. *Magn. Reson. Chem.* 49, S55–S60. doi:10.1002/mrc.2833

Van Genderen, E., Clabbers, M. T. B., Das, P. P., Stewart, A., Nederlof, I., Barentsen, K. C., et al. (2016). Ab Initiostructure Determination of Nanocrystals of Organic Pharmaceutical Compounds by Electron Diffraction at Room Temperature Using a Timepix Quantum Area Direct Electron Detector. *Acta Cryst. Sect A.* 72, 236–242. doi:10.1107/s2053273315022500

Verdegem, D., Lambrechts, D., Carmeliet, P., and Ghesquiere, B. (2016). Improved Metabolite Identification with MIDAS and MAGMa through MS/MS Spectral Dataset-Driven Parameter Optimization. *Metabolomics* 12, 16. doi:10.1007/s11306-016-1036-3

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., et al. (2016). Sharing and Community Curation of Mass Spectrometry Data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* 34, 828–837. doi:10.1038/nbt.3597

Warmack, R. A., Boyer, D. R., Zee, C.-T., Richards, L. S., Sawaya, M. R., Cascio, D., et al. (2019). Structure of Amyloid-β (20-34) with Alzheimer's-Associated Isomerization at Asp23 Reveals a Distinct Protofilament Interface. *Nat. Commun.* 10, 3357. doi:10.1038/s41467-019-11183-z

Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vázquez-Fresno, R., et al. (2018). HMDB 4.0: the Human Metabolome Database for 2018. *Nucleic Acids Res.* 46, D608–D617. doi:10.1093/nar/gkx1089

Wolfender, J.-L., Nuzillard, J.-M., Van Der Hooft, J. J. J., Renault, J.-H., and Bertrand, S. (2019). Accelerating Metabolite Identification in Natural Product Research: Toward an Ideal Combination of Liquid Chromatography-High-Resolution Tandem Mass Spectrometry and NMR Profiling,in SilicoDatabases, and Chemometrics. *Anal. Chem.* 91, 704–742. doi:10.1021/acs.analchem.8b05112

Yonekura, K., and Maki-Yonekura, S. (2016). Refinement of Cryo-EM Structures Using Scattering Factors of Charged Atoms. *J. Appl. Crystallogr.* 49, 1517. doi:10.1107/S1600576716011274

Yonekura, K., Matsuoka, R., Yamashita, Y., Yamane, T., Ikeguchi, M., Kidera, A., et al. (2018). Ionic Scattering Factors of Atoms that Compose Biological Molecules. *Int. Union Crystallogr. J.* 5, 348–353. doi:10.1107/s2052252518005237

Zee, C.-T., Saha, A., Sawaya, M. R., and Rodriguez, J. A. (2021). "Ab Initio Determination of Structures by MicroED," in *cryoEM: Methods and Protocols.* Editors T. Gonen and B. L. Nannenga (New York, NY: Springer US), 329–348. doi:10.1007/978-1-0716-0966-8_17

Zhou, H., Luo, F., Luo, Z., Li, D., Liu, C., and Li, X. (2019). Programming Conventional Electron Microscopes for Solving Ultrahigh-Resolution Structures of Small and Macro-Molecules. *Anal. Chem.* 91, 10996–11003. doi:10.1021/acs.analchem.9b01162