Precise. Reliable.

Sleek and Squeaky Clean.

Ensure excellence with PHCbi brand products.

PHCbi

Life Science Innovator Since 1966

Check for updates

Received: 22 August 2020 | Revised: 2 December 2020 | Accepted: 13 December 2020

DOI: 10.1002/bit.27661

ARTICLE

BIOTECHNOLOGY AND BIOENGINEERING WILEY

Creation and filtering of a recurrent spectral library of CHO cell metabolites and media components

Kelly H. Telu[1] | Ramesh Marupaka[1] | Nirina R. Andriamaharavo[1] | Yamil Simón-Manso[1] | Yuxue Liang[1] | Yuri A. Mirokhin[1] | Tallat H. Bukhari[1] | Renae J. Preston[2] | Lila Kashi[2] | Zvi Kelman[2] | Stephen E. Stein[1]

[1]Mass Spectrometry Data Center, Biomolecular Measurement Division, National Institute of Standards and Technology, Gaithersburg, Maryland, USA

[2]Biomolecular Labeling Laboratory, Institute for Bioscience and Biotechnology Research, National Institute of Standards and Technology and the University of Maryland, Rockville, Maryland, USA

Correspondence
Kelly H. Telu, Mass Spectrometry Data Center, Biomolecular Measurement Division, National Institute of Standards and Technology, Gaithersburg, MD 20899, USA.
Email: kelly.telu@nist.gov

Abstract

This paper reports the first implementation of a new type of mass spectral library for the analysis of Chinese hamster ovary (CHO) cell metabolites that allows users to quickly identify most compounds in any complex metabolite sample. We also describe an annotation methodology developed to filter out artifacts and low-quality spectra from recurrent unidentified spectra of metabolites. CHO cells are commonly used to produce biological therapeutics. Metabolic profiles of CHO cells and media can be used to monitor process variability and look for markers that discriminate between batches of product. We have created a comprehensive library of both identified and unidentified metabolites derived from CHO cells that can be used in conjunction with tandem mass spectrometry to identify metabolites. In addition, we present a workflow that can be used for assigning confidence to a NIST MS/MS Library search match based on prior probability of general utility. The goal of our work is to annotate and identify (when possible), all liquid chromatography-mass spectrometry generated metabolite ions as well as create automatable library building and identification pipelines for use by others in the field.

KEYWORDS

Chinese hamster ovary cells, global metabolite profiling, liquid chromatography-tandem mass spectrometry, nontargeted metabolomics, recurrent unidentified spectra

1 | INTRODUCTION

Chinese hamster ovary (CHO) cells are the predominant host cells for monoclonal antibody (mAb) production (Kunert & Reinhart, 2016). Metabolomics provides information on cellular phenotypes. Several metabolites have been demonstrated to be biomarkers of CHO cell status (Mohmad-Saberi et al., 2013). Metabolomic analysis of CHO cells has primarily been used in process or media/feed development and has predominantly focused on targeted metabolite analysis of major metabolites, although there are several studies that utilized global metabolite analysis (Stolfa et al., 2018). A comprehensive assessment of CHO cell metabolic profiles could lead to improvements in product yield and quality by providing further understanding of the CHO cell metabolome (Stolfa et al., 2018). Mass spectral libraries have been extremely popular

Ramesh Marupaka, Nirina R. Andriamaharavo, and Yuri A. Mirokhin are NIST Associates.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2021 The Authors. *Biotechnology and Bioengineering* Published by Wiley Periodicals LLC. This article has been contributed to by US Government employees and their work is in the public domain in the USA

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

for more than 40 years for identifying volatile chemical compounds using gas chromatography-mass spectrometry (GC-MS). They are used to locate the most similar spectra in the reference library and present the compounds that generated them in a "hit list" sorted by their similarity to the acquired spectrum (S. Stein, 2012). Liquid chromatography-mass spectrometry (LC-MS) is a widely practiced method for identifying the chemical components in metabolomics (Gowda & Djukovic, 2014). For confident metabolite identifications, liquid chromatography-tandem mass spectrometry (LC-MS/MS) can be performed and the fragmentation pattern can be compared to a MS/MS spectral library. Commercial MS/MS libraries that contain curated spectra (the NIST Tandem [MS/MS] Mass Spectral Library and the Wiley MSforID Library) as well as free libraries that facilitate data sharing (MassBank, MassBank of North America [MoNA], LipidBlast, METLIN, mzCloud, GNPS, etc.) are available and have been reviewed recently (Kind et al., 2018). These libraries contain experimental spectra of known compounds and spectra of unidentified compounds are not documented there. Other libraries such as LipidBlast, Greazy/LipidLama, CFM-ID, and so forth are based on *in silico* prediction of the spectra of known or predicted metabolites (Kind et al., 2018).

A comprehensive library of both known and unidentified CHO cell metabolites will be beneficial to the field of CHO cell metabolite analysis. In addition to producing the NIST MS/MS Library, the NIST Mass Spectrometry Data Center (MSDC) has recently begun creating material-oriented libraries that are generated from the analysis of complex mixtures such as human plasma and urine (https://chemdata.nist.gov/dokuwiki/doku.php?id = chemdata:arus) to address the issue of unknown metabolites (metabolites not identified by library searching), identify cross-platform metabolite signatures, and catalogue all spectra associated with a particular material of interest (Mallard et al., 2014; Remoroza et al., 2018; Simon-Manso et al., 2013; Simon-Manso et al., 2019; S. Stein, 2012; Telu et al., 2016). These material-oriented libraries contain recurrent spectra (spectra that occur repeatedly in the sample) for all detectable metabolites, both known and unknown that are processed to produce high-quality consensus spectra for the library. The MSDC has also created spectral libraries (Dong et al., 2018; Dong, Yan, Liang, & Stein, 2016) of the NISTmAb, a humanized IgG1κ Monoclonal Antibody Reference Material (RM 8671; https://www.nist.gov/programs-projects/nist-monoclonal-antibody-reference-material-8671).

The use of tandem mass spectral libraries in biomedical and biomanufacturing applications has been very limited until recently with the development of omics technologies. To date, there are no reports of libraries being used for optimizing biomanufacturing processes and very little for discovering new metabolic pathways. Here, we implemented recurrent spectral libraries for use in CHO cell metabolite analysis that allows users to quickly identify most compounds in any complex metabolite sample. We also developed an annotation strategy for these libraries to filter out artifacts and low-quality spectra from recurrent unidentified spectra of metabolites. These libraries are focused on metabolite analysis, however, small peptides that extract along with the metabolites are also present. Furthermore, the limited coverage of tandem libraries is somewhat ameliorated by the use of the recently developed hybrid search (Burke et al., 2017; Cooper et al., 2019; Moorthy et al., 2017), which can identify compounds similar to, but not present in the library. The recurrent spectral library is unique in that it can be used to determine if an ion has been seen before in other analyses, assign the class identification for compounds not found in a library or commercially available, and enables library evolution based upon feedback from users. As more experiments are done, the library can continue to grow in coverage. The library and the associated metabolite identifications are freely available for download for use in the analysis of CHO cell metabolites by LC-MS/MS. Although this study was demonstrated in CHO cells, the developed methods for filtering spectra and assigning match confidence can be applied to not only other cell types, but also other metabolomics studies. In addition, work is currently underway at NIST to create a metabolite identification pipeline and graphical user interface (GUI) that those in the biomanufacturing community can use to implement their own libraries.

2 | EXPERIMENTAL METHODS*

For the coverage of metabolites to be broad, CHO cells were extracted by four different methods available in the literature: (1) 50% acetonitrile in water (Dietmair et al., 2012), (2) Methanol (Dietmair, Timmins, Gray, Nielsen, & Kromer, 2010; Sellick et al., 2011), (3) methanol/methyl tert-butyl ether(MTBE)/water, and (4) methanol/dichloromethane(DCM)/water (Matyash et al., 2008). Metabolites were separated with three different LC methods (reversed-phase [C_{18}], hydrophilic interaction liquid chromatography [HILIC] and a reversed-phase method optimized for lipids [lipid C_{18}]), and analyzed in positive and negative ionization mode with both higher-energy C-trap dissociation (HCD**) over a range of collision energies and ion trap (IT) collision-induced dissociation. Media samples (fresh and spent) were resuspended in two different solvents (50% acetonitrile or pure methanol) after protein precipitation, separated with two different LC methods (C_{18} and HILIC), and analyzed with the same breadth of methods as the CHO cell metabolites.

2.1 | Sample preparation

CHO-S cells (Thermo Fisher Scientific) were grown in ProCHO5 protein-free medium (Lonza) supplemented with 4 mmol/L L-glutamine (Thermo Fisher Scientific). CHO cells and spent media were harvested and metabolite extractions were performed. Protein precipitation was performed on the media with 80% (vol/vol) methanol. After drying and before analysis, media samples were resuspended in either pure methanol or 50% acetonitrile (vol/vol). Metabolites were extracted by four different methods: 50%

*Certain commercial instruments are identified in this document. Such identification does not imply recommendation or endorsement by The National Institute of Standards and Technology, nor does it imply that the products identified are necessarily the best available for the purpose.

**HCD is a term specific to the orbitrap mass spectrometer (Thermo Fisher Scientific). HCD and QTOF spectra are equivalent as they both result from beam-type collision-induced dissociation (versus ion trap collision-induced dissociation).

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

acetonitrile in water, methanol, methanol/methyl tert-butyl ether (MTBE)/water, and methanol/dichloromethane (DCM)/water. Additional details regarding sample preparation can be found in the supporting information.

2.2 | LC-MS/MS analysis

The metabolites were separated by three different liquid chromatography methods. Extracts containing polar metabolites (50% acetonitrile, methanol, lower phase for the methanol/MTBE/water extraction, and upper phase for the methanol/DCM/water extraction) were separated by both C_{18} and HILIC. The organic phases of the two lipid extractions were separated by a lipid C_{18} method. Fresh and spent media samples were separated by C_{18} and HILIC. These separations were coupled to either a Q Exactive or Orbitrap Fusion Lumos (Thermo Fisher Scientific). The data were collected in positive and negative ionization mode with data-dependent MS/MS acquisition. To provide as many spectra as possible for the library, HCD spectra were collected over a range of normalized collision energies from 10 to 50 using nitrogen as the collision gas. In addition, low-resolution IT and high-resolution IT spectra were acquired on the Lumos at a normalized collision energy of 35% using helium as the collision gas. The collision gases used were those recommended by the equipment manufacturer. Additional details regarding analysis can be found in the supporting information.

2.3 | Data analysis

Data were analyzed to produce recurrent spectral libraries as reported previously (Telu et al., 2016). Briefly, all data were processed with the NIST MSCQ pipeline (see below under "Annotation of Spectra" for a description of the pipeline). Recurrent spectra were exported from the output of the pipeline with a perfect score cutoff (1.0) to ensure all spectra (even identified ones) were included. Following this, consensus spectra were created from the experimental data using in-house developed software after grouping the data by polarity, fragmentation type (HCD or IT), and collision energy. The similarity of the spectra was based on precursor and the dot-product (Yang et al., 2014). Only similar spectra (a cluster) were used to create the consensus spectrum. Spectra dissimilar to the given cluster were placed in another cluster or, if unique, were ignored. After the libraries were created, the consensus spectra were searched against the NIST17 Library to obtain metabolite identifications. In addition, an annotation strategy was developed following manual evaluation of a representative data file. The data file analyzed was a 50% acetonitrile extraction that was separated on a C_{18} column and fragmented at HCD 20. The file was searched against the NIST17 Library with the NIST MSPepSearch software to provide tandem mass spectral library identifications as discussed below.

3 | RESULTS AND DISCUSSION

3.1 | Identification of metabolites

The first goal of this study was to collect, organize, and to the degree possible, identify all measurable tandem mass spectra in CHO cell metabolite and growth media extracts acquired using electrospray LC-MS/MS methods. To do this, we developed an HCD and IT fragmentation spectral library containing consensus spectra in both positive and negative ionization mode using a spectral clustering method developed in-house. The libraries contain data from both CHO cell metabolite analyses as well as media analyses and are annotated to show the origin of the spectra. In addition to metabolites, peptides that are co-extracted are also present in the libraries, although these are not the focus of the work. The resulting HCD recurrent spectral libraries contain 109,601 and 61,677 spectra for the positive and negative ionization mode libraries, respectively. The IT libraries contain 15,703 and 12,499 spectra for the positive and negative ionization mode libraries, respectively. IT spectra are similar to low energy HCD spectra, except for their low mass cut-off at about one-third of the precursor mass and their higher degree of fragmentation at these low energies; IT fragment ions are therefore more intense than low energy HCD spectra. Note that low energy spectra are generally easier to interpret than higher energy spectra due to their simpler mechanisms. Additional information about the libraries, including collision energies, precursor ion types, and source (CHO cell, media, or both) of the consensus spectra can be found in the supporting information. The results of CHO cell metabolite and media analyses are highly orthogonal as only 8%–13% of the consensus spectra in the libraries originate from both samples. The overlap would likely be higher if a chemically defined media was used.

To identify spectra, we searched the consensus spectra generated for the recurrent spectral libraries against the NIST17 MS/MS library (Yang et al., 2014; Yang et al., 2017). To compare our results with those previously published in the literature, the CHO cell metabolite identifications were summarized and compared to a literature review of CHO cell metabolite identifications. To summarize the identifications, we sorted the library match identifications by name and library match score. We kept only the top-scoring hit of each identification and then manually validated the library match result. Any poor matches were removed. In addition, we curated the data to remove identifications that are not previously observed as endogenous metabolites by searching for the identification in the Human Metabolome Database (HMDB) (Wishart et al., 2013, 2018), PubChem (Kim et al., 2019), or the LIPID MAPS structure database (Sud et al., 2007) as no comprehensive CHO cell metabolite library is available. If there was no information on if an identification was a metabolite, it was not removed. Spreadsheet 1 of supporting information contains all the library match identifications and can be mined for new or unexpected metabolites by experts in CHO cell metabolism. Our curated list resulted in 365 CHO cell metabolites

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(the majority identified by multiple ions or in multiple libraries) and an additional 304 di- or tri-peptides. We split out the peptides into a separate list because they are likely less interesting than other metabolites. Metabolites identified are reported in Table 1. A literature search resulted in a list of 232 metabolites. Identifications made by HPLC, GC-MS, MALDI-MS, and LC-MS were included. Of these 232 reported metabolites, we identified 43% in our data. For ones that were not identified, the majority (66%) were represented in the NIST17 library, but not identified in our experiments, possibly because they were below the detection limit. The remaining literature identifications not present in the NIST17 library that are compatible with analysis by LC-MS can be added to future versions of the NIST MS/MS library. Lists of identified metabolites summarized from our data as well as the literature review can be found in Spreadsheets 2 and 3 of the supporting information, respectively. These spreadsheets also contain information demonstrating the percentages reported herein.

3.2 | Improvement of accuracy of pipeline identifications

We developed a procedure to improve the accuracy of identifications obtained using the NIST MSQC pipeline by modifying the order of identifications in a hit list. The NIST pipeline, by default, sorts hits entirely by their score which reflects the quality of the spectral match between the experimental and library spectra. We identified four categories of errors in identification. For clarity, we labeled these as category A–D errors. Additional information on the errors, examples, and solutions for these errors can be found in the supporting information.

3.3 | Hybrid search

To discover the identity of compounds not represented in the library, a hybrid search was performed. The hybrid search match is a new search strategy available in the 2017 release of NIST MS Search software (version 2.3) (Burke et al., 2017; Cooper et al., 2019; Moorthy et al., 2017). This search finds compounds that differ by an inert chemical group, hence, can often match unidentified spectra with members of the same chemical classes that are present in the library. The term delta mass is used to represent the difference in mass between the query spectrum and library entry. An example of a hybrid match in the CHO cell metabolite data is for the match of a spectrum (ion $m/z = 472.0011$) to a sodiated Adenosine 5'-diphosphate library spectrum with a delta mass of 21.9824 Da. This delta mass corresponds to a sodium, so the correct annotation of this ion is adenosine 5'-diphosphate $[M\text{-}H+2Na]^+$. The hybrid search was also utilized to assist in the identification of two groups of related spectra. Information on these identifications can be found in the supporting information.

3.4 | Utility of recurrent spectral libraries

There are multiple metabolomics analysis software tools available. A recent review summarized those that are freely available (Spicer et al., 2017). In addition, there are a variety of freely available packages for processing MS/MS spectra (Kind et al., 2018). One such tool, RAMClustR (Broeckling et al., 2014) can group features extracted via XCMS (Smith et al., 2006) into spectra in an unsupervised manner and therefore identify features that originate from the same compound in an indiscriminant MS/MS (idMS/MS) data acquisition. Spectra can then be searched against a reference library such as the NIST MS/MS Library. The NIST MSQC pipeline (Rudnick et al., 2010; https://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:msqcpipeline), a fully integrated software pipeline that was developed for the analysis of a tryptic protein digest to assist in the identification of variability caused by issues with analytical platforms, was used to process data files in this study. We have extended the application of the pipeline to identification of small molecule metabolites by modifying searching and scoring. The pipeline begins by reading a data file from a commercial instrument, extracting all spectral data, and searching the spectra against the NIST library using the NIST MS Search software. When multiple spectra are acquired for a single precursor ion, the most intense one is selected and its maximum MS1 abundance is recorded at its retention time. Figure 1 shows ion plots generated from the pipeline output after searching against the NIST17 MS/MS library or the Recurrent library and provides a visual representation of the data. Each object represents a clustered mass spectrum. More detailed plots of those shown in Figure 1 can be found in the supporting information. For these ion plots, the pipeline has found 5335 ion clusters in this data file. When searched against the NIST17 MS/MS library, 80% of these clusters have no identification. When searched against the positive ion HCD recurrent spectral library, the number of clusters with no identification drops to 23%. Thirty-eight percent of the clusters have a recurrent label, which indicates they have matched spectra in the recurrent spectral library by either direct or hybrid MS/MS search. This increase in cluster identification demonstrates the utility of the recurrent spectral libraries. As we are cataloguing every observed ion in the libraries instead of just previously identified metabolites, we can identify these ions in future analyses of the same or similar materials.

3.5 | Annotation of spectra

The second goal of this study was to develop a comprehensive, automatable approach to annotate the spectra in the libraries of recurrent spectra for the purpose of filtering out artifacts and low-quality spectra from recurrent unidentified spectra of metabolites. This type of filtering is important because unknowns can be redundant signals, artifacts (man-made signals), and contaminants (real chemicals), instead of metabolites that are not present in the library used for spectral matching (Sindelar & Patti, 2020). Credentialing

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Compounds identified in the Recurrent Spectral Library created from CHO cell metabolite extracts

Metabolite	Library	PubChem ID
10Z-Nonadecenoic acid	HCD-Pos	5312513
1-Methylnicotinamide	HCD-Pos	457
1-Methylxanthine	HCD-Pos	80220
2,3-Dehydro-2-deoxy-*N*-acetylneuraminic acid	HCD-Pos, IT-Pos	65309
2,3-Diaminopropionic acid	HCD-Pos	364
2-Arachidonyl glycerol ether	HCD-Pos	6483057
2'-Deoxyguanosine 5'-monophosphate	HCD-Neg, IT-Neg	645
2-hydroxy-2-(4-hydroxy-3-methoxyphenyl)acetic acid	HCD-Pos	1245
2-Hydroxyhexadecanoic acid	HCD-Neg	92836
2-Hydroxyphenethylamine	HCD-Pos	1000
2-Methylbutyrylcarnitine	HCD-Pos	6426901
2-Methylhippuric acid	HCD-Pos	91637
2'-*O*-Methyladenosine	HCD-Neg, IT-Neg	102213
2-Phospho-D-glyceric acid	HCD-Neg, IT-Neg	59
3,4-Dihydroxymandelic acid	HCD-Pos	85782
3'-AMP	HCD-Neg, HCD-Pos	41211
3'-CMP	HCD-Neg, HCD-Pos, IT-Pos	66535
3-Deoxy-D-glycero-D-galacto-2-nonulosonic acid	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	123691
3-Hexenedioic acid	HCD-Pos	107550
3-Oxoglutaric acid	HCD-Pos	68328
3-Phosphoglycerate	HCD-Neg	724
3-Sialyl-*N*-acetyllactosamine	HCD-Neg, HCD-Pos, IT-Neg	4150746
4-Coumaryl alcohol	HCD-Pos	5280535
4-Hydroxybutyric acid	HCD-Neg	10413
4-Hydroxyglutamic acid	HCD-Neg	439902
5α-cholest-7-en-3β-ol	HCD-Pos, IT-Pos	420
5-Aminovaleric acid	HCD-Pos	138
5-Hydroxyindole	HCD-Pos	16054
5-Hydroxylysine	HCD-Pos	439437
5'-Methylthioadenosine	HCD-Pos, IT-Pos	149
5-Phosphonatoribosyl 1-pyrophosphate	HCD-Neg	1041
5-Thymidylic acid	HCD-Neg, IT-Neg	1139
6-Phosphogluconic acid	HCD-Pos	91493
7-Ketocholesterol	HCD-Pos	91474
7-Methylguanine	HCD-Pos	135398679
7-Methylguanosine	HCD-Pos	135445750
9,10-Epoxyoctadecenoic acid	HCD-Pos	5283018
Acetylcholine	HCD-Pos	187
Acetyl-CoA	HCD-Neg, HCD-Pos, IT-Neg	6302
Adenine	HCD-Pos	190

(Continues)

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Metabolite	Library	PubChem ID
Adenosine	HCD-Pos	60961
Adenosine 2',3'-cyclic phosphate	HCD-Neg, HCD-Pos, IT-Pos	2024
Adenosine 2'-phosphate	HCD-Neg, HCD-Pos	94136
Adenosine diphosphate ribose	HCD-Neg, HCD-Pos, IT-Pos	30243
Adenosine monophosphate	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	6083
Adenosine phosphosulfate	HCD-Neg	10238
Adenosine triphosphate	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	5957
Adenylsuccinic acid	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	440122
ADP	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	6022
Agmatine	HCD-Pos	199
α-D-Glucose 1,6-bisphosphate	HCD-Neg	82400
α-Ionone	HCD-Pos	24680
α-Ketoisovaleric acid	IT-Neg	49
Aminoadipic acid	HCD-Pos	469
Arabinonic acid	HCD-Neg, IT-Neg	122045
Aspartylglycosamine	HCD-Pos	123826
Asymmetric dimethylarginine	HCD-Pos, IT-Pos	123831
β-Carboline	HCD-Pos	64961
β-Glycerophosphoric acid	HCD-Neg, IT-Neg	2526
Betaine	HCD-Pos	247
Biopterin	HCD-Pos	135403659
But-2-enoic acid	HCD-Neg	637090
Carnosine	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	439224
CDP	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	6132
Cer(d18:1/24:1(15Z))	HCD-Pos, IT-Pos	5283568
Cholest-5-en-3-one	HCD-Pos	9908107
Cholesta-4,6-dien-3-one	HCD-Pos, IT-Pos	3034666
Cholesterol	HCD-Pos	5997
Choline	HCD-Pos, IT-Pos	305
cis-Aconitic acid	HCD-Neg	309
cis-Vaccenic acid	HCD-Pos	5282761
Citicoline	HCD-Neg, HCD-Pos, IT-Pos	13804
Citraconic acid	HCD-Neg	643798
Citric acid	HCD-Neg, IT-Neg	311
Citrulline	HCD-Neg, HCD-Pos	833
Coenzyme A	HCD-Neg	87642
Coenzyme Q9	HCD-Pos, IT-Pos	5280473
Cyclic ADP-ribose	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	123847
Cyclic AMP	HCD-Neg	6076

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27761 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Metabolite	Library	PubChem ID
Cytidine	HCD-Neg, HCD-Pos	6175
Cytidine 5'-diphosphate ethanolamine	HCD-Neg, HCD-Pos, IT-Neg	123727
Cytidine monophosphate	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	6131
Cytidine monophosphate *N*-acetylneuraminic acid	HCD-Neg	448209
Cytidine triphosphate	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	6176
Cytosine	HCD-Pos	597
dCDP	HCD-Neg, IT-Neg	150855
dCMP	HCD-Neg, HCD-Pos, IT-Neg	13945
Deoxyadenosine monophosphate	HCD-Neg	12599
Deoxycytidine	HCD-Pos	13711
Deoxyinosine	HCD-Pos, IT-Pos	135398593
Dephospho-CoA	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	444485
D-Erythrose	HCD-Neg	94176
D-Fructose	HCD-Neg	5984
DG(14:0/14:0/0:0)	HCD-Pos	10369168
DG(16:0/16:0/0:0)	HCD-Pos	644078
DG(16:0/18:1(9Z)/0:0)	HCD-Pos, IT-Pos	5282283
DG(18:1(9Z)/18:1(9Z)/0:0)	HCD-Pos, IT-Pos	9543716
D-Galactose	HCD-Neg, IT-Neg	6036
D-Glucaro-1,4-lactone	HCD-Neg, IT-Neg	122306
D-Glucose	HCD-Neg	5793
D-Glucuronic acid	HCD-Neg	94715
Diadenosine triphosphate	HCD-Neg	165381
Dihydrobiopterin	HCD-Pos	135402011
D-Malic acid	HCD-Neg, IT-Neg	525
D-Maltose	HCD-Neg, IT-Pos	294
D-Mannose 1-phosphate	HCD-Neg, IT-Neg	644175
D-Ornithine	HCD-Pos	71082
D-Phenyllactic acid	HCD-Neg	643327
D-Pipecolinic acid	HCD-Pos	736316
ε-caprolactam	HCD-Pos, IT-Pos	7768
Erucamide	HCD-Pos, IT-Pos	5365371
Erucic acid	HCD-Pos, IT-Pos	8216
FAPy-adenine	HCD-Pos	114926
Flavin mononucleotide	HCD-Pos, IT-Pos	710
Folic acid	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	135398658
Fructose 1,6-bisphosphate	HCD-Neg, IT-Neg	10267
Fructose-6-phosphate	HCD-Neg	69507
Galactaric acid	HCD-Neg	3037582

(Continues)

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Metabolite	Library	PubChem ID
Galactinol	HCD-Neg, IT-Neg	11727586
Galactitol	HCD-Neg	11850
Galactonic acid	HCD-Neg	128869
Galactose 1-phosphate	HCD-Neg, IT-Neg	123912
Galactosylsphingosine	HCD-Pos	5280458
Galβ1,3GlcNAc	HCD-Pos	440994
ϒ-Aminobutyric acid	HCD-Pos	119
ϒ-Glutamylglutamic acid	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	92865
GDP-glucose	HCD-Pos	135398625
GDP-L-fucose	HCD-Neg	135398655
Glucaric acid	HCD-Neg, IT-Neg	33037
Glucose 1-phosphate	HCD-Neg	65533
Glucose 6-phosphate	HCD-Neg, HCD-Pos, IT-Neg	5958
Glutathione	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	124886
Glyceraldehyde 3-phosphate	IT-Neg	729
Glyceric acid	HCD-Neg	752
Glycerophosphocholine	HCD-Pos, IT-Pos	71920
Glyceryl monooleate	HCD-Pos	33022
Guanidinosuccinic acid	HCD-Neg, HCD-Pos	97856
Guanine	HCD-Neg, HCD-Pos	135398634
Guanosine	HCD-Neg, HCD-Pos, IT-Neg	135398635
Guanosine diphosphate	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	135398619
Guanosine diphosphate mannose	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	135398627
Guanosine monophosphate	HCD-Neg, HCD-Pos, IT-Pos	135398631
Guanosine triphosphate	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	135398633
Helicin	HCD-Pos	101799
Hexadecanedioic acid	HCD-Pos	10459
Hydroxyphenyllactic acid	HCD-Neg, HCD-Pos, IT-Neg	9378
Hypoxanthine	HCD-Pos	135398638
Indole-3-carboxylic acid	HCD-Pos	69867
Indolelactic acid	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	92904
Indolepyruvate	HCD-Pos	803
Inosine	HCD-Pos	135398641
Inosinic acid	HCD-Neg, HCD-Pos, IT-Neg	135398640
Inositol 1,3,4-trisphosphate	HCD-Neg	123680
Inositol 1,3-bisphosphate	HCD-Neg, IT-Neg	128419
Inositol 1,4,5-trisphosphate	HCD-Pos	55310
Inositol 1,4-bisphosphate	HCD-Neg, HCD-Pos	123903
Inositol 1-phosphate	IT-Pos	107737

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Metabolite	Library	PubChem ID
Inositol 3-phosphate	HCD-Pos	440194
Inositol 4-phosphate	HCD-Neg	440043
Isobutyryl-L-carnitine	HCD-Pos	168379
Isocitric acid	HCD-Neg	1198
Isomaltose	HCD-Neg	872
Isovaleryl coenzyme A	HCD-Neg	165435
Isovaleryl-L-carnitine	HCD-Pos	169235
Ketoleucine	HCD-Neg	70
L2-Hydroxyglutaric acid	HCD-Neg	43
Lacto-*N*-triaose	HCD-Pos	53477860
Lactose	HCD-Neg	6134
L-Arginine	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	6322
L-Asparagine	HCD-Neg, HCD-Pos	6267
L-Aspartic acid	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	424
L-Carnitine	HCD-Pos, IT-Pos	10917
L-Cystathionine	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	834
L-Cystine	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	67678
L-Erythrulose	HCD-Neg	162406
Lewis A trisaccharide	HCD-Pos	4139998
Lewis X trisaccharide	HCD-Pos, IT-Pos	4571095
L-Glutamic acid	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	33032
L-Glutamine	HCD-Neg, HCD-Pos	5961
L-Gulonolactone	HCD-Neg, IT-Neg	439373
L-Histidine	HCD-Neg, HCD-Pos	6274
L-Homoserine	HCD-Pos, IT-Pos	12647
L-Iditol	HCD-Pos	5460044
L-Isoleucine	HCD-Pos	6306
L-Kynurenine	HCD-Pos	161166
L-Leucine	HCD-Neg, HCD-Pos, IT-Neg	6106
L-Lysine	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	5962
L-Methionine	HCD-Pos, IT-Pos	6137
L-Phenylalanine	HCD-Neg, HCD-Pos, IT-Pos	6140
L-Proline	HCD-Neg, HCD-Pos	145742
L-Serine	HCD-Neg, HCD-Pos, IT-Pos	5951
L-Threonine	HCD-Neg, HCD-Pos, IT-Pos	6288
L-Tryptophan	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	6305
L-Tyrosine	HCD-Neg, HCD-Pos, IT-Pos	6057
L-Valine	HCD-Pos	6287
Maltotetraose	HCD-Neg, HCD-Pos, IT-Pos	870

(Continues)

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Metabolite	Library	PubChem ID
Maltotriose	HCD-Neg, IT-Neg	92146
Mannose 6-phosphate	HCD-Neg, HCD-Pos, IT-Neg	65127
Melibiose	HCD-Neg	219994
Methionine sulfoxide	HCD-Neg, HCD-Pos, IT-Pos	158980
MG(0:0/16:0/0:0)	HCD-Pos, IT-Pos	123409
myo-Inositol	HCD-Neg	892
N8-Acetylspermidine	HCD-Pos, IT-Pos	123689
N-Acetyl-D-galactosamine	HCD-Pos	35717
N-Acetyl-D-glucosamine	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	439174
N-Acetyl-D-glucosamine 6-phosphate	HCD-Neg	439219
N-Acetyl-D-lactosamine	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	9800166
N-Acetyl-L-aspartic acid	HCD-Pos	65065
N-Acetyl-L-carnosine	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	9903482
N-Acetyl-L-glutamic acid	HCD-Neg	70914
N-Acetyl-L-glutamine	HCD-Pos, IT-Pos	182230
N-Acetyl-L-methionine	HCD-Neg, IT-Neg	448580
N-Acetyl-L-phenylalanine	HCD-Neg, HCD-Pos, IT-Pos	74839
N-Acetyl-L-tyrosine	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	68310
N-Acetylmannosamine	HCD-Pos, IT-Pos	65150
N-Acetylneuraminic acid	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	906
NAD	HCD-Neg, HCD-Pos, IT-Pos	925
NADH	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	439153
NADP	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	4412
N-alpha-Acetyl-L-ornithine	HCD-Pos	907
N-Formyl-L-methionine	HCD-Neg, IT-Neg	439750
N-Glycolylneuraminic acid	HCD-Neg, IT-Neg	123802
Niacinamide	HCD-Pos, IT-Pos	936
Nicotinamide riboside	HCD-Pos	439924
Nicotinamide ribotide	HCD-Pos	16219737
Nicotinic acid adenine dinucleotide	HCD-Neg	165490
Nicotinic acid mononucleotide	HCD-Neg, IT-Neg	5288991
N-Methyl-L-glutamic acid	HCD-Pos	439377
N-Methyllysine	HCD-Pos	164795
N-Methyltyramine	HCD-Pos	9727
N-Palmitoyl-D-sphingosine	HCD-Pos, IT-Pos	5353456
Oleamide	HCD-Pos	5283387
Oleic acid	HCD-Pos	445639
Oleoyl glycine	HCD-Pos	6436908
Oleoyl serine	HCD-Neg, HCD-Pos, IT-Pos	44190514

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Metabolite	Library	PubChem ID
O-Phosphotyrosine	IT-Pos	30819
Orotic acid	HCD-Neg	967
O-Tyrosine	HCD-Pos	91482
Oxidized glutathione	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	65359
PA(16:0/16:0)	HCD-Neg	3099
PA(16:0/18:1(9Z))	HCD-Neg	5283523
PA(18:1/0:0)	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	5311263
Palmitic amide	HCD-Pos	69421
Palmitoyl ethanolamide	HCD-Pos	4671
Palmitoyl sphingomyelin	HCD-Neg, HCD-Pos, IT-Pos	9939941
Pantothenic acid	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	6613
Paullinic acid	HCD-Pos	5312518
PC(14:0/0:0)	HCD-Neg, HCD-Pos, IT-Pos	460604
PC(14:0/14:0)	HCD-Neg, HCD-Pos, IT-Neg	5459377
PC(14:0/16:0)	HCD-Neg, HCD-Pos	129657
PC(14:0/18:0)	HCD-Pos	131150
PC(15:0/15:0)	HCD-Pos, IT-Neg	24778654
PC(16:0/0:0)	HCD-Pos, IT-Pos	460602
PC(16:0/12:0)	HCD-Pos	10676014
PC(16:0/14:0)	HCD-Neg, HCD-Pos, IT-Neg	24778679
PC(16:0/16:0)	HCD-Pos	452110
PC(16:0/18:1(9Z))	HCD-Neg, HCD-Pos, IT-Pos	5497103
PC(16:0/18:2(9Z,12Z))	HCD-Pos	5287971
PC(16:1(9Z)/16:1(9Z))	HCD-Pos	24778764
PC(18:0/0:0)	HCD-Neg, HCD-Pos	497299
PC(18:0/14:0)	HCD-Pos	3082163
PC(18:0/18:0)	HCD-Pos	94190
PC(18:0/18:1(9Z))	HCD-Neg	24778825
PC(18:0/18:2(9Z,12Z))	HCD-Pos	6441487
PC(18:1(9Z)/0:0)	HCD-Neg, HCD-Pos, IT-Pos	16081932
PC(18:1(9Z)/14:0)	HCD-Pos, IT-Neg	24778931
PC(18:1(9Z)/16:0)	HCD-Neg, HCD-Pos	24778933
PC(20:1(11Z)/20:1(11Z))	HCD-Pos	24779063
PC(22:0/0:0)	HCD-Pos	24779479
PC(22:1(13Z)/22:1(13Z))	HCD-Pos	24779126
PC(24:0/0:0)	HCD-Pos	24779481
PC(O-16:0/0:0)	HCD-Pos, IT-Pos	162126
PC(O-16:0/18:1(9Z))	HCD-Pos	24779266
PC(O-16:0/2:0)	HCD-Pos	108156

(Continues)

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Metabolite	Library	PubChem ID
PC(O-16:0/20:3(8Z,11Z,14Z))	HCD-Pos	16759365
PC(O-18:0/0:0)	HCD-Pos	2733532
PC(P-16:0/0:0)	HCD-Pos	10917802
PC(P-18:0/0:0)	HCD-Neg, HCD-Pos	24779527
PC(P-18:0/18:1(9Z))	HCD-Pos	42607428
PE(14:0/0:0)	HCD-Neg, HCD-Pos, IT-Neg	9547070
PE(16:0/0:0)	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	9547069
PE(16:0/16:0)	HCD-Neg, IT-Neg	445468
PE(16:0/18:1(9Z))	HCD-Pos	5283496
PE(16:0/18:2(9Z,12Z))	HCD-Pos	9546747
PE(18:0/0:0)	HCD-Neg, HCD-Pos, IT-Pos	9547068
PE(18:0/18:1(9Z))	HCD-Pos	9546742
PE(18:0/18:2(9Z,12Z))	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	9546749
PE(18:1(9Z)/0:0)	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	9547071
PE(18:1(9Z)/18:1(9Z))	HCD-Pos	9546757
PE(O-16:0/18:1(9Z))	HCD-Neg, HCD-Pos	42607455
PE(P-18:0/18:1(9Z))	HCD-Neg	42607457
PE(P-18:0/22:6(4Z,7Z,10Z,13Z,16Z,19Z))	HCD-Neg, IT-Neg	42607458
PE-NMe2(18:1(9Z)/18:1(9Z))	HCD-Neg, IT-Neg	9547022
PG(18:0/18:1)	HCD-Neg, IT-Neg	24779551
Phenylacetic acid	HCD-Neg	999
Phenylacetylglutamine	HCD-Neg, HCD-Pos, IT-Pos	92258
Phosphoadenosine phosphosulfate	HCD-Neg, IT-Neg	10214
Phosphorylcholine	HCD-Pos, IT-Pos	135437
Phosphoserine	HCD-Neg	106
PI(16:0/18:1(9Z))	HCD-Neg, IT-Neg	5771758
Pip(18:1(9Z)/18:1(9Z))	HCD-Neg	53480169
p-Octopamine	HCD-Pos	4581
Proline betaine	IT-Pos	115244
PS(16:0/18:1(9Z))	HCD-Neg, HCD-Pos, IT-Neg	5283499
PS(16:0/20:4)	HCD-Neg	24779544
PS(18:0/18:0)	HCD-Neg	9547096
PS(18:0/18:1(9Z))	HCD-Neg, HCD-Pos, IT-Neg	9547087
PS(18:0/20:4(5Z,8Z,11Z,14Z))	HCD-Neg	24779545
PS(18:1(9Z)/18:1(9Z))	HCD-Neg, HCD-Pos, IT-Neg	6438639
Pterin	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	73000
Pyridoxal	HCD-Pos, IT-Pos	1050
Pyridoxal 5'-phosphate	HCD-Neg, HCD-Pos	1051
Pyridoxamine	HCD-Pos, IT-Pos	1052

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Metabolite	Library	PubChem ID
Pyroglutamic acid	HCD-Neg, HCD-Pos, IT-Pos	7405
Raffinose	HCD-Neg, HCD-Pos	10542
Ribitol	HCD-Neg	6912
Riboflavin	HCD-Neg, HCD-Pos, IT-Pos	493570
Ribono-ϒ-lactone	HCD-Neg	111064
Ribose 1-phosphate	HCD-Neg, IT-Neg	1074
Ribose 5-phosphate	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	77982
Ribulose 5-phosphate	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	439184
S-Adenosylhomocysteine	HCD-Pos, IT-Pos	13792
S-Adenosylmethionine	HCD-Pos, IT-Pos	34755
Sebacic acid	HCD-Neg, IT-Neg	5192
Sedoheptulosan	HCD-Neg	5460956
Serotonin	HCD-Pos	5202
S-Glutathionyl-L-cysteine	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	10455148
SM(d18:1/18:0)	HCD-Pos	6453725
SM(d18:1/18:1(9Z))	HCD-Pos	6443882
SM(d18:1/24:1(15Z))	HCD-Pos	53481791
Sorbitol	HCD-Neg, IT-Neg	5780
Spermine	HCD-Pos	1103
Stachyose	HCD-Pos, IT-Pos	439531
Stearoyl ethanolamide	HCD-Pos	27902
Succinic acid	HCD-Neg, IT-Neg	1110
Succinic acid semialdehyde	HCD-Neg	1112
Sucrose	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	5988
Tetradecanoyl-CoA	HCD-Neg	11966124
Thiamine	HCD-Pos, IT-Pos	1130
Thiamine monophosphate	HCD-Pos	3382778
Thiamine pyrophosphate	HCD-Neg, HCD-Pos, IT-Pos	5431
Threonic acid	HCD-Neg, IT-Neg	151152
trans-13-Octadecenoic acid	HCD-Pos	6161490
trans-Vaccenic acid	HCD-Pos	5281127
Trehalose	HCD-Neg	7427
Trigonelline	HCD-Pos	5570
Triolein	HCD-Pos, IT-Pos	5497163
Tripalmitolein	HCD-Pos	9543989
Ubiquinone-1	HCD-Pos	4462
Undecanedioic acid	HCD-Neg, IT-Neg	15816
Uracil	HCD-Pos	1174
Uric acid	HCD-Neg, HCD-Pos, IT-Neg	1175

(Continues)

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 (Continued)

Metabolite	Library	PubChem ID
Uridine	HCD-Pos	6029
Uridine 5'-diphosphate	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	6031
Uridine 5'-monophosphate	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	6030
Uridine 5'-triphosphate	HCD-Neg, HCD-Pos, IT-Neg	6133
Uridine diphosphate glucose	HCD-Pos, IT-Pos	8629
Uridine diphosphate glucuronic acid	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	17473
Uridine diphosphategalactose	HCD-Neg, HCD-Pos, IT-Neg, IT-Pos	6857410
Urocanic acid	HCD-Pos	736715
Xanthine	HCD-Pos, IT-Pos	1188
Xanthosine	HCD-Pos	64959
Xanthylic acid	HCD-Neg, HCD-Pos, IT-Pos	73323
Xylulose 5-phosphate	HCD-Neg, HCD-Pos, IT-Pos	439190
Zymosterol	HCD-Pos	92746

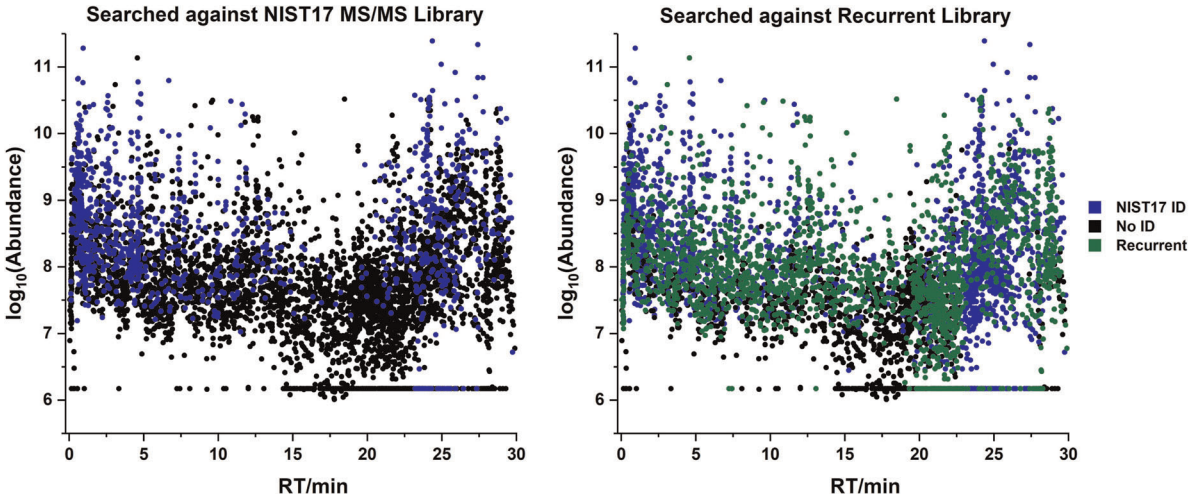

FIGURE 1 Plot of a single LC-MS/MS analysis of a 50% acetonitrile extract of CHO cell metabolites after searching against the NIST17 MS/MS library (left) or Recurrent Library (right). LC, liquid chromatography; MS, mass spectrometry [Color figure can be viewed at wileyonlinelibrary.com]

features (Mahieu et al., 2014) and isotopic ratio outlier analysis (IROA; de Jong & Beecher, 2012) are two isotopic-labeling techniques that have been developed to provide further confidence in metabolite identifications. Such techniques were not used in the creation of the mass spectral libraries. Therefore, an annotation strategy based on the comparison of the extracted ion chromatogram (EIC) between the sample and blank runs was developed to filter library spectra. Once spectra are filtered, efforts can be focused on identifying compounds that are likely to be unidentified recurrent spectra originating from CHO cells and/or media (vs. the environment or instrument) by searching the spectra against other available tandem mass spectral libraries and *in silico* prediction libraries. MassBank of North America (MoNA) in combination with the NIST MS/MS Library and *in silico* fragmentation tools CSI:FingerID and LipidBlast has been demonstrated to be effective in assigning structural annotation to MS/MS spectra (Blaženović et al., 2019).

To develop the annotation strategy, the search results of all the mass spectra contained in a representative data file were manually evaluated. Figure 2 is a graphical summary of the annotation strategy developed for filtering. First, spectra are removed if they do not have a sufficiently narrow chromatographic peak width (<30 s), unless they are identified. Second, spectra without sufficiently high spectral purity (>80%) are removed. Third, spectra without sufficient fragment ion abundances (summed product ion abundance/precursor ion

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

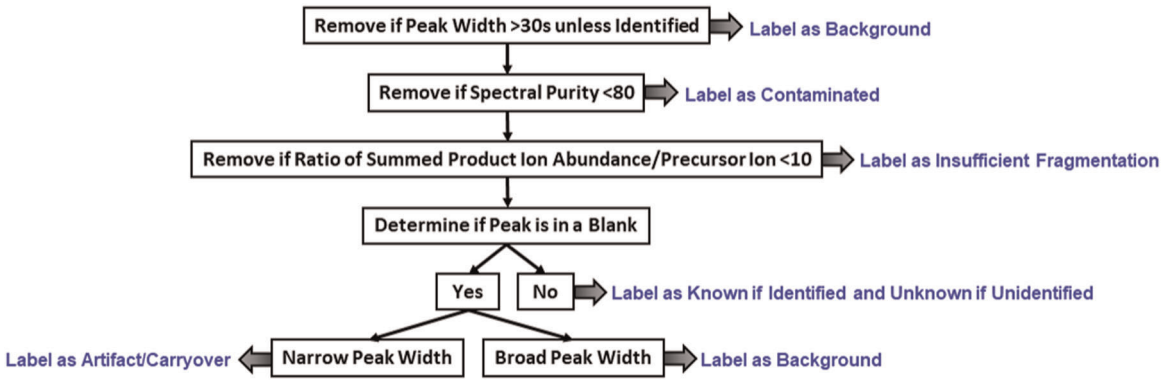

FIGURE 2 Workflow for annotation of spectra [Color figure can be viewed at wileyonlinelibrary.com]

abundance < 10) are removed. The data shown in Figure 1 was filtered using these parameters, which eliminated two-thirds of the 5335 spectra. The spreadsheet used for sorting and eliminating spectra can be found in Spreadsheet 4 of the supporting information. Of those eliminated, 9.5% of were background, 77.8% were possibly contaminated (due to the presence of another peak close in mass to the parent ion), and 12.8% contained insufficient fragmentation. Figure 3 shows the distribution of abundances of the 1752 identified (by direct and hybrid MS/MS match) and unidentified ion clusters. This figure shows that less abundant compounds are less likely to be identified. These unidentified ion clusters could be comprised of spectra of previously unidentified metabolites or metabolites that are not represented in the library as well as spectra of background and artifacts, which is why annotation is crucial.

In the next step in spectral annotation after filtering, the EIC of the ion of the corresponding spectrum is compared to the EIC of the same ion in a blank run via visual inspection. If the peak is not present in the blank with intensity within 100x that of the sample, then this spectrum is labeled as either a known (if it is identified by MS/MS match) and annotated with the identification or as an unknown (metabolite not identified by library searching). If the peak is in the blank, then the spectrum can either be due to an artifact/carryover or background. During manual evaluation of EICs, we found that for the purposes of spectral classification, artifact/carryover ions can be

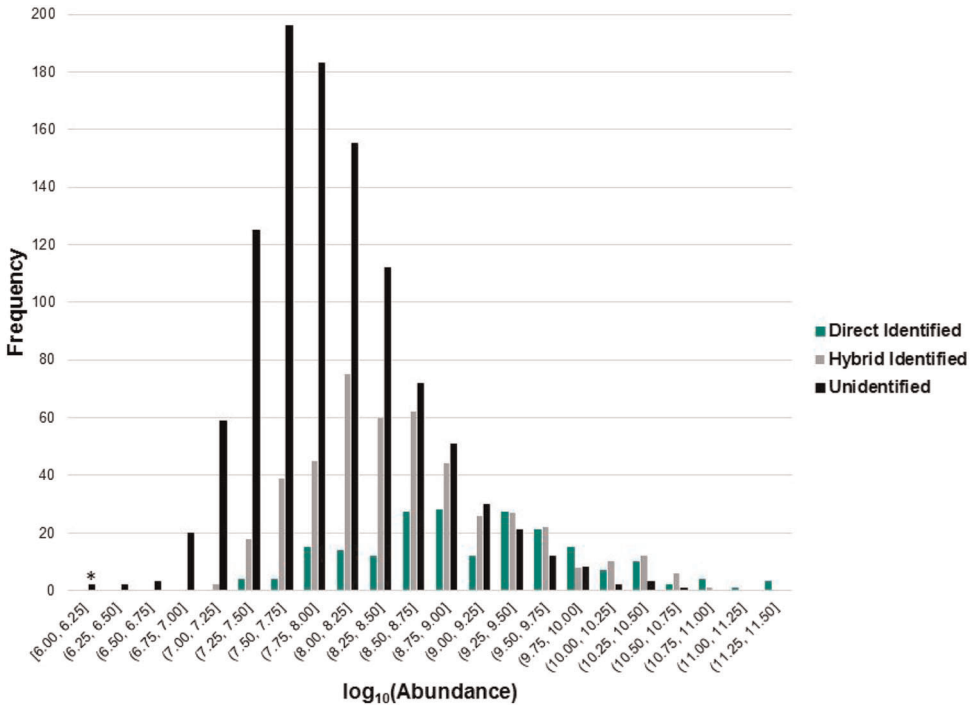

FIGURE 3 Distribution of identified and unidentified ion clusters after filtering.
*0.2% of hybrid identified and 2.9% of unidentified spectra that were assigned abundance in this bin because the abundance could not be calculated were removed [Color figure can be viewed at wileyonlinelibrary.com]

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

separated from background ions by examining the peak width. The background has a broad peak width (in regions where both hydrophilic and hydrophobic compounds elute) while an artifact/carryover has a narrow peak width. For most of the cases, differentiating between background and artifact/carryover was straightforward, however, for cases that were difficult to differentiate, we labeled ions as background if there was a substantial signal in both halves of the chromatogram (0–15 and 15–30 min). Separation using peak width is a quick method to classify spectra, but more accurate methods could be applied in an automated pipeline. Multiple algorithms (Cleary et al., 2019; Ho, Kuo, Wang, Chen, & Tseng, 2013; Zhang & Yang, 2008; Zhu et al., 2009) have been developed for the purpose of subtracting the background from LC-MS data. In addition, a hierarchical cluster analysis technique was developed to identify chemical interferants that are not removable by background subtraction (Caesar, Kvalheim, & Cech, 2018). Figure 4 shows examples of each of the above-mentioned classifications. For the unidentified recurrent spectra, these classifications are an effective way (that can be automated) to annotate the spectra for the library. These labels allow us to prioritize spectra needing identification first through library and literature searching. Unknowns represent compounds that originate from the CHO cells or cell culture media and are the highest priority to attempt to identify. Artifacts/carryover are the next priority because these may still be compounds that originate from the CHO cells or cell culture media. Background spectra are likely not worth an analyst's time to try to identify as the background will be different in analyses from different labs. Table S4 shows the resultant annotation of the 20 most abundant unidentified ion clusters. Fifteen percent of the clusters are unknowns and would be the most useful to search the literature and online databases for the identities. Fifty percent of the clusters are artifacts/carryover and the remaining 35% are background.

3.6 | Confidence in library match identifications

A framework for reporting confidence in metabolomics identifications was proposed in 2007 by the Chemical Analysis Working Group of the Metabolomics Standards Initiative (MSI) and is composed of four levels of metabolite confidence. These are identified compounds (Level 1), putatively annotated compounds (Level 2), putatively characterized compound classes (Level 3), and unknown compounds (Level 4) (Sumner et al., 2007). There has been discussion in the metabolomics community about providing more information about confidence by modifying/expanding the level system, introducing a quantitative system, or providing alphanumeric identification metrics, but no consensus has been reached (Creek et al., 2014; Schrimpe-Rutledge et al., 2016; Schymanski et al., 2014; Sumner et al., 2014; Viant et al., 2017). Schymanski et. al. (2014) proposed a framework for reporting confidence that was based on the MSI levels

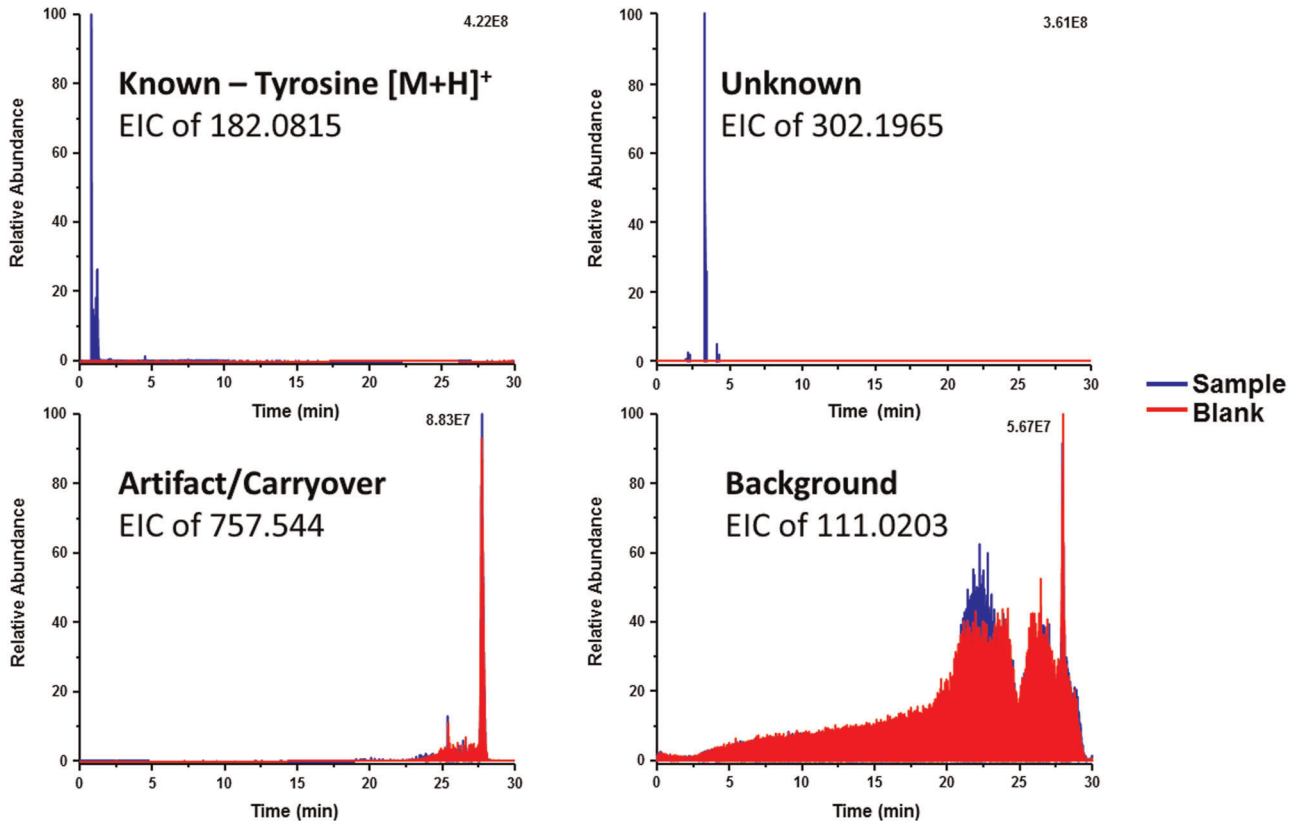

FIGURE 4 Examples of each type of annotated ion [Color figure can be viewed at wileyonlinelibrary.com]

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and adapted for high-resolution mass spectrometry (HR-MS). These HR-MS specific confidence levels are most appropriate for our data and consist of five confidence levels. These are confirmed structure (Level 1), probable structure (Level 2), tentative candidate(s) (Level 3), unequivocal molecular formula (Level 4), and exact mass (Level 5). In this study, we have Level 2, 3, and 5 confidences. Level 1 is confirmed using two or more properties of reference standards using the same experimental conditions. Although the NIST17 MS/MS library is acquired using reference standards, the experimental data in this paper was not acquired on the same platform, so it is not a Level 1 confidence. This type of confirmation is unrealistic for our work where we are trying to catalogue all metabolites and identify as many as possible. The direct identifications reported in this study represent Level 2 confidence structure identifications as they are obtained with library matching. Hybrid match identifications are Level 3 because they are chemical class identifications made with library searching. To assign a Level 4 confidence, we would need to attempt to assign a chemical structure to the spectra in the libraries, which we have not done to date. All the spectra in the libraries are associated with accurate mass data, and spectra annotated as unknowns would have a Level 5 confidence. Some of the spectra annotated as artifact/carryover could be originating from the sample and have a Level 5 confidence but finding these could be challenging and a method for doing this would require further development.

To provide additional detail about the confidence of both our direct and hybrid library MS/MS matches, we developed a workflow to assign a qualitative confidence level to each metabolite identification. The workflow starts with the match score and incorporates prior probability information about whether the identified compound has been previously observed as a metabolite. The workflow also incorporates the annotation as described above to ensure the identified spectrum originates from the sample. Match scoring performed by the pipeline is well documented in the literature for the NIST Tandem MS library and for the NIST MS Search program and is based upon the dot product of the spectra being compared (S. E. Stein, 1999). The match score has been validated by manual inspection of matches and correlates very well with the match quality as determined by visual inspection. A score cut-off of 400 removes essentially all poor matches and has been chosen as the default cutoff for metabolites. To assign confidence, an identification can initially be classified as high, medium, or low confidence, depending on the match score. Scores of 400–599, 600–799, and 800–999 correspond to low, medium, and high confidence, respectively. Of course, in cases of isomers with similar spectra, distinguishing them may not be possible without the use of reference standards. Prior probability as well the spectrum annotation can be used to raise or lower the qualitative level of confidence and can, to some degree, assist in isomer identification. Figure 5 depicts the workflow that was developed for assigning confidence. The workflow starting with a medium confidence is depicted at the top of the figure and is described below. The workflow starting with a low or high confidence is depicted at the bottom of the figure with the differences from the medium highlighted in green. The first step in the workflow is to determine if the identified compound is a known metabolite. For this study, we performed a literature search for reported CHO cell metabolites and searched the Human Metabolome Database

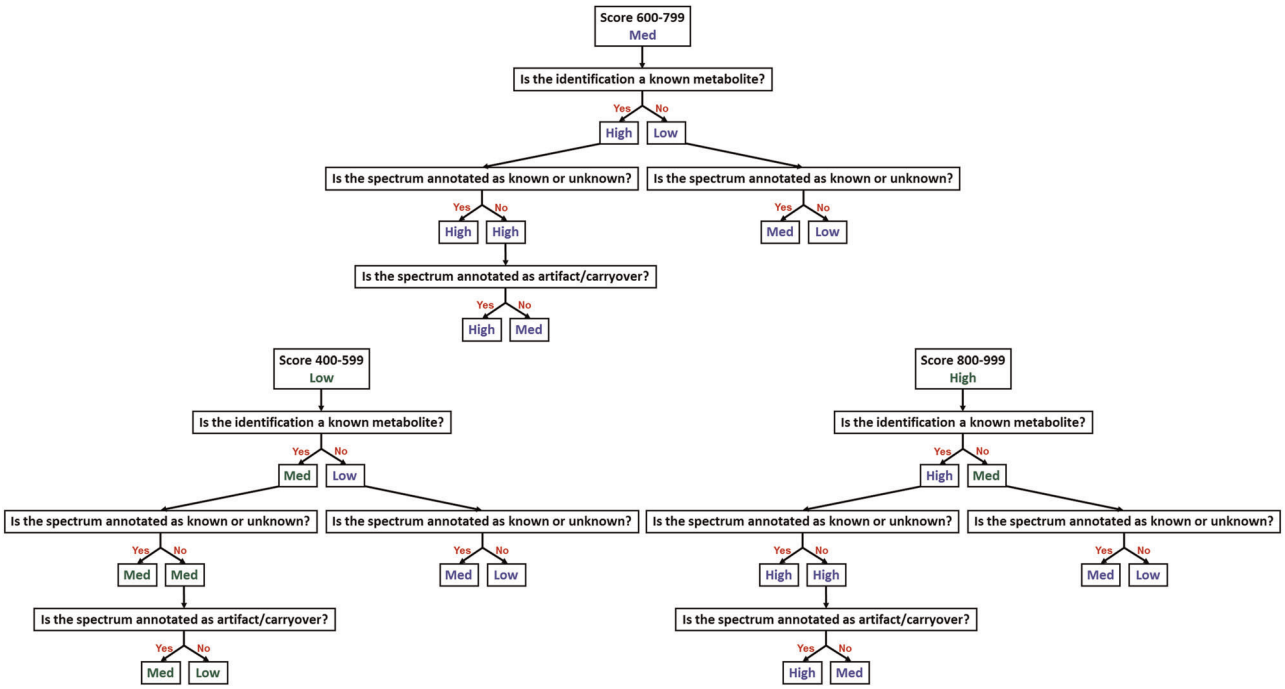

FIGURE 5 Workflow for assigning confidence in MS/MS identifications. Initial confidence level is determined by the match score and initial medium confidence is shown at the top. MS, mass spectrometry [Color figure can be viewed at wileyonlinelibrary.com]

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(HMDB; Wishart et al., 2013, 2018) and/or PubChem (Kim et al., 2019) to see if the compound was a reported human metabolite (it was not considered a metabolite if it was on HMDB, but not endogenous). In addition, we searched for lipids using the LIPID MAPS structure database (Sud et al., 2007). If it was found in any of these places, the qualitative confidence level was increased and if not, it was decreased. For the initial confidence of medium, an identification was elevated to high confidence if it was a known metabolite and lowered to low confidence if it was not. The next step is determining if the spectrum is annotated as a known/unknown. For the right side of the workflow, if the spectrum is not a known/unknown, confidence remains low and if it is, confidence is elevated to medium. On the left side of the workflow, if the spectrum is a known/unknown, confidence remains high and if it is not, it is determined if the spectrum is annotated as an artifact/carryover. If the spectrum is an artifact/carryover, then confidence remains high, and if it is not, confidence is lowered to medium. Confidence is only elevated once in the workflow to prevent a match with a low score from being elevated to high confidence. Table S5 shows the 20 most abundant identified ions from the data in Figure 1 and their associated confidence.

3.7 | Automation

One of the goals of this study was to develop tools that could be automated after initial development. The two workflows for annotation of spectra in the library and assignment of a qualitative confidence level for library identifications are amenable to automation via development of software tools. This will drastically increase the speed at which annotation and confidence assignment can occur. In addition, development of software tools for assessing prior probability and tools for automatic detection of spectra that are likely to be originating from the Pluronic F-68 in the cell culture media will be beneficial. However, expert evaluation of the output of any developed software tools will be required until the methods become routine.

4 | CONCLUSIONS

We have created the first recurrent spectral library for use in identifying CHO cell metabolites and outlined a procedure for future extensions. The library contains metabolites originating from a single CHO cell variety in a single cell culture media and represents the spectra of all compounds repeatedly observed in these samples and can be used as a tool by others in the field to quickly identify compounds in a CHO cell metabolite sample. During this analysis, we have developed a method capable of identifying all components commonly found in the LC-MS analysis of CHO cell metabolite extracts and media. An extension of this approach is expected to lead to both an automated way to extend this library and to develop similar libraries for other metabolite materials. Finally, we developed a strategy to assign qualitative confidence to NIST MS/MS library identifications. Although methods of representing the confidence of measurement have been developed for reporting individual metabolite identifications, this scheme could not adequately represent the confidence needed to properly annotate the identification made here—many of which cannot be regarded as definitive. The next step for this project will be automation of the workflows and release of the recurrent spectral libraries. The libraries can then be used in metabolomics studies of CHO cell metabolites using LC-MS/MS analyses.

ACKNOWLEDGMENT

The authors thank Dr. Arun Moorthy of the NIST MSDC for aiding in writing the pseudocode for hit list sorting. The authors also thank Dr. Tytus Mak of the NIST MSDC as well as Dr. Michael J. Betenbaugh, Harnish Mukesh Naik, and Venkata Gayatri Dhara from Johns Hopkins University for discussions regarding the CHO cell metabolite analysis project.

AUTHOR CONTRIBUTIONS

Kelly H. Telu and Stephen E. Stein contributed intellectually to project conceptualization and experiment design. Renae J. Preston and Lila Kashi grew the CHO cells used in the experiments. Zvi Kelman supervised CHO cell growth. Kelly H. Telu, Ramesh Marupaka, and Nirina R. Andriamaharavo performed the experiments. Kelly H. Telu, Yamil Simón-Manso, and Yuxue Liang contributed to LC-MS/MS method development. Yuri A. Mirokhin developed the algorithm Tallat H. Bukhari used to create the recurrent spectral libraries. The manuscript was drafted by Kelly H. Telu, revised by Stephen E. Stein, and then critiqued and approved by all co-authors. Stephen E. Stein supervised the project.

DATA AVAILABILITY STATEMENT

All programs and data are freely available at chemdata@nist.gov or via request to the corresponding author.

REFERENCES

Blaženović, I., Kind, T., Sa, M. R., Ji, J., Vaniya, A., Wancewicz, B., Roberts, B. S., Torbašinović, H., Lee, T., Mehta, S. S., Showalter, M. R., Song, H., Kwok, J., Jahn, D., Kim, J., & Fiehn, O. (2019). Structure annotation of all mass spectra in untargeted metabolomics. *Analytical Chemistry*, *91*(3), 2155–2162. https://doi.org/10.1021/acs.analchem.8b04698

Broeckling, C. D., Afsar, F. A., Neumann, S., Ben-Hur, A., & Prenni, J. E. (2014). RAMClust: A novel feature clustering method enables spectral-matching-based annotation for metabolomics data. *Analytical Chemistry*, *86*(14), 6812–6817. https://doi.org/10.1021/ac501530d

Burke, M. C., Mirokhin, Y. A., Tchekhovskoi, D. V., Markey, S. P., Heidbrink Thompson, J., Larkin, C., & Stein, S. E. (2017). The hybrid search: A mass spectral library search method for discovery of modifications in proteomics. *Journal of Proteome Research*, *16*(5), 1924–1935. https://doi.org/10.1021/acs.jproteome.6b00988

Caesar, L. K., Kvalheim, O. M., & Cech, N. B. (2018). Hierarchical cluster analysis of technical replicates to identify interferents in untargeted mass spectrometry metabolomics. *Analytica Chimica Acta*, *1021*, 69–77. http://doi.org/10.1016/j.aca.2018.03.013

Cleary, J. L., Luu, G. T., Pierce, E. C., Dutton, R. J., & Sanchez, L. M. (2019). BLANKA: An Algorithm for blank subtraction in mass spectrometry of

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

complex biological samples. *Journal of the American Society for Mass Spectrometry*, *30*(8), 1426–1434.

Cooper, B. T., Yan, X., Simon-Manso, Y., Tchekhovskoi, D. V., Mirokhin, Y. A., & Stein, S. E. (2019). Hybrid search: A method for identifying metabolites absent from tandem mass spectrometry libraries. *Analytical Chemistry*, *91*(21), 13924–13932. https://doi.org/10.1021/acs.analchem.9b03415

Creek, D. J., Dunn, W. B., Fiehn, O., Griffin, J. L., Hall, R. D., Lei, Z., Mistrik, R., Neumann, S., Schymanski, E. L., Sumner, L. W., Trengove, R., & Wolfender, J. L. (2014). Metabolite identification: Are you sure? And how do your peers gauge your confidence? *Metabolomics*, *10*(3), 350–353. https://doi.org/10.1007/s11306-014-0656-8

Dietmair, S., Hodson, M. P., Quek, L. E., Timmins, N. E., Chrysanthopoulos, P., Jacob, S. S., Gray, P., & Nielsen, L. K. (2012). Metabolite profiling of CHO cells with different growth characteristics. *Biotechnology and Bioengineering*, *109*(6), 1404–1414. https://doi.org/10.1002/bit.24496

Dietmair, S., Timmins, N. E., Gray, P. P., Nielsen, L. K., & Kromer, J. O. (2010). Towards quantitative metabolomics of mammalian cells: Development of a metabolite extraction protocol. *Analytical Biochemistry*, *404*(2), 155–164. https://doi.org/10.1016/j.ab.2010.04.031

Dong, Q., Liang, Y., Yan, X., Markey, S. P., Mirokhin, Y. A., Tchekhovskoi, D. V., Bukhari, T. H., & Stein, S. E. (2018). The NISTmAb tryptic peptide spectral library for monoclonal antibody characterization. *mAbs*, *10*(3), 354–369. https://doi.org/10.1080/19420862.2018.1436921

Dong, Q., Yan, X., Liang, Y., & Stein, S. E. (2016). In-Depth characterization and spectral library building of glycopeptides in the tryptic digest of a monoclonal antibody using 1D and 2D LC-MS/MS. *Journal of Proteome Research*, *15*(5), 1472–1486. https://doi.org/10.1021/acs.jproteome.5b01046

Gowda, G. A., & Djukovic, D. (2014). Overview of mass spectrometry-based metabolomics: Opportunities and challenges. *Methods in Molecular Biology*, *1198*, 3–12.

Ho, T. J., Kuo, C. H., Wang, S. Y., Chen, G. Y., & Tseng, Y. J. (2013). True ion pick (TIPick): A denoising and peak picking algorithm to extract ion signals from liquid chromatography/mass spectrometry data. *Journal of Mass Spectrometry*, *48*(2), 234–242. https://doi.org/10.1002/jms.3154

de Jong, F. A., & Beecher, C. (2012). Addressing the current bottlenecks of metabolomics: Isotopic ratio outlier analysis, an isotopic-labeling technique for accurate biochemical profiling. *Bioanalysis*, *4*(18), 2303–2314. https://doi.org/10.4155/bio.12.202

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., & Bolton, E. E. (2019). PubChem 2019 update: Improved access to chemical data. *Nucleic Acids Research*, *47*(D1), D1102–D1109. https://doi.org/10.1093/nar/gky1033

Kind, T., Tsugawa, H., Cajka, T., Ma, Y., Lai, Z., Mehta, S. S., Wohlgemuth, G., Barupal, D. K., Showalter, M. R., Arita, M., & Fiehn, O. (2018). Identification of small molecules using accurate mass MS/MS search. *Mass Spectrometry Reviews*, *37*(4), 513–532. https://doi.org/10.1002/mas.21535

Kunert, R., & Reinhart, D. (2016). Advances in recombinant antibody manufacturing. *Applied Microbiology and Biotechnology*, *100*(8), 3451–3461. https://doi.org/10.1007/s00253-016-7388-9

Mahieu, N. G., Huang, X., Chen, Y. J., & Patti, G. J. (2014). Credentialing features: A platform to benchmark and optimize untargeted metabolomic methods. *Analytical Chemistry*, *86*(19), 9583–9589. https://doi.org/10.1021/ac503092d

Mallard, W. G., Andriamaharavo, N. R., Mirokhin, Y. A., Halket, J. M., & Stein, S. E. (2014). Creation of libraries of recurring mass spectra from large data sets assisted by a dual-column workflow. *Analytical Chemistry*, *86*(20), 10231–10238. https://doi.org/10.1021/ac502379x

Matyash, V., Liebisch, G., Kurzchalia, T. V., Shevchenko, A., & Schwudke, D. (2008). Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *Journal of Lipid Research*, *49*(5), 1137–1146. https://doi.org/10.1194/jlr.D700041-JLR200

Mohmad-Saberi, S. E., Hashim, Y. Z., Mel, M., Amid, A., Ahmad-Raus, R., & Packeer-Mohamed, V. (2013). Metabolomics profiling of extracellular metabolites in CHO-K1 cells cultured in different types of growth media. *Cytotechnology*, *65*(4), 577–586. https://doi.org/10.1007/s10616-012-9508-4

Moorthy, A. S., Wallace, W. E., Kearsley, A. J., Tchekhovskoi, D. V., & Stein, S. E. (2017). Combining fragment-ion and neutral-loss matching during mass spectral library searching: A new general purpose algorithm applicable to illicit drug identification. *Analytical Chemistry*, *89*(24), 13261–13268. https://doi.org/10.1021/acs.analchem.7b03320

Remoroza, C. A., Mak, T. D., De Leoz, M. L. A., Mirokhin, Y. A., & Stein, S. E. (2018). Creating a mass spectral reference library for oligosaccharides in human milk. *Analytical Chemistry*, *90*(15), 8977–8988. https://doi.org/10.1021/acs.analchem.8b01176

Rudnick, P. A., Clauser, K. R., Kilpatrick, L. E., Tchekhovskoi, D. V., Neta, P., Blonder, N., Billheimer, D. D., Blackman, R. K., Bunk, D. M., Cardasis, H. L., Ham, A. J. L., Jaffe, J. D., Kinsinger, C. R., Mesri, M., Neubert, T. A., Schilling, B., Tabb, D. L., Tegeler, T. J., Vega-Montoto, L., … Stein, S. E. (2010). Performance metrics for liquid chromatography-tandem mass spectrometry systems in proteomics analyses. *Molecular & Cellular Proteomics*, *9*(2), 225–241. https://doi.org/10.1074/mcp.M900223-MCP200

Schrimpe-Rutledge, A. C., Codreanu, S. G., Sherrod, S. D., & McLean, J. A. (2016). Untargeted metabolomics strategies-challenges and emerging directions. *Journal of the American Society for Mass Spectrometry*, *27*(12), 1897–1905. https://doi.org/10.1007/s13361-016-1469-y

Schymanski, E. L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H. P., & Hollender, J. (2014). Identifying small molecules via high resolution mass spectrometry: Communicating confidence. *Environmental Science and Technology*, *48*(4), 2097–2098. https://doi.org/10.1021/es5002105

Sellick, C. A., Hansen, R., Stephens, G. M., Goodacre, R., & Dickson, A. J. (2011). Metabolite extraction from suspension-cultured mammalian cells for global metabolite profiling. *Nature Protocols*, *6*(8), 1241–1249. https://doi.org/10.1038/nprot.2011.366

Simon-Manso, Y., Marupaka, R., Yan, X., Liang, Y., Telu, K. H., Mirokhin, Y., & Stein, S. E. (2019). Mass spectrometry fingerprints of small-molecule metabolites in biofluids: Building a spectral library of recurrent spectra for urine analysis. *Analytical Chemistry*, *91*, 12021–12029. https://doi.org/10.1021/acs.analchem.9b02977

Simón-Manso, Y., Lowenthal, M. S., Kilpatrick, L. E., Sampson, M. L., Telu, K. H., Rudnick, P. A., Mallard, W. G., Bearden, D. W., Schock, T. B., Tchekhovskoi, D. V., Blonder, N., Yan, X., Liang, Y., Zheng, Y., Wallace, W. E., Neta, P., Phinney, K. W., Remaley, A. T., & Stein, S. E. (2013). Metabolite profiling of a NIST standard reference material for human plasma (SRM 1950): GC-MS, LC-MS, NMR, and clinical laboratory analyses, libraries, and web-based resources. *Analytical Chemistry*, *85*(24), 11725–11731. https://doi.org/10.1021/ac402503m

Sindelar, M., & Patti, G. J. (2020). Chemical discovery in the era of metabolomics. *Journal of the American Chemical Society*, *142*(20), 9097–9105. https://doi.org/10.1021/jacs.9b13198

Smith, C. A., Want, E. J., O'Maille, G., Abagyan, R., & Siuzdak, G. (2006). XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Analytical Chemistry*, *78*(3), 779–787. https://doi.org/10.1021/ac051437y

Spicer, R., Salek, R. M., Moreno, P., Canueto, D., & Steinbeck, C. (2017). Navigating freely-available software tools for metabolomics analysis. *Metabolomics*, *13*(9), 106. https://doi.org/10.1007/s11306-017-1242-7

Stein, S. (2012). Mass spectral reference libraries: An ever-expanding resource for chemical identification. *Analytical Chemistry*, *84*(17), 7274–7282. https://doi.org/10.1021/ac301205z

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Stein, S. E. (1999). An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. *Journal of the American Society for Mass Spectrometry*, *10*(8), 770–781. https://doi.org/10.1016/S1044-0305(99)00047-1

Stolfa, G., Smonskey, M. T., Boniface, R., Hachmann, A. B., Gulde, P., Joshi, A. D., Pierce, A. P., Jacobia, S. J., & Campbell, A. (2018). CHO-omics review: The impact of current and emerging technologies on Chinese hamster ovary based bioproduction. *Biotechnology Journal*, *13*(3), e1700227. https://doi.org/10.1002/biot.201700227

Sud, M., Fahy, E., Cotter, D., Brown, A., Dennis, E. A., Glass, C. K., & Subramaniam, S. (2007). LMSD: LIPID MAPS structure database. *Nucleic Acids Research*, *35*(Database issue), D527–D532. https://doi.org/10.1093/nar/gkl838

Sumner, L. W., Amberg, A., Barrett, D., Beale, M. H., Beger, R., Daykin, C. A., Fan, T. W. M., Fiehn, O., Goodacre, R., Griffin, J. L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A. N., Lindon, J. C., Marriott, P., Nicholls, A. W., … Viant, M. R. (2007). Proposed minimum reporting standards for Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*, *3*(3), 211–221. https://doi.org/10.1007/s11306-007-0082-2

Sumner, L. W., Lei, Z. T., Nikolau, B. J., Saito, K., Roessner, U., & Trengove, R. (2014). Proposed quantitative and alphanumeric metabolite identification metrics. *Metabolomics*, *10*(6), 1047–1049. https://doi.org/10.1007/s11306-014-0739-6

Telu, K. H., Yan, X., Wallace, W. E., Stein, S. E., & Simon-Manso, Y. (2016). Analysis of human plasma metabolites across different liquid chromatography/mass spectrometry platforms: Cross-platform transferable chemical signatures. *Rapid Communications in Mass Spectrometry*, *30*(5), 581–593. https://doi.org/10.1002/rcm.7475

Viant, M. R., Kurland, I. J., Jones, M. R., & Dunn, W. B. (2017). How close are we to complete annotation of metabolomes? *Current Opinion in Chemical Biology*, *36*, 64–69. https://doi.org/10.1016/j.cbpa.2017.01.001

Wishart, D. S., Feunang, Y. D., Marcu, A., Guo, A. C., Liang, K., Vázquez-Fresno, R., Sajed, T., Johnson, D., Li, C., Karu, N., Sayeeda, Z., Lo, E., Assempour, N., Berjanskii, M., Singhal, S., Arndt, D., Liang, Y., Badran, H., Grant, J., … Scalbert, A. (2018). HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Research*, *46*(D1), D608–D617. https://doi.org/10.1093/nar/gkx1089

Wishart, D. S., Jewison, T., Guo, A. C., Wilson, M., Knox, C., Liu, Y., Djoumbou, Y., Mandal, R., Aziat, F., Dong, E., Bouatra, S., Sinelnikov, I., Arndt, D., Xia, J., Liu, P., Yallou, F., Bjorndahl, T., Perez-Pineiro, R., Eisner, R., … Scalbert, A. (2013). HMDB 3.0—The human metabolome database in 2013. *Nucleic Acids Research*, *41*, D801–D807. https://doi.org/10.1093/nar/gks1065

Yang, X., Neta, P., & Stein, S. E. (2014). Quality control for building libraries from electrospray ionization tandem mass spectra. *Analytical Chemistry*, *86*(13), 6393–6400. https://doi.org/10.1021/ac500711m

Yang, X., Neta, P., & Stein, S. E. (2017). Extending a tandem mass spectral library to include MS(2) spectra of fragment ions produced in-source and MS(n) spectra. *Journal of the American Society for Mass Spectrometry*, *28*(11), 2280–2287. https://doi.org/10.1007/s13361-017-1748-2

Zhang, H., & Yang, Y. (2008). An algorithm for thorough background subtraction from high-resolution LC/MS data: Application for detection of glutathione-trapped reactive metabolites. *Journal of Mass Spectrometry*, *43*(9), 1181–1190. https://doi.org/10.1002/jms.1390

Zhu, P., Ding, W., Tong, W., Ghosal, A., Alton, K., & Chowdhury, S. (2009). A retention-time-shift-tolerant background subtraction and noise reduction algorithm (BgS-NoRA) for extraction of drug metabolites in liquid chromatography/mass spectrometry data from biological matrices. *Rapid Communications in Mass Spectrometry*, *23*(11), 1563–1572. https://doi.org/10.1002/rcm.4041

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Telu KH, Marupaka R, Andriamaharavo NR, et al. Creation and filtering of a recurrent spectral library of CHO cell metabolites and media components. *Biotechnology and Bioengineering*. 2021;118:1491–1510. https://doi.org/10.1002/bit.27661

10970290, 2021, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/bit.27661 by National Institute Of Standard & Technology (NIST), Wiley Online Library on [10/03/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License