In silico MS/MS spectra for identifying unknowns: A critical examination using CFM-ID algorithms and ENTACT mixture samples

Journal:	*Analytical and Bioanalytical Chemistry*
Manuscript ID	ABC-01909-2019.R1
Type of Paper:	Research Paper
Date Submitted by the Author:	n/a
Complete List of Authors:	Chao, Alex; Oak Ridge Institute for Science and Education (ORISE) Participant; Student Contractor, U.S. Environmental Protection Agency Al-Ghoul, Hussein; Oak Ridge Institute for Science and Education (ORISE) Participant; Student Contractor, U.S. Environmental Protection Agency McEachran, Andrew; Oak Ridge Institute for Science and Education (ORISE) Participant; Agilent Technologies Balabin, Ilya; General Dynamics Information Technology Transue, Tom; General Dynamics Information Technology Cathey, Tommy; General Dynamics Information Technology Grossman, Jarod; Oak Ridge Institute for Science and Education (ORISE) Participant, 109 T.W. Alexander Drive, Research Triangle Park, NC 27711; Agilent Technologies Singh, Randolph; Oak Ridge Institute for Science and Education, US Environmental Protection Agency; University of Luxembourg, Luxembourg Centre for Systems Biomedicine Ulrich, Elin; United States Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory Williams, Antony; United States Environmental Protection Agency, Office of Research and Development, National Center for Computational Toxicology Sobus, Jon; United States Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory
Keywords:	Non-targeted analysis, High-resolution mass spectrometry, CFM-ID, ENTACT, ToxCast, DSSTox

November 26, 2019

Re: Submission of Research Paper

Dear Dr. Nicola Oberbeckmann-Winter and Prof. Gérard Hopfgartner,

The manuscript entitled "*In silico* MS/MS spectra for identifying unknowns: A critical examination using CFM-ID algorithms and ENTACT mixture samples" is hereby resubmitted for consideration as Research Paper in ***Analytical and Bioanalytical Chemistry***.

We thank the reviewers for their thorough and thoughtful critique of our initial submission. We have now addressed all reviewer comments and prepared a revised version of the manuscript using track changes. Individual responses to each of the reviewers' comments, as well as specific edits to the manuscript, can be found in the following pages.

We assure you that this manuscript is an original work, that is has not been previously published whole or in part, and that it is not currently under consideration for publication elsewhere. All authors have read the manuscript, agree that it is ready for submission, and accept responsibility for the manuscript's contents. We declare that there are no conflicts of interest.

Thank you for your consideration of this revised manuscript. Please do not hesitate to contact us for any further information.

Sincerely,

Alex Chao
ORD/CCTE/CCED/AACMB
Oak Ridge Institute for Science and Education
Research Triangle Park, NC 27711
chao.alex@epa.gov

****Please consider the following comments.

Editor's comment:
This manuscript requires major revision, as per the reviewers' comments prior to consideration for publication.

Response: We thank the Editor for the opportunity to respond to reviewer comments and submit a revised manuscript.

Referees' comments

Reviewer A:

1) Chao et al., present an interesting approach to use Experimental spectra from EPA's Non-Targeted Analysis Collaborative Trial (ENTACT), in order to test the performance of Competitive Fragmentation Modeling-ID, for generation and query of in silico spectral libraries. The authors use a combination of candidate filtering and score aggregation that results in higher sensitivity and specificity, and point to the advantageous use of in silico libraries in combination with experimental spectral libraries. The paper is clear and well written, however, there are some points that should be addressed before the publication, to improve clarity, and to ensure that the community can be benefited from data and code generated in this project.

Response: We thank the reviewer for their complimentary review of the manuscript and provide detailed responses to specific questions below.

2) Are the ENTACT, 1269 unique substances present in public spectral libraries (MassBank or GNPS)? With different collision energies? I believe it should be, as it would benefit the community.

Response: We thank the reviewer for highlighting this important issue. During the initial planning stages of ENTACT, we anticipated a need for the generation and public release of MS2 data for the ENTACT substances. As such, we provided each ENTACT substance, individually plated on multi-well plates, to several participants for the purpose of generating and distributing reference MS data. Our study design article (https://link.springer.com/article/10.1007%2Fs00216-018-1435-6) describes these activities in detail, stating that "*Perhaps the largest need of the NTA research community, as articulated by the workshop attendees, is quality reference spectra for high-interest compounds. When included in reference libraries, experimental spectra can enable broad and accurate suspect screening. They can further function as a training set when building spectra prediction models (e.g., Competitive Fragmentation Modeling for Metabolite Identification, CFM-ID: http://cfmid.wishartlab.com/). Shortly after the 2015 workshop, EPA decided to make all ToxCast substances available to a group of vendors and software developers to facilitate the development of reference libraries and NTA tools. Furthermore, EPA has the ability to provide a subset of the individual ToxCast substances (i.e., those included in the ten synthetic mixtures) to a limited number of labs participating in ENTACT. These materials enable the generation of reference data (e.g., MS2 spectra, method specific retention times, collision cross sections), and are intended to facilitate rigorous self-evaluation of NTA results for ENTACT mixtures and spiked samples*." At present, we have received MS2 data generated on ENTACT compounds across several instrument platforms and methods. We are working to make these data

available to the public in a readily usable format but have no time table for initial release. We do anticipate eventual deposition of these data into public spectral libraries, such as MassBank and GNPS.

3) Is the python code described available? Where will be the prototype search tool be available? Is it possible to make it available during the development? The code is very important to the reproducibility of the work, and the availability of a prototype would add value to the work, as an additional contribution.

Response: We completely agree with the reviewer that the availability of the code is desirable to ensure reproducibility of the work described here, as well as future work. As such, the python code used for both the searching/scoring of the experimental data, as well as processing of the search results, has been uploaded to a GitHub repository with associated documentation (https://github.com/NTA-Code/cfmid). This code is further being incorporated into a Web application for public use, with a release date yet to be determined. The prototype search and visualization tool is scheduled for release in Spring 2020.

4) The authors discussed the different performance of in silico tools for different datasets, could that be related to the training set?

Response: The specific training sets used to develop fragmentation models will affect the predictions and subsequent performance of those predictions. Ideally, a given *in silico* tool would be trained with a dataset large enough to capture an extremely diverse chemical space, leading to predictable and acceptable performance in any application. We note that some *in silico* prediction tools (described in the Introduction section), such as rule-based fragmentation predictors, do not get "trained" but rather operate based on manually specified criteria. Here performance is less dependent on a specific training set, but still affected by existing knowledge of fragmentation behavior for known compounds.

5) 'CFM-ID prediction source code (http://sourceforge.net/projects/cfm-id) with pre-trained parameters.' The authors should discuss how CFM-ID was trained, and how a new training would impact the performance. In the good summary of the working principles of the in silico fragmentation tools, made by the authors, the fingerprint based tools, such as CSI:FingerID, are currently the best performing tools. There is evidence in the literature (PLoS One. 2011;6(12):e28966) that natural metabolites (unlike synthetic compounds, that should be abundant in DSSTox), differ from non-metabolites by distinct structural signatures. The fingerprint based tools seem to be able to capture these signatures, and do a better job classifying metabolites, would CFM-ID trained in a non-metabolite spectral library perform significantly better?

Response: The training of CFM-ID has been described in our earlier work (McEachran, Andrew D., et al. "Linking in silico MS/MS spectra with chemistry data to improve identification of unknowns." Scientific data 6.1 (2019): 1-9.) with a citation given in the current article. Substantive background information regarding CFM-ID is also available on the CFM-ID website and in published articles, with the website link/article references provided in our manuscript. We agree that a new training of CFM-ID could affect performance and have added the following text in the Discussion section to reflect this possibility: "*In certain cases, sub-optimal performance of CFM-ID may reflect dissimilarities in structures between compounds used to train CFM-ID and those included in ENTACT. A re-training of the CFM-ID models with*

***an expanded set of compounds has the potential to improve scoring and ranking results for the ENTACT mixture compounds*".**

We also agree that fingerprint-based tools, such as CSI:FingerID, can outperform CFM-ID. While we acknowledge that a re-training of CFM-ID could improve overall performance, we are very hesitant to speculate on the level of potential improvement. We are currently investigating this matter and feel that any demonstrable performance improvements, based on training sets of synthetic chemicals, are beyond the scope of the current article, which focuses exclusively on the performance of the existing CFM-ID (v2) model.

6) When the authors state: ' … and then processed using a custom script written in the Python programming language. Processing of MGF files was performed to improve data formatting and to de-duplicate MS2 spectra. Regarding de-duplication, any single chemical feature with an associated precursor mass may generate multiple MS2 spectra during acquisition.' One of the main concerns for this processing, for which there is no code available is, how is this useful for someone that wants to replicate it for their work?

Response: Based on the reviewer's suggestion, we have provided our data processing code and supporting documentation in a GitHub repository (https://github.com/NTA-Code/cfmid). We further provide the link to the source code within section 3.6 of the manuscript.

7) If the author should consider mentioning open source tools to easily export MS2 data (https://www.biorxiv.org/content/10.1101/812404v1.abstract) and a brief discussion of the use of the in house script versus more friendly implementations, such as with MZmine, should be included.

Response: For the purposes of this work, only a simple re-formatting of the exported .mgf files was necessary to make the MS2 data searchable. Thus, we wrote a basic script to perform this function as opposed to relying on an outside software package.

8) Page 6 – line 37 – True Positive Rates (TPR) and False Positive Rates (FPR) - why not call it Precision and Recall?

Response: To illustrate the performance of CFM-ID cut-off filters, Receiver Operating Characteristic curves (ROC) were plotted, which typically have True Positive Rates (TPR) and False Positive Rates (FPR) on the y- and x-axes, respectively. There are certainly alternate terms for both TPR (e.g., "recall" and "sensitivity") and FPR (e.g., "fall-out" and "1-specificity") that are more well-established in the -omics communities, but we believe that using "TPR" and "FPR" allows for the most direct interpretation by the widest audience. We note that ROC curves are not equivalent to Precision-Recall curves, which examine the relationship between the TPR (i.e., "Recall") and the positive predictive value (i.e., "Precision"). Importantly, true negative results are not considered in Precision-Recall curves. We believe these results to be of considerable importance when examining the performance of CFM-ID cut-off filters.

9) How the spectral search described in: "Agilent MassHunter Qualitative Analysis (version B.08) software with forward and reverse scoring thresholds of 0 and 20, respectively." Differs from the: "for all candidates against the experimental MS2 spectrum using a dot- product algorithm [28] with"

Response: The reference MS2 libraries used in this analysis are proprietary libraries created by Agilent, and the searching/scoring software for these libraries are proprietary algorithms within Agilent's MassHunter software. Forward and reverse scoring refers to minimum thresholds for forward and reverse scores to be considered as a match by Agilent's software. For CFM-ID spectra scoring, a cosine-dot product algorithm was used to compare experimental spectra to CFM-ID predicted spectra, generating a match score between 0 and 1, with 1 being the best match score. Since the reference libraries and vendor algorithms were proprietary, we were simply not able to identify reference library matches in the exact manner as the *in silico* library matches. We did, however, manually review all reference library matches to ensure a high-quality final match list.

10) Would be important to have the spectral library search and in silico search both done by freely available tools, one easy way to prepare spectral libraries is https://www.biorxiv.org/content/10.1101/804401v1.

Response: We completely agree that it would be ideal to have searching/matching procedures performed in the exact same manner using the reference and *in silico* MS2 libraries. In this instance, however, all reference spectra were maintained in a proprietary vendor library. Furthermore, proprietary matching algorithms were the basis for compound scoring based on reference spectra. Given the proprietary nature of the reference spectra and algorithms, we were not able to use identical approaches when matching against reference and *in silico* spectra. Importantly, we had no means of extracting reference spectra from the proprietary libraries for analysis using freely available tools. Nevertheless, our methods allowed for a rigorous evaluation of the number of compounds that could be identified using both vendor and *in silico* libraries.

11) In the discussion the authors should provide recommendations on how the findings could be used for future research, for instance, should other groups acquire spectra in three collision energies (there can be restrictions for practical applications) or just 20 V? Should the 20 V be compared to all 3 collision energies in the predicted spectra?

Response: Based on the reviewer's suggestion, we have added text in the Discussion section. First, regarding acquisition of experimental data, we state that: "*For an NTA workflow where the compounds are unknown, the recommended practice is to acquire experimental MS2 data at all three CE levels in order to capture suitable spectra on the widest range of compounds".* Second, regarding the use of predicted spectra, we state that: "*Moving forward, when using the CFM-ID database as a screening-level tool, we recommend an aggregated approach wherein each experimental spectrum is compared to all three CE levels of predicted spectra…*".

12) The authors state: 'In silico MS2 libraries are not meant to replace reference libraries. Rather, they are meant to supplement reference library matching procedures.' There are recent applications using this concept that should be cited (Anal Chem. 2016 Mar 15;88(6):3317-23.)

Response: We have added this citation, as well as citations to several other relevant works.

Reviewer B:

1) The main results of this paper are similar to what has been reported/published in the past for NTA: 1) If an unknown was not in a database it will not be identified. 2) If an unknown is in a database and it's experimental MS spectrum is also included in the database it will likely be identified. 3) If an unknown is in a small database (i.e. < 1x10^6 cpds) computational methods can be used to identify it approximately 35% of the time. 4) There is an optimum combination of collision energies that gives better results, and this typically happens when there are more peaks in the MS spectrum. 5) limiting searches to a specific formula improves results. 6) Utilizing results from in silico library searching is a balance between sensitivity (TPR) and specificity (FPR).

Response: We agree with the reviewer that our observations and results are consistent with those published in previous works. But we view this consistency very much as a positive. In other words, we would be concerned if our results were not in agreement with those of previous studies. The present work was not intended to pioneer a new algorithm or matching technique. Rather, it was intended as a rigorous evaluation of a comprehensive method applied to complex chemical mixtures developed as part of ENTACT; such rigorous testing is critical to support future NTA applications that rely on our methods, data, and tools. We remind the reviewer that the existing CFM-ID web application (http://cfmid.wishartlab.com/**) allows users to predict spectra, annotate spectral peaks, and identify compounds, but that searches cannot be performed in batches, and dataset queries are restricted to HMDB and KEGG. Given these limitations, we have extended the functionality of the web-based CFM-ID tool by enabling batch searching of experimental MS2 spectra against *in silico* MS2 spectra based on the entirety of the US EPA DSSTox Database (~765,000 substances at the time of analysis). This database is unique in that it focuses on environmentally relevant compounds that are of increasing interest to environmental health scientists and public health officials. Given the public release of the *in silico* spectra and data processing scripts, we felt it necessary to evaluate and communicate (through peer-reviewed publication) the performance of the methods and tools. We strongly believe that this article provides a strong foundation for guiding future applications of *in silico* spectra for DSSTox substances.**

2) The manuscript was difficult to review because figures were not labeled and there were no figure captions included.

Response: We sincerely apologize for this omission. We submitted the figure captions as instructed by ABC but didn't realize they were not included in the .pdf version of the submission. We have made sure that all Table and Figure captions are included in the revised submission. We again sincerely apologize for this oversight.

3) The authors do not compare their results with previous/similar studies. And they fail to emphasize/discuss the fact that their in silico identification results (unlike the reference MS2 spectral library results) are at Level 3 and thus of marginal value.

Response: We agree with the reviewer that a comparison of our results to those of similar studies is of value and have therefore added text and relevant citations to the Discussion section. Specifically, we now state that: "*Utilizing CFM-ID results from Approach 3, 34% of the 377 ENTACT mixture compounds were identified as the best matching compound. This result is comparable to those reported from the*

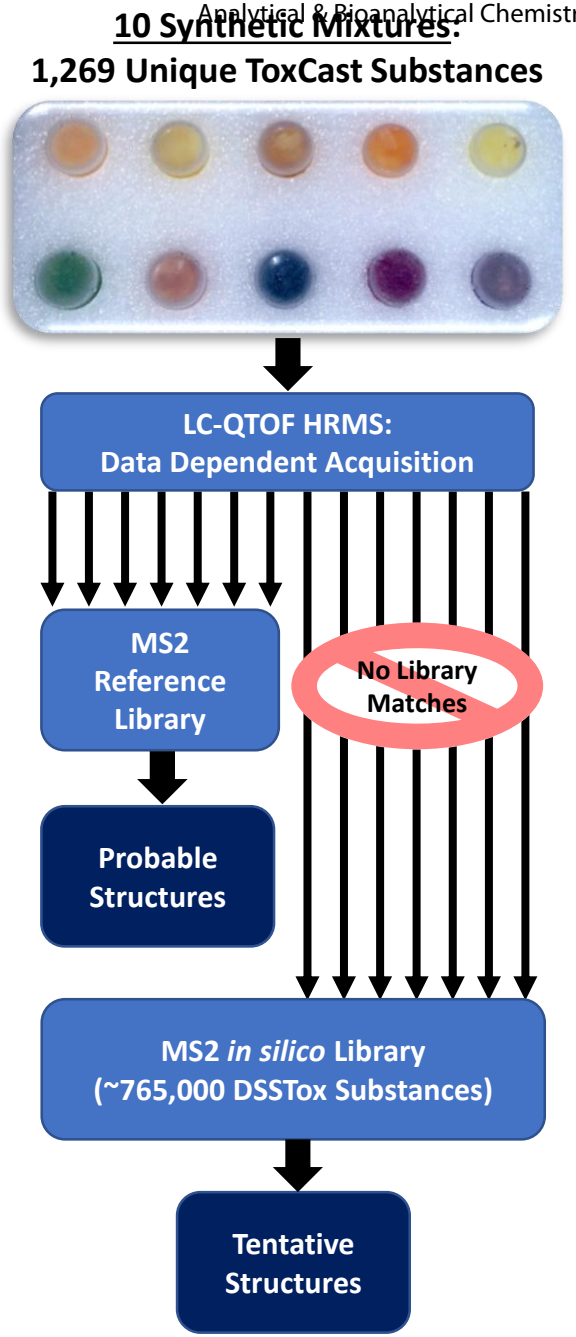
2016 CASMI contest, in which 12% to 34% of correct candidates were identified as the best matching compound*". We also agree that *in silico* identification results are less confident than those obtained via reference library matching. We have therefore added the following text to the Discussion section: *"In silico library matches are inherently less confident than reference library matches. As such, in silico MS2 libraries are not meant to replace reference libraries, but to enable supplementary matching procedures."

4) In tables 2 and 3 it is not clear what is meant by "Per Bin".

Response: Within tables 2 and 3, results for ENTACT mixture compounds were totaled based upon whether they were within the Top 1, Top 5, or Top 20 CFM-ID scoring results for each specified scoring approach. That is, the results were "binned" into the specified ranking ranges. Each row of results within the specified sections corresponds to a specific bin of ranks. We have removed "Per Bin" from Tables 2 and 3 to avoid unnecessary confusion.

5) The manuscript is well written, but contains results that are mostly already known from previous work.

Response: We kindly refer the reviewer to our response to comment #1 (from Reviewer B)


10 Synthetic Mixtures:
1,269 Unique ToxCast Substances
LC-QTOF HRMS:
Data Dependent Acquisition
MS2
Reference
Library
No Library
Matches
Probable
Structures
MS2 *in silico* Library
(~765,000 DSSTox Substances)
Tentative
Structures

In silico MS/MS spectra for identifying unknowns: A critical examination using CFM-ID algorithms and ENTACT mixture samples

Authors: Alex Chao[1,2]*, Hussein Al-Ghoul[1,2], Andrew D. McEachran[1,3], Ilya Balabin[4], Tom Transue[4], Tommy Cathey[4], Jarod N. Grossman[2,3], Randolph Singh[1,5], Elin M. Ulrich[6], Antony J. Williams[7] and Jon R. Sobus[6]*

Affiliations:

[1]Oak Ridge Institute for Science and Education (ORISE) Participant, 109 T.W. Alexander Drive, Research Triangle Park, NC 27711, USA

[2]Student Contractor, U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, 109 T.W. Alexander Drive, Research Triangle Park, NC 27711, USA

[3]Present address: Agilent Technologies Inc., Santa Clara, CA 95051, USA

[4]General Dynamics Information Technology, 79 T.W. Alexander Drive, Research Triangle Park, NC 27709

[5]Present address: Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Luxembourg

[6]U.S. Environmental Protection Agency, Office of Research and Development, National Exposure Research Laboratory, 109 T.W. Alexander Drive, Research Triangle Park, NC 27711, USA

[7]U.S. Environmental Protection Agency, Office of Research and Development, National Center for Computational Toxicology, 109 T.W. Alexander Drive, Research Triangle Park, NC 27711, USA

***Corresponding author(s):**

Alex Chao (chao.alex@epa.gov)
Jon Sobus (sobus.jon@epa.gov)

1.0 Abstract:

High-resolution mass spectrometry (HRMS) enables rapid chemical annotation via accurate mass measurements and matching of experimentally derived spectra with reference spectra. Reference libraries are generated from chemical standards and are therefore limited in size relative to known chemical space. To address this limitation, *in silico* spectra (i.e., MS/MS or MS2 spectra), predicted via Competitive Fragmentation Modeling-ID (CFM-ID) algorithms, were generated for compounds within the U.S. Environmental Protection Agency's (EPA) Distributed Structure-Searchable Toxicity (DSSTox) database (totaling, at the time of analysis, ~765,000 substances). Experimental spectra from EPA's Non-Targeted Analysis Collaborative Trial (ENTACT) mixtures (n=10) were then used to evaluate the performance of the *in silico* spectra. Overall, MS2 spectra were acquired for 377 unique compounds from the ENTACT mixtures. Approximately 53% of these compounds were correctly identified using a commercial reference library, whereas up to 50% were correctly identified as the top hit using the *in silico* library. Together, the reference and *in silico* libraries were able to correctly identify 73% of the 377 ENTACT substances. When using the *in silico* spectra for candidate filtering, an examination of binary classifiers showed a true positive rate (TPR) of 0.90 associated with false positive rates (FPRs) of 0.10 to 0.85, depending on the sample and method of candidate filtering. Taken together, these findings show the abilities of *in silico* spectra to correctly identify true positives in complex samples (at rates comparable to those observed with reference spectra), and efficiently filter large numbers of potential false positives from further consideration.

2.0 Introduction:

The exposome was originally conceived as the sum of all exposures encountered by an individual during their lifetime [1]. Despite more than ten years of dedicated research, the exposome is not well-characterized for individuals or populations, owing (in part) to a lack of suitable monitoring tools. Traditional exposure monitoring has relied on targeted analytical methods, developed and validated for specific high-interest compounds. These methods have generally proven impractical for exposome studies, where a goal is to characterize previously unknown compounds that may be of eventual interest. Time and resource limitations simply prohibit the development of enough targeted methods to cover the expanse of the exposome.

Advancements in analytical and computational technologies have enabled a shift from targeted monitoring methods to non-targeted analysis (NTA) methods. High-resolution mass spectrometers (HRMS), utilizing Orbitrap and quadrupole time-of-flight (Q-TOF) mass analyzers, now provide the combination of resolution, sensitivity, and speed needed to support NTA studies. Whereas targeted methods only monitor specific compounds during data acquisition, HRMS instruments generate data with sufficient quality that compound selection/identification can be performed at later stages of analysis, without reliance on pre-conceived chemical target lists. The confidence in eventual chemical identifications depends, in part, on the experimental HRMS data available for analysis. Accurate mass and isotope pattern data may enable chemical characterization at the molecular formula level, whereas tandem fragmentation data (i.e., MS/MS or MS2 spectra) may enable characterization at the structure level [2]. Highly confident identifications are generally those in which experimental MS2 data are matched to reference MS2 data contained within a well-curated library (with confirmation ultimately requiring use of a chemical standard). Numerous

reference libraries exist (e.g., mzCloud, MassBank, NIST) and enable confident identifications in NTA studies; these range from proprietary vendor-generated libraries to public repositories reflecting the collaborative efforts of many contributors. Recent reviews highlight the breadth of these MS2 reference libraries, which include spectra for up to tens-of-thousands of compounds [3-5]. Compared to chemical listings within ChemSpider and PubChem (numbering in the millions), however, these libraries cover only a small fraction of potential chemicals of interest [6, 7].

Chemical coverage within reference libraries is unlikely to change dramatically in the near future; the requirement for chemical synthesis followed by MS analysis is rate-limiting in the growth of said libraries. To address this challenge, researchers have turned to computational approaches, wherein computer-generated spectra (or fragment ions) are the basis for comparison against experimental data. Using these *in silico* approaches, library coverage is limited only by the size of the database from which the predictions are based.

A variety of approaches currently exist for spectra/fragment prediction and comparison. Approaches like MS-Finder and Mass Frontier use specific fragmentation rules to predict MS2 spectra for database compounds [8]. An inherent limitation of this approach is a bias towards compounds for which the known rules apply. Other approaches like MetFrag and MAGMA use combinatorial fragmentation. Here, rather than predicting spectra for a given compound, each bond of that compound is systematically broken *in silico* to yield possible molecular fragments. Experimental fragment ions are then matched against possible molecular fragment ions to generate a weighted score for that compound [9-11].

Molecular fingerprinting is another computational technique, and is being utilized by ChemDistiller and CSI:FingerID. With this approach, predictive analysis is performed on experimental data [12-14]. Specifically, fragment ions within an experimental spectrum are used to predict specific structural features (i.e., substructures) of the unknown compound, which together yield a "fingerprint" for that compound. The predicted fingerprint for the unknown compound is compared with discrete fingerprints for database compounds to yield a list of scored matches. Recent reviews highlight the merits and limitations of these computational approaches for the analysis of experimental MS2 data [3, 15, 16].

Competitive Fragmentation Modeling-ID (CFM-ID) is an approach wherein experimental MS2 spectra are searched and scored against predicted spectra based on similarity [17, 18]. CFM-ID algorithms are trained on experimental data and used to discover fragmentation rules and eventual predictive models for MS2 spectra. Relative to previously described computational approaches, CFM-ID exists in a middle ground; predicted spectra are more complex than those based on specific fragmentation rules, while avoiding the explosion of fragmentation possibilities from combinatorial methods. CFM-ID further predicts peak intensities, which can be incorporated into spectral similarity searches and match scores. The source code for CFM-ID is publicly available, allowing for incorporation into in-house databases. Predictions can thus be pre-processed on the entirety of a chemical database, reducing computational time during actual searching of experimental data.

With several computational approaches available, numerous performance comparisons have been conducted in recent years [11, 13, 17]. Unsurprisingly, results have varied from assessment to

assessment, as the tested datasets have differed from one study to the next. To address this challenge, the Critical Assessment of Small Molecule Identification (CASMI) contest was founded in 2012 with the goal of enabling a more accurate comparison between methods. For each CASMI contest, an MS-based data set of challenge compounds unknown to the participants was made publicly available for examination [19, 20]. Specifically, previously acquired MS2 spectra (with accompanying metadata, in some instances) for individual compounds was shared for blinded evaluation. Results for each completed contest year have been compiled and are available online (http://casmi-contest.org), along with the challenge data sets, allowing for additional testing of new/refined computational approaches.

The data sets and results available through CASMI are an excellent resource for evaluating specific computational tools and *in silico* libraries. Since the CASMI contests were focused on evaluating spectra of individual compounds, a logical extension is to consider many spectra from a complex mixture as part of a performance evaluation. Along these lines, EPA's Non-Targeted Analysis Collaborative Trial (ENTACT) was launched in 2016 to evaluate the current status and landscape of NTA approaches, from data acquisition through results processing, with a focus on xenobiotic compounds in complex mixtures [21, 22]. Ten ENTACT mixtures were ultimately prepared, encompassing over 1,200 chemical substances from EPA's Toxicity Forecaster (ToxCast) library, and sent to participating labs for analysis. Much like CASMI, participants were allowed freedom in the selection of NTA approaches. While initially blinded, labs were eventually informed of the contents of each mixture to enable self-evaluation.

Within EPA's Office of Research and Development (ORD), initial analysis of the ENTACT mixtures has been performed and results of self-evaluation reported [23]. The purpose of the current article is to describe the incorporation of CFM-ID predicted spectra into the existing EPA workflow, and to evaluate overall method performance using the ENTACT mixture data. CFM-ID was selected for this investigation given the availability of the source code and its documented performance in previous CASMI contests. This article describes: 1) workflows for processing and searching experimental MS2 spectra against CFM-ID predicted spectra; 2) approaches for utilizing CFM-ID search scores in NTA workflows; 3) assessment of CFM-ID performance on ENTACT mixture compounds; and 4) comparison of reference library performance vs. CFM-ID library performance. This analysis serves as the initial proof-of-concept for adding CFM-ID predictions to an established NTA workflow. Future analyses that utilize this addition will benefit from increased library coverage and enhanced confidence in compound identifications.

3.0 Methods:

Figure 1 displays the overall NTA workflow utilized in our analyses of the ENTACT mixtures. This workflow outlines the main components of data acquisition and processing (left), as well as database generation and matching (center). It further lists the confidence levels associated with each type of match (right). Our previously reported results for the ENTACT mixtures were based on matching feature data to mass lists, formula lists, and reference MS2 libraries (highlighted in blue) [23]. The current examination incorporates searching against CFM-ID predicted spectra (highlighted in purple).

3.1 Sample Preparation and Data Acquisition:

Sample preparation and analysis procedures have been previously described [23]. Briefly, a total of 1,269 unique substances were spiked across ten separate synthetic mixtures (labelled 499 through 508), with each mixture receiving between 95 and 365 substances. Each mixture was analyzed via liquid-chromatography/mass-spectrometry (LC/MS), utilizing an Agilent 1290 Infinity II LC coupled to an Agilent 6530B accurate-mass quadrupole-time-of-flight (Q-TOF) mass spectrometer with a Dual AJS ionization source. An Agilent ZORBAX Eclipse Plus C8 column (2.1 × 50 mm, 1.8 µm) was used along with mobile phases consisting of 0.4 mM ammonium formate buffer in water and methanol. MS1 and MS2 data were collected in a scan range of 100-1000 *m/z* in both positive and negative ionization mode. Reference solution consisting of purine, hexakis(1H,1H,3H-tetrafluoropropoxy)phosphazene, and trifluoroacetic acid (TFA) were infused into the source during the course of the run for auto-correction of mass drift. MS2 data were acquired using Auto MS2 acquisition with the following settings: 3 max precursors per cycle, minimum threshold 3000 counts, scan rate 4 spectra/second. MS2 exclusion lists were generated to exclude ions corresponding to the reference solution from selection for fragmentation. MS2 inclusion lists were generated to increase preference for ions corresponding to substances previously observed using MS1 data. Each sample was acquired three times to generate MS2 data, with each acquisition collecting at one of the three collision energy (CE) levels: 10, 20 or 40 V.

3.2 Chemical Substance Database:

EPA's Distributed Structure-Searchable Toxicity (DSSTox) Database is a public chemistry resource containing data on (at the time of analysis) ~765,000 chemical substances and serves as the foundation for EPA's CompTox Chemicals Dashboard, hereafter referred to as the Dashboard (https://comptox.epa.gov/dashboard) [24, 25]. Each chemical substance within DSSTox is identified by a unique DSSTox substance identifier (DTXSID) and is also mapped to a "MS-Ready" structure corresponding to the form that would be observed by MS analysis. "MS-Ready" structures are identified by DSSTox chemical identifiers (DTXCID) [26]. The entirety of the 1,269 unique ENTACT mixture substances are registered within DSSTox, with unique DTXSIDs and associated MS-Ready DTXCIDs.

3.3 Substance Selection for MS2 Matching:

In a previous analysis of the ENTACT mixtures, initial substance identification was performed without the use of individual reference standards. Thus, for any given spiked substance, determination of presence vs. absence could not be made with absolute certainty (i.e., Schymanski *et al*. Level 1) [23]. Features that could be linked to spiked substances with enough diagnostic evidence (e.g., MS1 and MS2 data corroborating an identification at the "probable structure" level [2]) were classified as "passes", indicating that there was strong evidence of their presence. The set of "pass" substances, spanning all ten mixtures, was the basis for all analyses in the current study. Specifically, these "pass" substances were first used to generate lists of expected monoisotopic masses, considering only $[M+H]^+$ and $[M-H]^-$ ion species for positive and negative ESI modes, respectively. These lists of expected masses were then searched (with a 10 ppm accuracy window) against MS2 precursor ion lists to identify "pass" substances for which MS2 data were acquired.

3.4 Reference Library Preparation:

Reference MS2 spectra were contained in Agilent Personal Compound Database and Library (PCDL) format. Six Agilent PCDLs (i.e., Environmental water screening, Pesticides, Forensic toxicology, Veterinary drugs, Metlin, and Extractables and leachables) were combined and used for the current analysis. Experimental MS2 data [23] were searched against the composite PCDL using Agilent MassHunter Qualitative Analysis (version B.08) software with forward and reverse scoring thresholds of 0 and 20, respectively. All matches were manually reviewed to increase confidence in compound identifications.

Compound information from each of the six PCDLs was exported using Agilent PCDL Manager software. Specifically, compound name, formula, mass, CAS number, and number of MS2 spectra were exported for all compounds in each PCDL. This list of compounds was filtered for those containing at least one MS2 spectrum, and then batch searched by CAS number on the Dashboard to retrieve a DTXSID for each compound in the PCDLs. MS-Ready DTXCIDs were then retrieved for each compound by querying a DSSTox MS-Ready mapping file. In some cases, a PCDL compound was not able to be mapped to a DTXSID/DTXCID, either due to the compound not being registered in DSSTox, or due to an incorrect CAS number preventing a mapping. PCDL compounds were compared against the ENTACT mixture compounds by MS-Ready DTXCID to estimate the approximate coverage of ENTACT mixture compounds within the searched PCDLs.

3.5 *In silico* Library Preparation:

In silico MS2 spectra were computed for the majority of MS-Ready structures in DSSTox using the publicly available CFM-ID 2.0 algorithms [17]. Predictions were based on electrospray ionization, in positive and negative modes, at three CE levels (10, 20, and 40 V). Briefly, SMILES strings for MS-Ready structures in DSSTox were input into the CFM-ID prediction source code (http://sourceforge.net/projects/cfm-id) with pre-trained parameters. Resulting predicted spectra were then linked with MS-Ready structure metadata such as DTXCID, molecular formula, and monoisotopic mass. The resulting database of CFM-ID predicted spectra is hereafter referred to as the "CFM-ID database" [27].

3.6 *In silico* Library Matching:

Figure S1 describes the workflow for searching ENTACT MS2 spectra against the CFM-ID database (Source code used for *in silico* library matching, scoring and processing of results is available at https://github.com/NTA-Code/cfmid). Acquired MS2 spectra were first exported from Agilent .d files in MGF format, and then processed using a custom script written in the Python programming language. Processing of MGF files was performed to improve data formatting and to de-duplicate MS2 spectra. Regarding de-duplication, any single chemical feature with an associated precursor mass may generate multiple MS2 spectra during acquisition. The spectrum with the highest signal was considered most representative of the chemical feature for spectral matching purposes. Thus, for a given precursor mass, the spectrum with the highest sum intensity of ions was retained for analysis. Once MS2 spectra were processed, the Python script searched the CFM-ID database for all candidate compounds (as identified by MS-Ready DTXCID) within

a 10-ppm mass window of each MS2 spectrum precursor mass, considering only $[M+H]^+$ and $[M-H]^-$ ion species for positive and negative mode, respectively. The Python script then scored predicted spectra (for CE 10, 20 and 40 V) for all candidates against the experimental MS2 spectrum using a dot-product algorithm [28] with a fragment mass window of 0.02 Da, with scores ranging from 0 to 1.

Once scores were generated for candidate compounds, three approaches for using the scores were evaluated (Figure 2). In approach 1, only the score of the CFM-ID spectrum with the same CE level as the experimental spectrum was used. In approach 2, scores for CFM-ID spectra at all three CE levels were summed as a new score. In approach 3, scores for CFM-ID spectra at all CE levels were summed as a new score, and these new scores were summed across all experimental CE levels. Scores from each approach were used to rank ENTACT mixture compounds against other candidate compounds for each MS2 spectrum. Scores were also used to generate percentile and quotient values for all candidate compounds, with quotient values defined as the score of the candidate compound divided by the maximum score among all candidate compounds for a given experimental MS2 spectrum.

Only MS2 spectra corresponding to "pass" ENTACT mixture compounds were evaluated by CFM-ID library matching. For each MS2 spectrum, the ENTACT mixture compound represents a true positive (TP) and the remaining candidate compounds represent potential false positives (FP). When a cut-off filter is applied to CFM-ID results based on either a percentile or quotient value, the ENTACT mixture compound is considered either a potential TP (if above the cut-off value) or a false negative (FN; if below the cut-off value). Other candidate compounds which are above the cut-off value are considered potential FPs, and those below the cut-off value are considered true negatives (TN). Examples of cut-off filtering of CFM-ID results are shown in Figure S2. True Positive Rates (TPR) and False Positive Rates (FPR) were calculated using the following equations:

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

To identify an optimal threshold for candidate filtering, cut-off values were incremented throughout the entire range by hundredths of the value range (i.e., percentile cutoffs were set to 0, 1, 2 … 100; quotient cutoffs were set to 0, 0.01, 0.02 … 1). At each level, TP, FP, TN, and FP counts were tallied and used to calculate TPR and FPR. Receiver operating characteristic (ROC) curves were then generated, using TPR and FPR values, for the global ENTACT data set (i.e., all ten mixtures). Using the global curves, the percentile value and quotient value that would result in a minimum TPR of 0.90 was determined. These global percentile and quotient cut-offs were applied to each ENTACT mixture's results to calculate the mixture-specific TPR and FPR based on the global cut-off. The mixture-specific TPRs and FPRs ultimately serve as performance metrics for the proposed methods.

Some NTA workflows base predicted library matching on monoisotopic mass queries, whereas others restrict the candidate compound set to those matching a specific formula (deduced from MS1 spectra or other orthogonal methods). All procedures described in section 3.6 were performed separately based either on monoisotopic mass queries or mass queries followed by formula filtering (where the MS-Ready formula of all candidates was forced to match that of the "pass" substance). It is noteworthy that, for this investigation of ENTACT mixtures, a single formula was previously assigned to each "pass" substance with a high level of confidence. Formula assignments for features in true unknown samples are subject to considerably larger error rates. Thus, results of our formula-based analysis represent a "best case scenario" and yield the smallest expected FPRs. Nevertheless, comparison of results based on mass vs. formula queries will help establish best practices and performance targets for predicted library matching protocols.

4.0 Results:

4.1 Reference Library Matching:

For a given ENTACT compound, identification via reference library matching requires that the compound is ionizable (given the experimental source conditions), selected for MS2 acquisition, and present in the reference library. As described above, our previous analysis of the ENTACT mixtures yielded a list of "pass" substances that were identified with sufficient diagnostic evidence; this list of substances (Table S1) represents the starting point for the current evaluation. It is noteworthy that certain substances were included in multiple mixtures as part of the ENTACT design to help evaluate method reproducibility [21, 23]. For the purposes of this analysis, the focus of which was to evaluate performance of *in silico* library matching across a broad range of substances, each substance was ultimately evaluated only once even if it was acquired in multiple mixtures. Initial results (*vide infra*), however, are provided without deduplication to preserve statistics specific to each individual ENTACT mixture.

Overall, 44% of spiked ENTACT substances were classified with a "pass" rating (Table 1). Certain ENTACT mixtures (e.g., 507 and 508) had a very low proportion of "pass" compounds owing, in part, to a high number of spiked isomers that could not be resolved even with MS2 data. Out of 845 total "pass" compounds, 500 (59%) were included in the composite PCDL (including reference MS2 data), 453 (54%) had acquired MS2 data, and 300 (36%) had both reference and acquired MS2 data (Table 1). Ultimately, 246 of these 300 "pass" compounds were correctly identified with a Level 2a designation [2]. Thus, an 82% success rate was observed when considering "pass" compounds with both experimental and reference MS2 data (n=300). A 54% success rate, however, was observed when considering all "pass" compounds with experimental MS2 data (n=453), regardless of whether they were in the composite PCDL.

4.2 In silico Library Matching:

4.2.1 Evaluation by Collision Energy

Regarding the use of *in silico* spectra for compound identification, initial goals of this evaluation were to determine whether 1:1 matching (i.e., one experimental spectrum vs. one *in silico*

spectrum) is best performed at a common CE level, and whether a specific CE level (10, 20, or 40 V data) would stand out as yielding the best results. To achieve these goals, MS2 spectra for "pass" compounds were scored against their respective CFM-ID spectra at all three CE levels. As shown in Figure S3, the highest match scores (where $CE_{experimental} = CE_{in\ silico}$) were generally observed at a CE of 10V, followed by those observed at 20V and 40V. These results likely reflect: 1) the presence and matching of intact precursor ions at lower CE levels; and 2) greater spectral complexity and number of fragments (with some below the experimental mass range) at higher CE levels.

Figure S4 shows, at each $CE_{experimental}$ for each "pass" compound, the quotient of the CFM-ID score when $CE_{experimental} = CE_{in\ silico}$ vs. the CFM-ID score when $CE_{experimental} \neq CE_{in\ silico}$. For each comparison group (n=6), the estimated median value was significantly greater than one (Wilcoxon Signed Rank Test; $p<0.0001$ in all cases), reflecting higher CFM-ID scores when $CE_{experimental} = CE_{in\ silico}$. Not surprisingly, median quotients were highest when the $CE_{experimental}$ and $CE_{in\ silico}$ were most dissimilar (e.g., $10V_{score}/40V_{score}$). Examination of the range of quotients shows that, for some "pass" compounds, the CFM-ID scores were over 1,000 times higher when $CE_{experimental} = CE_{in\ silico}$ vs. when $CE_{experimental} \neq CE_{in\ silico}$. In other cases, however, the CFM-ID scores were up to 100 times lower when $CE_{experimental} = CE_{in\ silico}$. These results highlight the potential value in utilizing *in silico* spectra at non-matching CE levels as part of a composite score. The value of such a proposition is examined below via scoring approaches 2 and 3.

4.2.2 Evaluation by Scoring Method

Three different scoring approaches were compared (Figure 2), with scores based on: 1) 1:1 matching between experimental and *in silico* spectra (where $CE_{experimental} = CE_{in\ silico}$); 2) 1:3 matching with summation across three CFM-ID match scores for a given experimental spectrum; and 3) summation of scores across all possible combinations (n=9) of experimental vs. *in silico* spectra. Each approach was evaluated for all "pass" compounds across all ten ENTACT mixtures.

Distributions of ranks for "pass" compounds amongst all candidate compounds retrieved from the CFM-ID database are given in Table 2 (without formula filtering) and Table 3 (with formula filtering). For approaches 1 and 2, the best results were observed when $CE_{experimental} = 20V$. Results using approach 3 were very comparable to the best results from approaches 1 and 2. Overall, when database matching was performed without formula filtering (Table 2), the spiked compound was ranked as the top candidate up to 38% of the time, within the top 5 candidates up to 60% of the time, and within the top 20 candidates up to 79% of the time. Using approach 3, the spiked compound ranked in the 81st percentile of all candidate compounds, on average, when considering CFM-ID match scores.

As expected, results were markedly better, regardless of the scoring approach, when implementing formula filtering as part of candidate ranking (Table 3). Again, results for approach 3 were very similar to those for approaches 1 and 2 when $CE_{experimental} = 20V$. This time, however, the spiked compound was ranked as the top candidate up to 50% of the time, within the top 5 candidates up to 71% of the time, and within the top 20 candidates up to 85% of the time. On average, using approach 3, the spiked compound was in the 84th percentile of all candidate CFM-ID match scores.

Individual results for each "pass" compound (without and with formula filtering), including the CFM-ID rank of the TP along with number of total candidate compounds, are shown in Figure S5.

Regarding approaches 1 and 2, where a single experimental spectrum is considered at one defined $CE_{experimental}$, performance results generally favor the use of CE = 20V (Tables 2 and 3). A comparative analysis for approach 1, however, shows benefit of considering all three CE results (Figure 3A). Specifically, out of 325 unique compounds identified (without formula filtering) as being within the top 20 CFM-ID hits (at one or more CE), 279 were identified at CE = 20V and 46 were not identified at CE = 20V (Figure 3A). Using approach 3, 298 unique compounds were correctly identified as being within the top 20 CFM-ID hits. Approach 3 coverage exceeded that of approach 1 by 31 compounds when CE = 10V, 19 compounds when CE = 20V, and 83 compounds when CE = 40V (Figure 3B). Considering these findings, composite scoring via approach 3 was used for all remaining evaluations of *in silico* MS2 spectra.

4.2.3. Evaluation of Filtering Criteria

ROC curves in Figure 4A show relationships between TPRs and FPRs, at various percentile and quotient cut-points, when candidates from the CFM-ID database were matched to experimental spectra using precursor mass or predicted formula. In general, results based on quotient cut-offs (in pink) are superior to those based on percentile cut-offs (in green). That is, a lower FPR is associated with a given TPR when using a quotient cut-off at a pre-defined test increment. This result is a function of the right-skewed distribution of quotient values vs. the uniform distribution of percentile values (Supplemental Figure S6). As expected, results based on formula matching (solid) are superior to those based on precursor mass matching (dotted). This result reflects the smaller number of candidate compounds when implementing a formula filter.

As shown in Figure 4A, a global TPR of 0.90 (horizontal gray dashed line) yielded percentile-based FPRs (green vertical dotted lines) of 0.67 (by mass) and 0.36 (by formula), and quotient-based FPRs (pink vertical dotted lines) of 0.57 (by mass) and 0.32 (by formula). This global TPR of 0.90 is associated with percentile cut-off values of 32 (by mass) and 38 (by formula), and quotient cut-off values of 0.13 (by mass) and 0.18 (by formula). Figure 4B shows distributions of TPR and FPR values for individual ENTACT mixtures based on these four cut-off values; these distributions highlight expected ranges of TPRs and FPRs when using the CFM-ID database to investigate unknowns in individual samples. Overall, individual mixture TPRs ranged from 0.72 to 1.0, and FPRs ranged from 0.10 to 0.85. Interestingly, more variability in FPRs was observed in analyses utilizing quotient cut-offs. Thus, FPRs are generally expected to be lower, on average, using quotient cut-offs, but more consistent using percentile cut-offs.

4.3 Comparison of Performance Across Reference and *In Silico* Libraries

Figure 5 shows a comparison of deduplicated "pass" compounds (n=377) that were correctly identified by PCDL reference library matching (n=199) vs. CFM-ID database matching (with formula filtering, n=188). When considering only the top hit from library matching, 88 compounds (23%) were identified only using the composite PCDL, 111 compounds (29%) were identified using both the composite PCDL and the CFM-ID database, and 77 compounds (20%) were identified using only the CFM-ID database. One-hundred-one (27%) compounds were not

identified as the top hit using either the composite PCDL or the CFM-ID database. Ultimately, 53% of "pass" substances were correctly identified by the composite PCDL, and 50% were correctly identified as the top hit using the CFM-ID database. Percentile and quotient-based cut-offs can be used to increase the potential TPR (up to 100%), but at the expense of increasing FPR, as described above. The implementation of cut-off values is at the discretion of the investigator, who must carefully consider the overall objectives of the research study when deciding on a selection strategy.

5.0 Discussion:

Targeted methods have long been the gold standard for chemical analysis. As such, they have been implemented in a wide number of scientific fields where chemical detection and/or quantitation is critical. The focused nature of targeted analytical methods has proven limiting in discovery research fields, where chemicals of eventual interest may not yet be known. NTA methods seek to address this shortcoming by enabling discovery and identification of unknown chemicals and informing follow-up targeted investigations.

Confidence in chemical identifications is a function of the experimental information available [2]. As the amount of information supporting an identification increases, the ambiguity surrounding that identification decreases, resulting in more confident annotations. Targeted methods produce data at the highest confidence level, as they utilize chemical standards for which reference MS1, MS2 and chromatographic data can be acquired. NTA methods can benefit from these reference data to the extent that they have been previously acquired and stored in a usable format. Six Agilent PCDLs were used in this analysis as the source of reference MS2 data for matching; the composite of these PCDLs included 11,324 unique compounds with reference MS2 spectra. The ten ENTACT mixtures contained a total of 1,269 unique substances, of which 610 (48%) were contained within the composite PCDL. The other 52% of compounds represent a "blind spot" in the reference libraries searched. Clearly, *in silico* predicted spectra are needed to enable MS2 matching for compounds not captured in empirical libraries. At the time of analysis, CFM-ID predicted spectra were available for ~765,000 unique DSSTox compounds, representing a >60-fold increase in search space over the composite PCDL. Given the obvious advantage of size, careful evaluation of performance is required to ensure proper use and maximum benefit of these predicted spectra.

Experimental MS2 data for ENTACT mixture compounds were collected and CFM-ID spectra predicted at three CE levels (10, 20 and 40 V). The specificity of CE level when matching experimental and predicted spectra was evaluated across all ten ENTACT mixtures. The highest CFM-ID scores were observed when $CE_{experimental} = CE_{in\ silico}$ (Figure S4). Furthermore, the best performance, in terms of compound ranking, was generally observed when CE=20V (Tables 2 and 3). For some compounds, however, it was more advantageous to acquire and match spectra at CE=10 or 40V (Figure 3A). This is most likely due to variability in compound lability, where different compounds have distinct optimal CE levels needed to generate a spectrum with fragment ions in high abundance. For an NTA workflow where the compounds are unknown, the recommended practice is to acquire experimental MS2 data at all three CE levels in order to capture suitable spectra on the widest range of compounds

It is difficult to anticipate, for a given compound of interest, whether scoring/ranking results at one CE should be preferred over another. Thus, aggregated scoring approaches were evaluated wherein summed scores were considered across multiple CEs (Figure 2). It was generally observed that the quality of matching results increased with the amount of data considered, in terms of both experimental and predicted spectra. Specifically, scoring results from Approaches 2 and 3 were shown to surpass those from Approach 1 at each individual CE (Table 2 and 3, and Figure 3B). Approach 3 tended to yield the best overall results and was therefore the basis for performance evaluations regarding ~~method sensitivity (TRP) and specificity (FPR)~~ TPR and FPR. Moving forward, ~~this is our recommended scoring methodology when using the CFM-ID database as a screening-level tool.~~ when using the CFM-ID database as a screening-level tool, we recommend an aggregated approach wherein each experimental spectrum is compared to all three CE levels of predicted spectra (i.e., Approach 3).

Utilizing CFM-ID results from Approach 3, 34% of the 377 ENTACT mixture compounds were identified as the best matching compound. This result is comparable to those reported from the 2016 CASMI contes~~x~~t, in which 12% to 34% of correct candidates were identified as the best matching compound [20]. In certain cases, sub-optimal performance of CFM-ID may reflect dissimilarities in structures between compounds used to train CFM-ID and those included in ENTACT [27]. A re-training of the CFM-ID models with an expanded set of compounds has the potential to improve scoring and ranking results for the ENTACT mixture compounds. Future work will examine the extent to which re-trained models can better identify ENTACT compounds (and potentially other xenobiotics) amongst other candidate chemicals.

Reference libraries are created from empirical spectra and generally yield matches with high accuracy. That is, the best match from a reference library search is often the TP. Predicted libraries are less accurate, and as such, do not always correctly identify the TP as having the best match score. Utilizing results from *in silico* library searching is therefore a balance between ~~sensitivity (TPR) and specificity (FPR)~~TPR and FPR. Considering only the highest matching compounds will limit the number of FPs, but at a greater risk of missing a TP. A less-stringent cut-off allows for more potential FPs, but also a higher likelihood of retaining the TP. The cut-off threshold depends on the desired goal(s) of the analysis; whether ~~sensitivity or specificity~~ retaining true compounds or eliminating false compounds is of most importance. For this analysis, cut-offs based on percentiles and quotients were evaluated, with candidate selection based on mass matching, with or without additional formula filtering. Our results show a preference for quotient-based cut-offs, and for filtering candidate lists based on molecular formula (Figure 4A). Specifically, the lowest FPR is expected for a given TPR when using a quotient-based cut-off and formula filtering. Better performance using quotient values is attributed to the skewed (i.e., right-tailed) distribution of quotient values (versus the uniform distribution of percentile values), where most candidates have very low CFM-ID match scores, and fewer have moderate to high scores (Figure S6). This allows for more incorrect candidates to be correctly removed from consideration at even a modest cut-point. Interestingly, wider distributions of FPRs were observed when using quotient-based cut-offs vs. percentile-based cut-offs (Figure 4B). This again stems from the skewed distributions of quotient values and underscores the variable nature of FPRs when using quotient cut-offs. More stable FPRs can be achieved with percentile-based cut-offs; these FPRs are expected to be higher, however, when aiming for a high TPR (~0.90).

In silico MS2 libraries are not meant to replace reference libraries. *In silico* library matches are inherently less confident than reference library matches. As such, *in silico* MS2 libraries are not meant to replace reference libraries, but Rather, they are meant to enable supplementary reference library matching procedures [3, 16, 29]. Figure 5 shows that, using either the reference library (composite PCDL) or the *in silico* library (CFM-ID database), about half of the "pass" compounds could be correctly identified as the top match. Using both libraries, however, yielded 73% correct identifications. A hybrid approach is therefore highly desirable for the most comprehensive and accurate analysis. For example, in a hypothetical study, MS2 spectra could be matched to both the reference and *in silico* libraries. Top matches based on the reference library would not require additional support from *in silico* match scores. Yet, these *in silico* match scores could serve as the basis for quotient- or percentile-based cut-points. These cut-points would then be used to filter unlikely candidates retrieved from the CFM-ID database. The use of additional supporting information, such as retention time predictions [30, 31] and metadata source counts [20, 32], has been shown to improve NTA identifications; incorporation of these data with CFM-ID ranking results could further improve candidate filtering, thus increasing the overall accuracy and performance of the workflow. Future investigations will aim to incorporate these various data streams into a unified workflow, and to optimize filtering criteria for maximum TPRs and minimum FPRs.

Since the time of this original analysis, EPA's DSSTox database has increased from ~765,000 to ~875,000 unique substances; CFM-ID predictions have been generated for the majority of these substances based on their associated "MS-Ready" structures. The dynamic nature of *in silico* libraries is a highly desirable feature when compared to reference libraries, which are relatively static due to the need for pure standards. This dependence on standards is a significant drawback when investigating new and rapidly emerging chemicals-of-concern, as the analyses are not able to keep up with the analytes. *In silico* libraries can be generated at a much more rapid pace, on both known and predicted structures (e.g., those of expected metabolites and transformation products) within a given database. EPA's DSSTox database is freely available to the public via the Dashboard (https://comptox.epa.gov/dashboard) [24]. Future Dashboard development will provide additional functionality to support HRMS-based NTA workflows (i.e., retention time predictions, media occurrence data, experimental substructure filtering). Updates to the CFM-ID processing and searching workflow are also being explored, including aggregation of multiple experimental spectra into a single spectrum (rather than selecting only the spectrum of highest sum ion intensity), and implementation of intensity threshold filters (for experimental and predicted spectra) prior to CFM-ID matching/scoring. A prototype web-based tool for searching an experimental spectrum against the CFM-ID database has been developed and is undergoing testing; users will see both the candidate results returned for the spectrum as well as visualizations of the predicted vs. experimental spectrum (Figure S7). CFM-ID batch searching is also being incorporated into existing NTA workflows, with plans to publicly release a stand-alone web service for processing of NTA data. Finally, implementation of CFM-ID 3.0 algorithms (not available at the start of the current project) will likely result in enhanced performance based on an improved *in silico* library [33].

6.0 Conclusions:

Confident identification of unknowns in NTA studies often requires the use of reference library spectra. The relatively modest size of existing reference libraries limits the number of possible identifications for any given study. Use of *in silico* fragmentation libraries can expand coverage into areas not reached by reference libraries alone. Analyses of the ENTACT mixture data shows promising results for the performance of *in silico* spectra towards aiding chemical identification strategies. The expansion of NTA workflows to incorporate *in silico* spectra for >800K DSSTox compounds will enable more rapid and certain identifications of xenobiotics and other emerging compounds.

Acknowledgments
The authors thank James McCord and Ariel Wallace for their thoughtful reviews of this manuscript.

Funding information
The United States Environmental Protection Agency (U.S. EPA), through its Office of Research and Development (ORD), funded and managed the research described here. Partial support for this work was provided by awards from ORD's Pathfinder Innovation Program (PIP) and EPA's Environmental Monitoring and Visualization Laboratory (EMVL). The work has been subjected to Agency administrative review and approved for publication. Alex Chao, Hussein Al-Ghoul, Andrew McEachran and Randolph Singh were supported by an appointment to the Internship/Research Participation Program at the Office of Research and Development, U.S. Environmental Protection Agency, administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and EPA.

Compliance with ethical standards~~Conflict of interest~~
The authors declare that they have no conflicts of interest.

Disclaimer
The views expressed in this paper are those of the authors and do not necessarily represent the views or policies of the U.S. EPA.

References:

1. Wild, CP, Complementing the genome with an "exposome": The outstanding challenge of environmental exposure measurement in molecular epidemiology. Cancer Epidem Biomar, 2005. **14**(8): p. 1847-1850.
2. Schymanski, EL, J Jeon, R Gulde, K Fenner, M Ruff, HP Singer, et al., Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ Sci Technol, 2014. **48**(4): p. 2097-8.
3. Blazenovic, I, T Kind, J Ji, and O Fiehn, Software Tools and Approaches for Compound Identification of LC-MS/MS Data in Metabolomics. Metabolites, 2018. **8**(2).
4. Kind, T, H Tsugawa, T Cajka, Y Ma, ZJ Lai, SS Mehta, et al., Identification of small molecules using accurate mass MS/MS search. Mass Spectrom Rev, 2018. **37**(4): p. 513-532.
5. Peisl, BYL, EL Schymanski, and P Wilmes, Dark matter in host-microbiome metabolomics: Tackling the unknowns-A review. Anal Chim Acta, 2018. **1037**: p. 13-27.
6. Little, JL, AJ Williams, A Pshenichnov, and V Tkachenko, Identification of "Known Unknowns" Utilizing Accurate Mass Data and ChemSpider. J Am Soc Mass Spectr, 2012. **23**(1): p. 179-185.

7. Kim, S, J Chen, TJ Cheng, A Gindulyte, J He, SQ He, et al., PubChem 2019 update: improved access to chemical data. Nucleic Acids Res, 2019. **47**(D1): p. D1102-D1109.
8. Tsugawa, H, T Kind, R Nakabayashi, D Yukihira, W Tanaka, T Cajka, et al., Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. Anal Chem, 2016. **88**(16): p. 7946-7958.
9. Wolf, S, S Schmidt, M Muller-Hannemann, and S Neumann, In silico fragmentation for computer assisted identification of metabolite mass spectra. Bmc Bioinformatics, 2010. **11**.
10. Ridder, L, JJJ van der Hooft, S Verhoeven, RCH de Vos, R van Schaik, and J Vervoort, Substructure-based annotation of high-resolution multistage MSn spectral trees. Rapid Commun Mass Sp, 2012. **26**(20): p. 2461-2471.
11. Ruttkies, C, EL Schymanski, S Wolf, J Hollender, and S Neumann, MetFrag relaunched: incorporating strategies beyond in silico fragmentation. J Cheminformatics, 2016. **8**.
12. Laponogov, I, N Sadawi, D Galea, R Mirnezami, and KA Veselkov, ChemDistiller: an engine for metabolite annotation in mass spectrometry. Bioinformatics, 2018. **34**(12): p. 2096-2102.
13. Duhrkop, K, HB Shen, M Meusel, J Rousu, and S Bocker, Searching molecular structure databases with tandem mass spectra using CSI:FingerID. P Natl Acad Sci USA, 2015. **112**(41): p. 12580-12585.
14. Shen, HB, K Duhrkop, S Bocker, and J Rousu, Metabolite identification through multiple kernel learning on fragmentation trees. Bioinformatics, 2014. **30**(12): p. 157-164.
15. Hufsky, F and S Bocker, Mining Molecular Structure Databases: Identification of Small Molecules Based on Fragmentation Mass Spectrometry Data. Mass Spectrom Rev, 2017. **36**(5): p. 624-633.
16. Hufsky, F, K Scheubert, and S Bocker, Computational mass spectrometry for small-molecule fragmentation. Trac-Trend Anal Chem, 2014. **53**: p. 41-48.
17. Allen, F, R Greiner, and D Wishart, Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. Metabolomics, 2015. **11**(1): p. 98-110.
18. Allen, F, A Pon, M Wilson, R Greiner, and D Wishart, CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. Nucleic Acids Res, 2014. **42**(W1): p. W94-W99.
19. Schymanski, EL and S Neumann, The Critical Assessment of Small Molecule Identification (CASMI): Challenges and Solutions. Metabolites, 2013. **3**(3): p. 517-38.
20. Schymanski, EL, C Ruttkies, M Krauss, C Brouard, T Kind, K Duhrkop, et al., Critical Assessment of Small Molecule Identification 2016: automated methods. J Cheminformatics, 2017. **9**.
21. Ulrich, EM, JR Sobus, CM Grulke, AM Richard, SR Newton, MJ Strynar, et al., EPA's non-targeted analysis collaborative trial (ENTACT): genesis, design, and initial findings. Anal Bioanal Chem, 2019. **411**(4): p. 853-866.
22. Sobus, JR, JF Wambaugh, KK Isaacs, AJ Williams, AD McEachran, AM Richard, et al., Integrating tools for non-targeted analysis research and chemical safety evaluations at the US EPA. J Expo Sci Environ Epidemiol, 2018. **28**(5): p. 411-426.
23. Sobus, JR, JN Grossman, A Chao, R Singh, AJ Williams, CM Grulke, et al., Using prepared mixtures of ToxCast chemicals to evaluate non-targeted analysis (NTA) method performance. Anal Bioanal Chem, 2019. **411**(4): p. 835-851.
24. Williams, AJ, CM Grulke, J Edwards, AD McEachran, K Mansouri, NC Baker, et al., The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. J Cheminformatics, 2017. **9**.
25. Grulke, CM, AJ Williams, I Thillanadarajah, and AM Richard, EPA's DSSTox database: History of development of a curated chemistry resource supporting computational toxicology research. Comput Toxicol, 2019. **12**: p. 100096.
26. McEachran, AD, K Mansouri, C Grulke, EL Schymanski, C Ruttkies, and AJ Williams, "MS-Ready" structures for non-targeted high-resolution mass spectrometry screening studies. J Cheminform, 2018. **10**(1): p. 45.
27. McEachran, AD, I Balabin, T Cathey, TR Transue, H Al-Ghoul, C Grulke, et al., Linking in silico MS/MS spectra with chemistry data to improve identification of unknowns. Sci Data, 2019. **6**.
28. Stein, SE and DR Scott, Optimization and Testing of Mass-Spectral Library Search Algorithms for Compound Identification. J Am Soc Mass Spectr, 1994. **5**(9): p. 859-866.
29. Allard, PM, T Peresse, J Bisson, K Gindro, L Marcourt, VC Pham, et al., Integration of Molecular Networking and In-Silico MS/MS Fragmentation for Natural Products Dereplication. Anal Chem, 2016. **88**(6): p. 3317-23.

30. Aalizadeh, R, MC Nika, and NS Thomaidis, Development and application of retention time prediction models in the suspect and non-target screening of emerging contaminants. J Hazard Mater, 2019. **363**: p. 277-285.
31. McEachran, AD, K Mansouri, SR Newton, BEJ Beverly, JR Sobus, and AJ Williams, A comparison of three liquid chromatography (LC) retention time prediction models. Talanta, 2018. **182**: p. 371-379.
32. McEachran, AD, JR Sobus, and AJ Williams, Identifying known unknowns using the US EPA's CompTox Chemistry Dashboard. Anal Bioanal Chem, 2017. **409**(7): p. 1729-1735.
33. Djoumbou-Feunang, Y, A Pon, N Karu, JM Zheng, C Li, D Arndt, et al., CFM-ID 3.0: Significantly Improved ESI-MS/MS Prediction and Compound Identification. Metabolites, 2019. **9**(4).

In silico MS/MS spectra for identifying unknowns: A critical examination using CFM-ID algorithms and ENTACT mixture samples

Alex Chao, Hussein Al-Ghoul, Andrew D. McEachran, Ilya Balabin, Tom Transue, Tommy Cathey, Jarod N. Grossman, Randolph Singh, Elin M. Ulrich, Antony J. Williams, Jon R. Sobus

Figure Captions

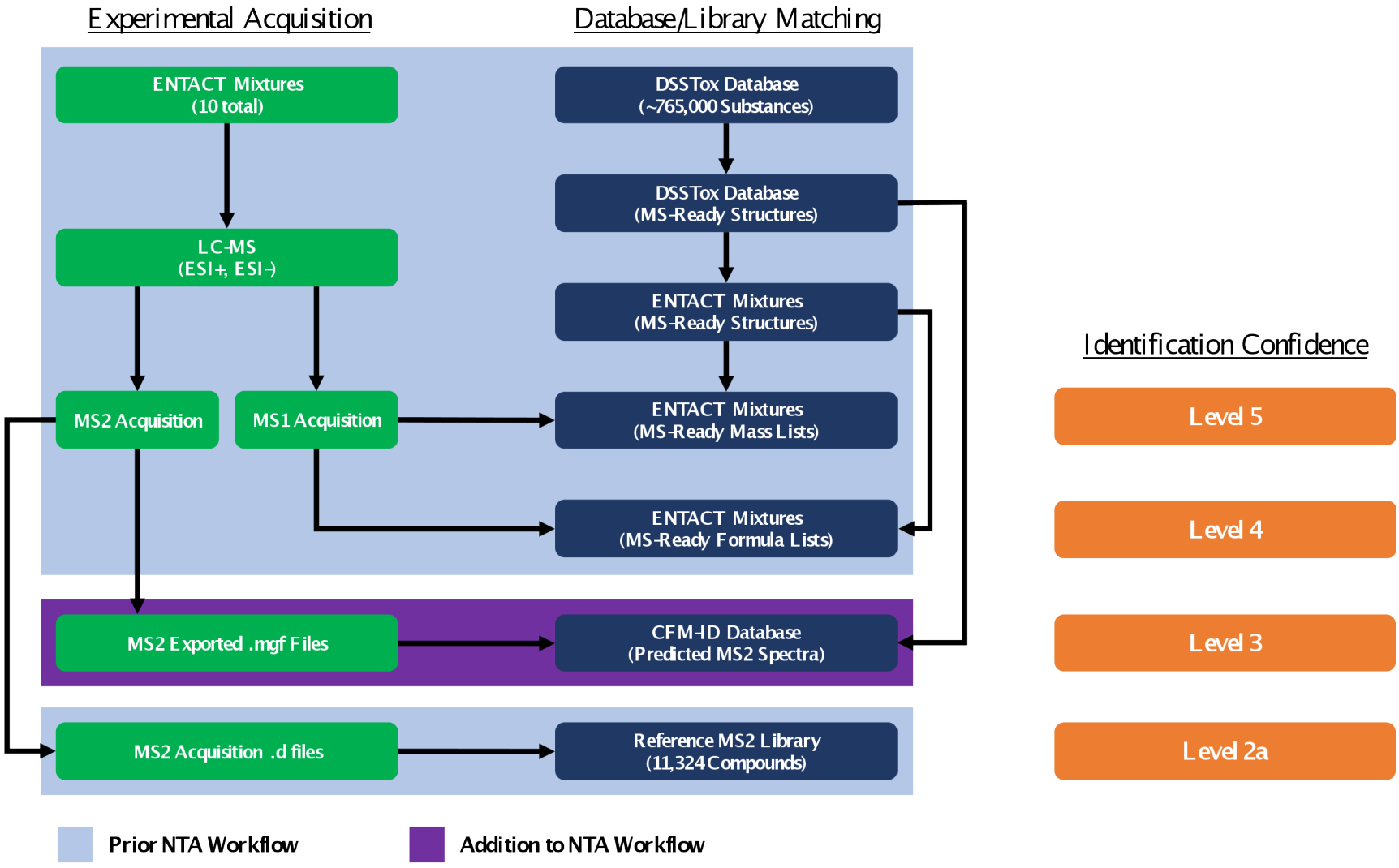
Figure 1. Overall workflow for data acquisition and compound identification. Sections outlined in blue show aspects of the workflow previously implemented for the analysis of ENTACT mixtures. The section outlined in purple shows additions to the workflow that involve matching experimental MS2 spectra with CFM-ID predicted spectra. Identification confidence levels [2] for each match of experimental data to a corresponding database/library entry are shown alongside the specified match in the workflow.

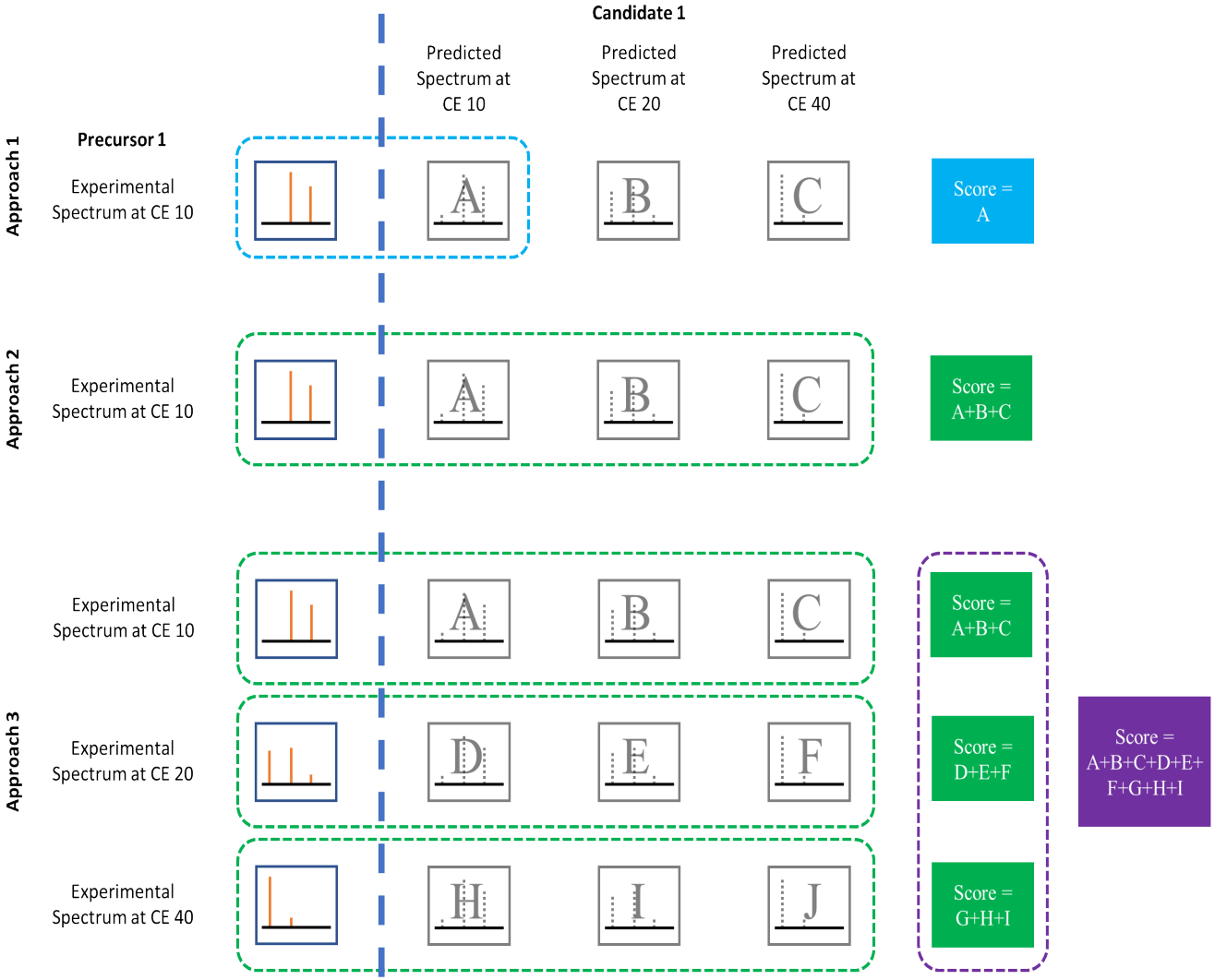
Figure 2. Three approaches for utilizing CFM-ID scores. Each combination of experimental spectrum vs. CFM-ID predicted spectrum generates a unique score via the dot-product algorithm, designated by a unique letter assignment. In approach 1, only one score is generated at the designated collision energy (CE, where $CE_{experimental} = CE_{in\ silico}$). In approach 2, scores from all three $CE_{in\ silico}$ levels are summed. In approach 3, scores are summed across all three $CE_{in\ silico}$ levels, and then across all three $CE_{experimental}$ levels.

Figure 3. Number of "pass" compounds within the top 20 CFM-ID hits using Approach 1 at CE=10V vs. 20V vs. 40V (A). Number of "pass" compounds within the top 20 CFM-ID hits using Approach 3 vs. Approach 1 at CE=10, 20 or 40V (B).

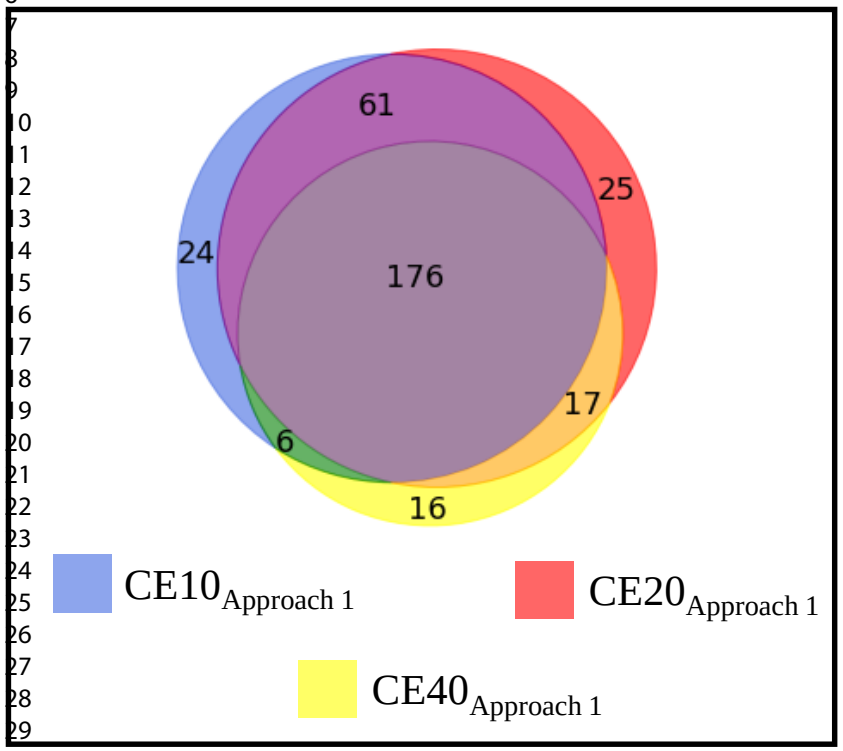
Figure 4. ROC curves (A) for ENTACT mixture data (all "pass" compounds from all ten mixtures) when using percentile and quotient cut-off values, and when filtering the CFM-ID database matches by mass or molecular formula. A global TPR of 0.90 (horizontal gray dashed line) results in percentile-based FPR values (green vertical dotted lines) of 0.67 (by mass) and 0.36 (by formula), and quotient-based FPR values (pink vertical dotted lines) of 0.57 (by mass) and 0.32 (by formula). Distributions (B) of True Positive Rates (TPR) and False Positive Rates (FPRs) across individual ENTACT mixtures (n=10) when selecting cut-off values based on a global TPR of 0.90 (from [A]).

Figure 5. Comparison of "pass" compounds (n=377) correctly identified by reference library matching (using a composite Agilent PCDL) vs. CFM-ID database matching (when filtering by molecular formula).

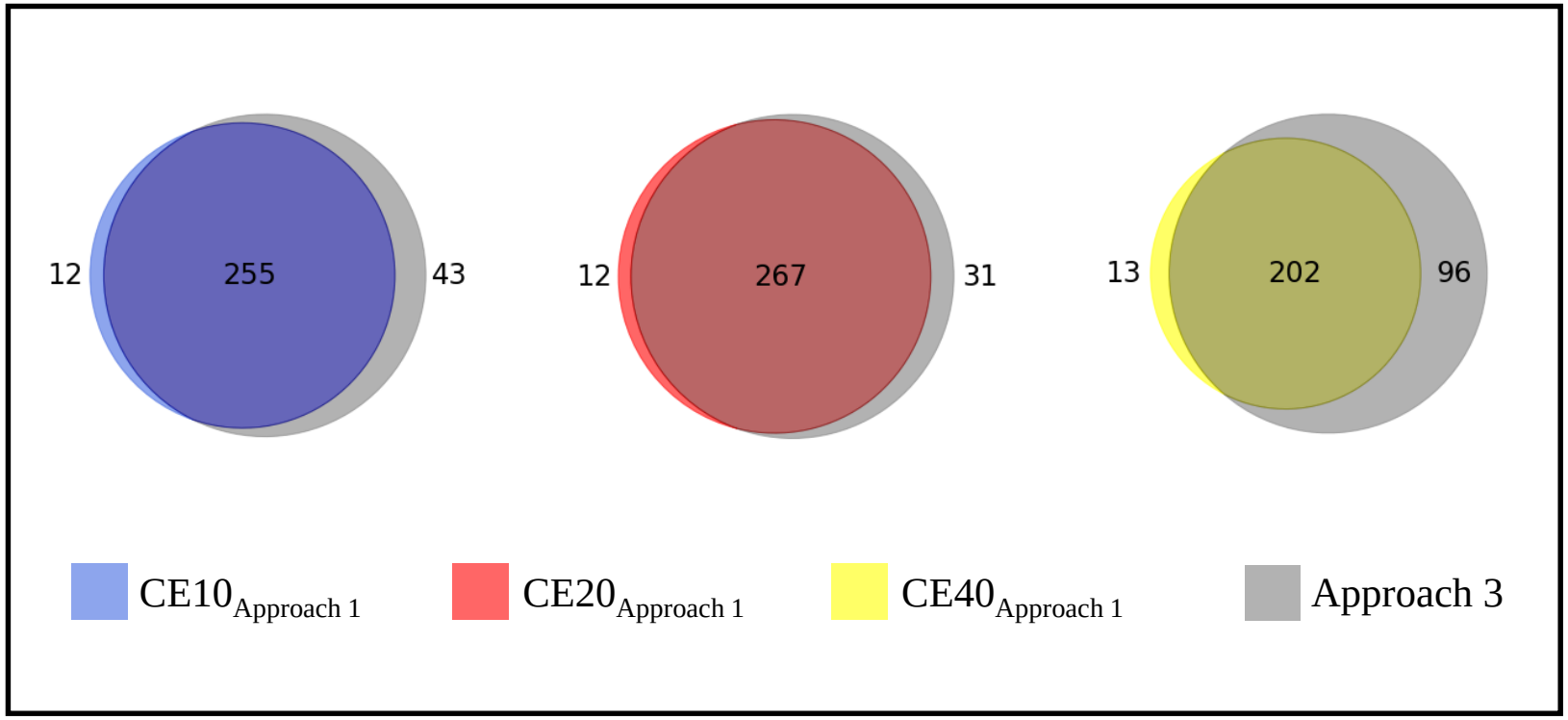

Experimental Acquisition
Database/Library Matching
ENTACT Mixtures
(10 total)
LC-MS
(ESI+, ESI-)
MS2 Acquisition
MS1 Acquisition
MS2 Exported .mgf Files
MS2 Acquisition .d files
DSSTox Database
(~765,000 Substances)
DSSTox Database
(MS-Ready Structures)
ENTACT Mixtures
(MS-Ready Structures)
ENTACT Mixtures
(MS-Ready Mass Lists)
ENTACT Mixtures
(MS-Ready Formula Lists)
CFM-ID Database
(Predicted MS2 Spectra)
Reference MS2 Library
(11,324 Compounds)
Identification Confidence
Level 5
Level 4
Level 3
Level 2a
Prior NTA Workflow
Addition to NTA Workflow

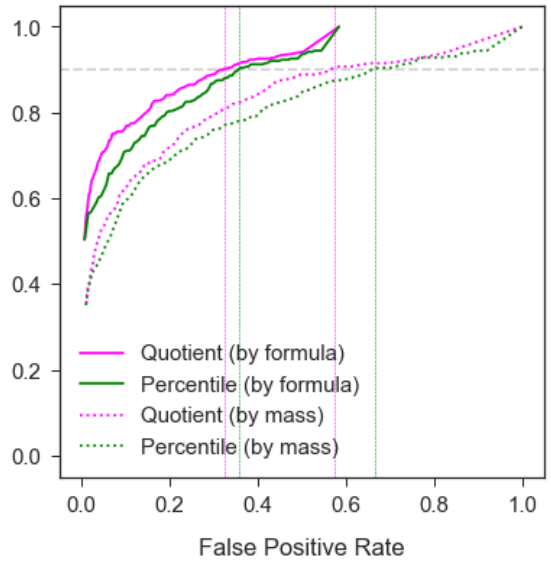

Candidate 1
Predicted Spectrum at CE 10
Predicted Spectrum at CE 20
Predicted Spectrum at CE 40
Approach 1
Precursor 1
Experimental Spectrum at CE 10
A
B
C
Score = A
Approach 2
Experimental Spectrum at CE 10
A
B
C
Score = A+B+C
Approach 3
Experimental Spectrum at CE 10
A
B
C
Score = A+B+C
Experimental Spectrum at CE 20
D
E
F
Score = D+E+F
Experimental Spectrum at CE 40
H
I
J
Score = G+H+I
Score = A+B+C+D+E+F+G+H+I

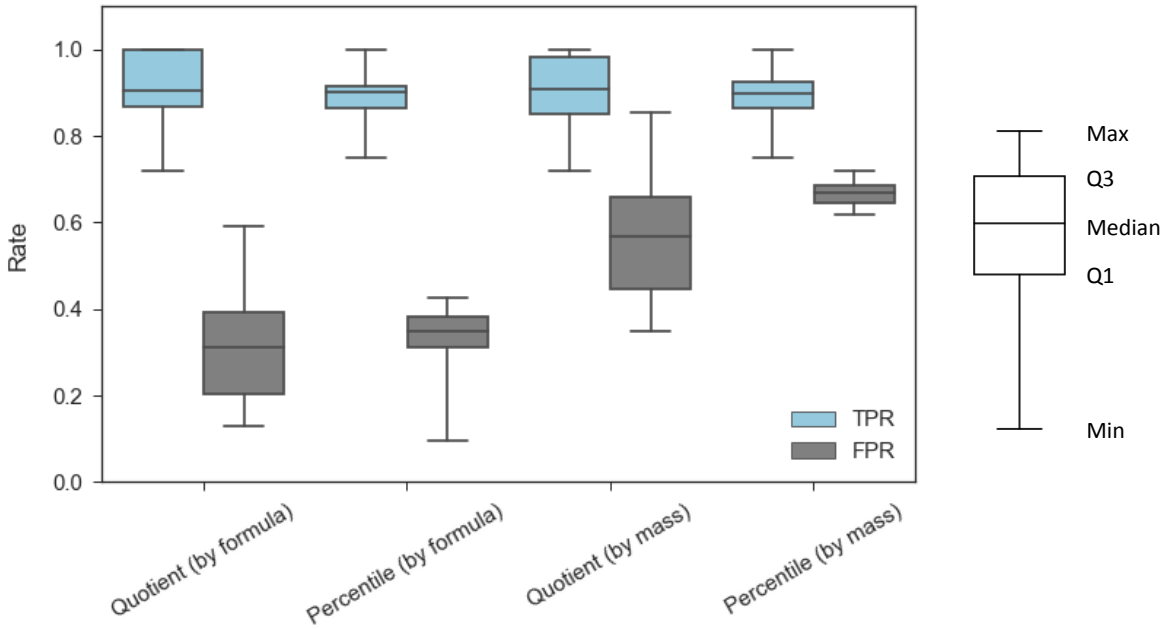
A.


61
25
24
176
17
6
16
CE10Approach 1
CE20Approach 1
CE40Approach 1

B.


12
255
43
12
267
31
13
202
96
CE10Approach 1
CE20Approach 1
CE40Approach 1
Approach 3


A.
True Positive Rate
False Positive Rate
Quotient (by formula)
Percentile (by formula)
Quotient (by mass)
Percentile (by mass)


B.
Rate
TPR
FPR
Quotient (by formula)
Percentile (by formula)
Quotient (by mass)
Percentile (by mass)
Max
Q3
Median
Q1
Min

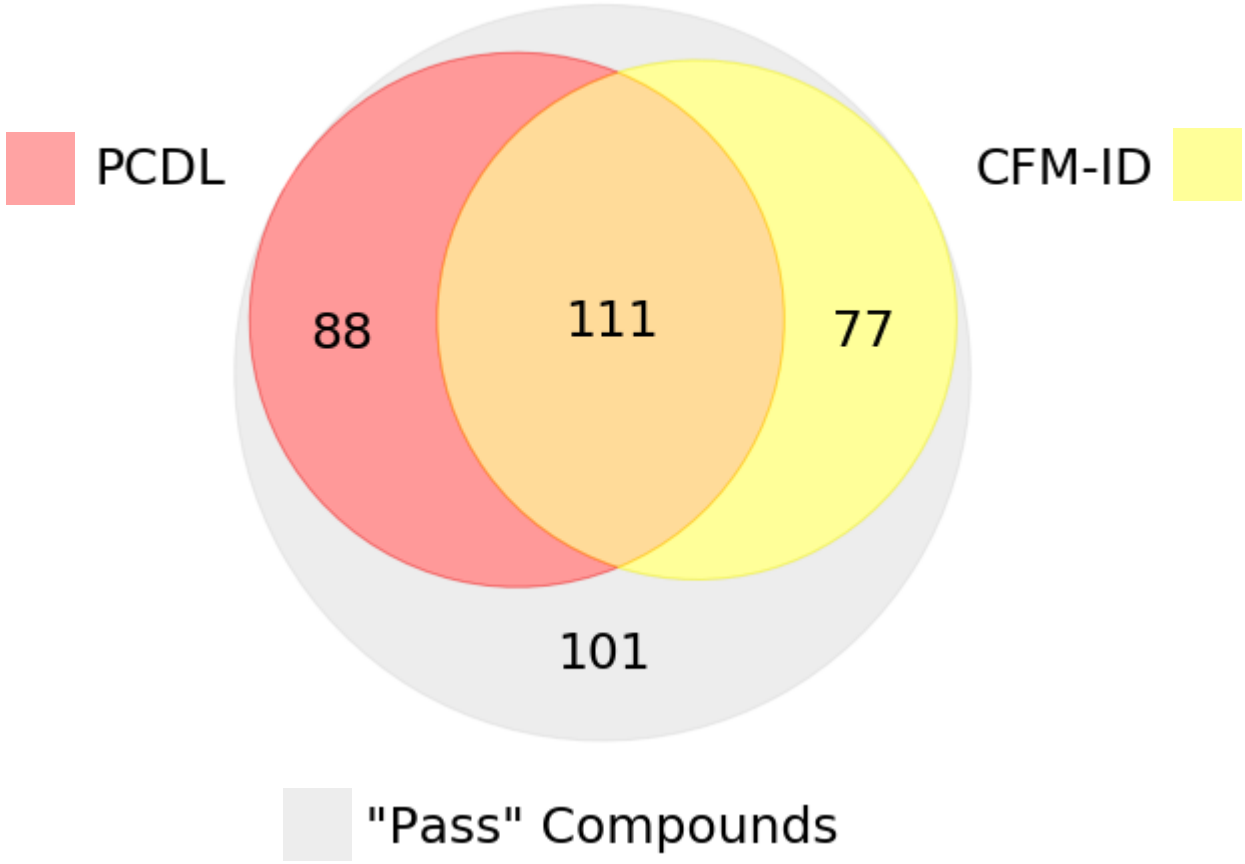

PCDL
CFM-ID
88
111
77
101
"Pass" Compounds

Table 1: Numbers of spiked ENTACT substances meeting specific research criteria.

Mixture	Spiked Substances	Passes	Passes in PCDL[1]	Passes w/ MS2	Passes in PCDL and w/ MS2	Passes Matched by PCDL
499	95	46	28	37	23	18
500	95	19	14	14	11	7
501	95	47	28	34	25	23
502	95	58	42	22	17	15
503	185	103	59	67	43	34
504	185	103	55	68	41	34
505	365	224	128	64	44	40
506	365	195	114	113	74	61
507	95	19	13	14	9	7
508	364	31	19	20	13	7
Total	1939	845	500	453	300	246
% of Total	NA	44%	26%	23%	15%	13%
% of Passes	NA	NA	59%	54%	36%	29%

[1] Composite "Personal Compound Database and Library" (PCDL) containing compounds from six individual Agilent PCDLs (i.e., Environmental water screening, Pesticides, Forensic toxicology, Veterinary drugs, Metlin, and Extractable and leachables).

Table 2: CFM-ID results for ENTACT mixture compounds across three scoring approaches (Figure 2). Candidate compounds from the CFM-ID database were limited to those having an MS-Ready monoisotopic mass matching (within 10 ppm) that of the known (spiked) substance.

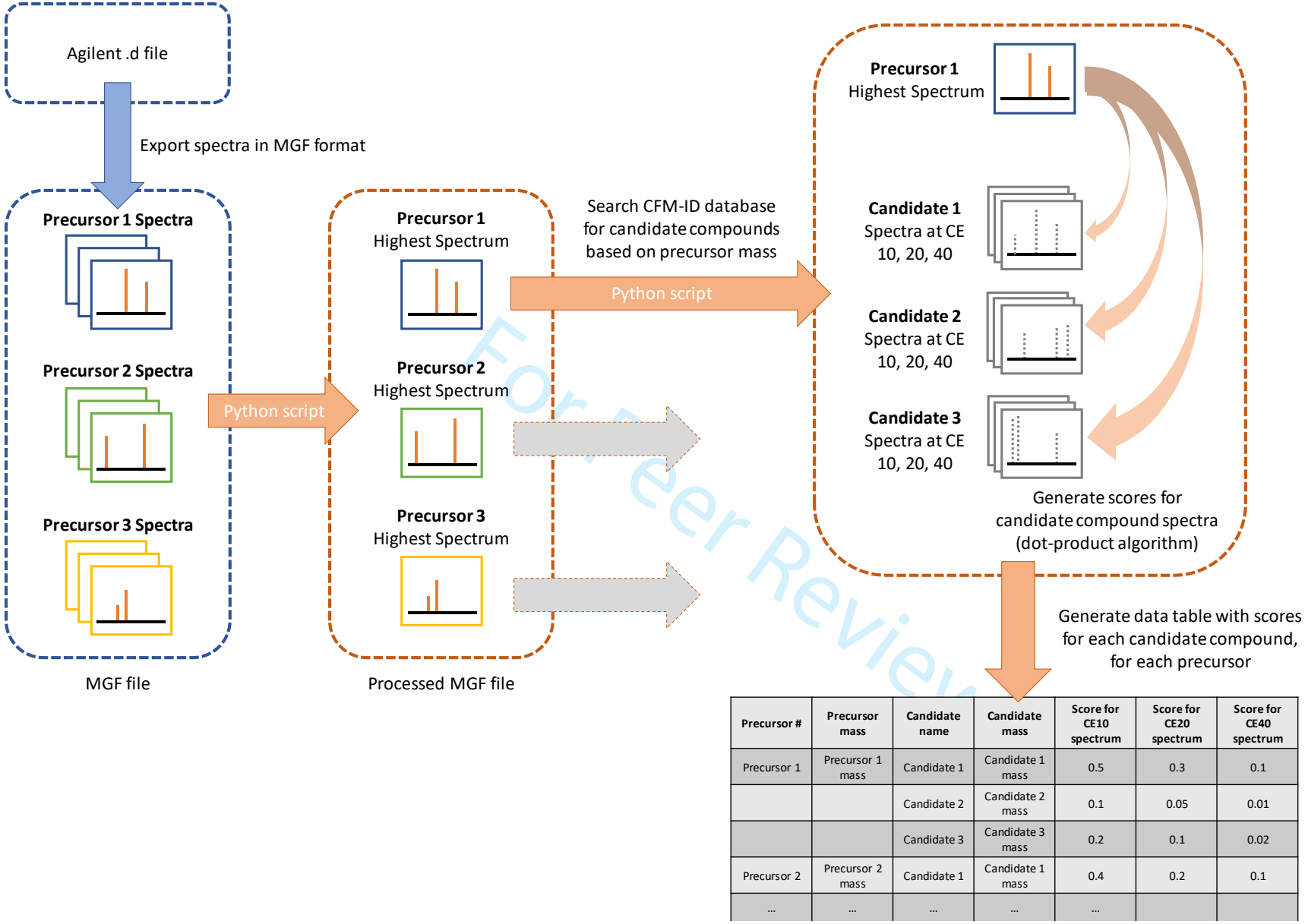
	Approach 1			Approach 2			Approach 3
$CE_{experimental}$	10	20	40	10	20	40	Σ^a
$CE_{in\ silico}$	10	20	40	Σ	Σ	Σ	Σ
# of Compounds Scored	363	368	360	363	368	360	377
Number of True Positives							
Top Hit	102	129	93	100	139	100	129
Within Top 5	187	219	162	188	221	162	224
Within Top 20	267	279	215	275	283	213	298
Percentage of True Positives							
Top Hit	28%	35%	26%	28%	38%	28%	34%
Within Top 5	52%	60%	45%	52%	60%	45%	59%
Within Top 20	74%	76%	60%	76%	77%	59%	79%
Average Percentile for True Positives	77th	81st	72nd	78th	82nd	73rd	81st
Average Quotient for True Positives	0.67	0.62	0.45	0.64	0.65	0.47	0.69

[a] Sum of three CEs

Table 3: CFM-ID results for ENTACT mixture compounds across three scoring approaches (Figure 2). Candidate compounds from the CFM-ID database were limited to those having an MS-Ready formula matching that of the known (spiked) substance.

	Approach 1			Approach 2			Approach 3
$CE_{experimental}$	10	20	40	10	20	40	Σ^a
$CE_{in\ silico}$	10	20	40	Σ	Σ	Σ	Σ
# of Compounds Scored	363	368	360	363	368	360	377
Number of True Positives							
Top Hit	159	178	123	171	180	128	188
Within Top 5	239	250	194	243	252	194	268
Within Top 20	284	291	232	295	292	232	321
Percentage of True Positives							
Top Hit	44%	48%	34%	47%	49%	36%	50%
Within Top 5	66%	68%	54%	67%	68%	54%	71%
Within Top 20	78%	79%	64%	81%	79%	64%	85%
Average Percentile for True Positives	82nd	83rd	76th	83rd	84th	77th	84th
Average Quotient for True Positives	0.77	0.73	0.57	0.77	0.75	0.59	0.79

[a] Sum of three CEs



Precursor #	Precursor mass	Candidate name	Candidate mass	Score for CE10 spectrum	Score for CE20 spectrum	Score for CE40 spectrum
Precursor 1	Precursor 1 mass	Candidate 1	Candidate 1 mass	0.5	0.3	0.1
		Candidate 2	Candidate 2 mass	0.1	0.05	0.01
		Candidate 3	Candidate 3 mass	0.2	0.1	0.02
Precursor 2	Precursor 2 mass	Candidate 1	Candidate 1 mass	0.4	0.2	0.1
...	...	...	...	...		

Figure S1. Workflow for parsing experimental (MS2) spectra and matching to CFM-ID predicted spectra. Blue outlines indicate vendor software steps and orange outlines indicate steps taken via custom Python scripts.

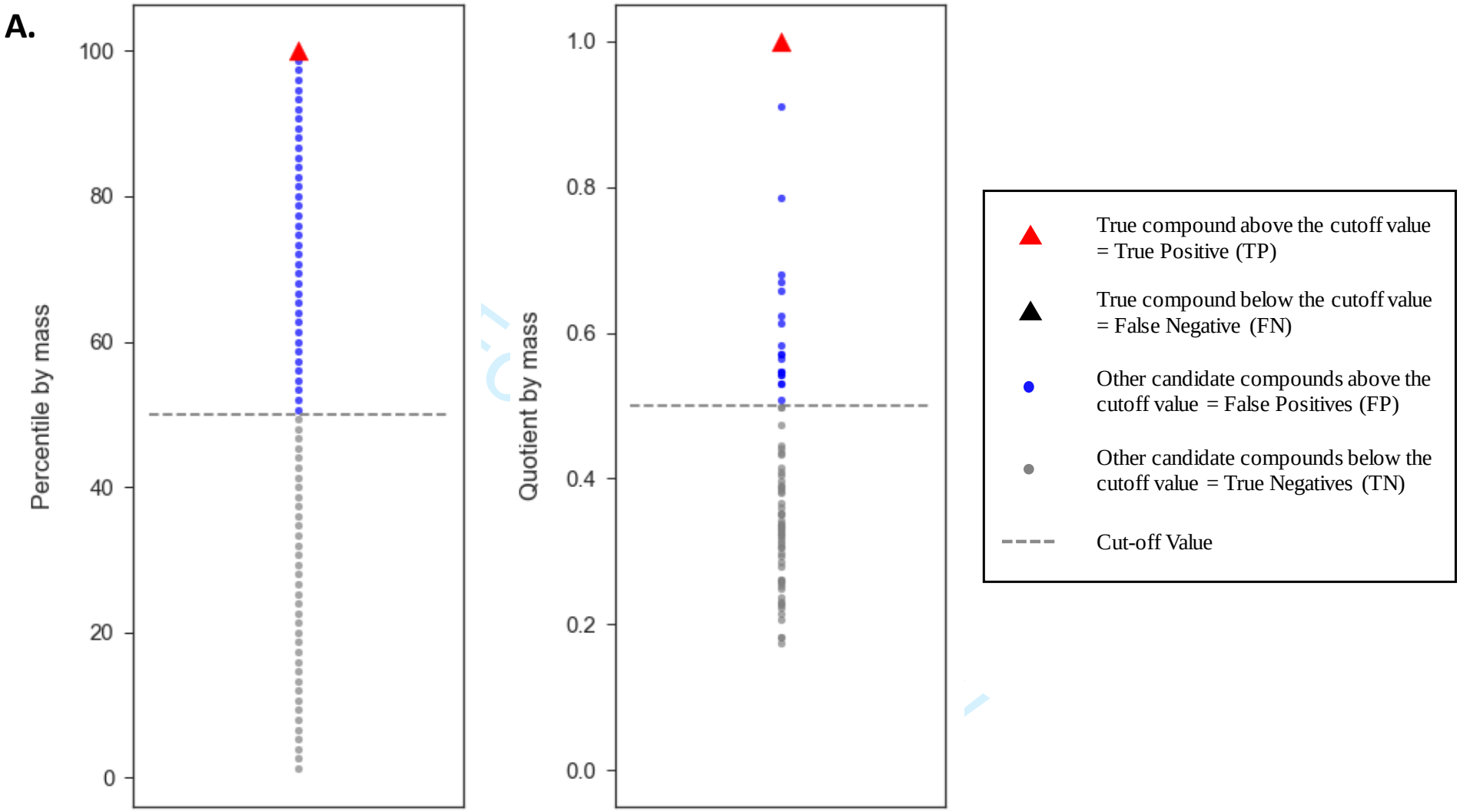


Figure S2A. Distributions of candidate compounds (n=75) for a single CFM-ID query (by mass) based on calculated percentile (left) and quotient (right) values. Here, the "pass" compound was correctly kept above the selected cut-off values, resulting in a "True Positive" (TP) assignment.

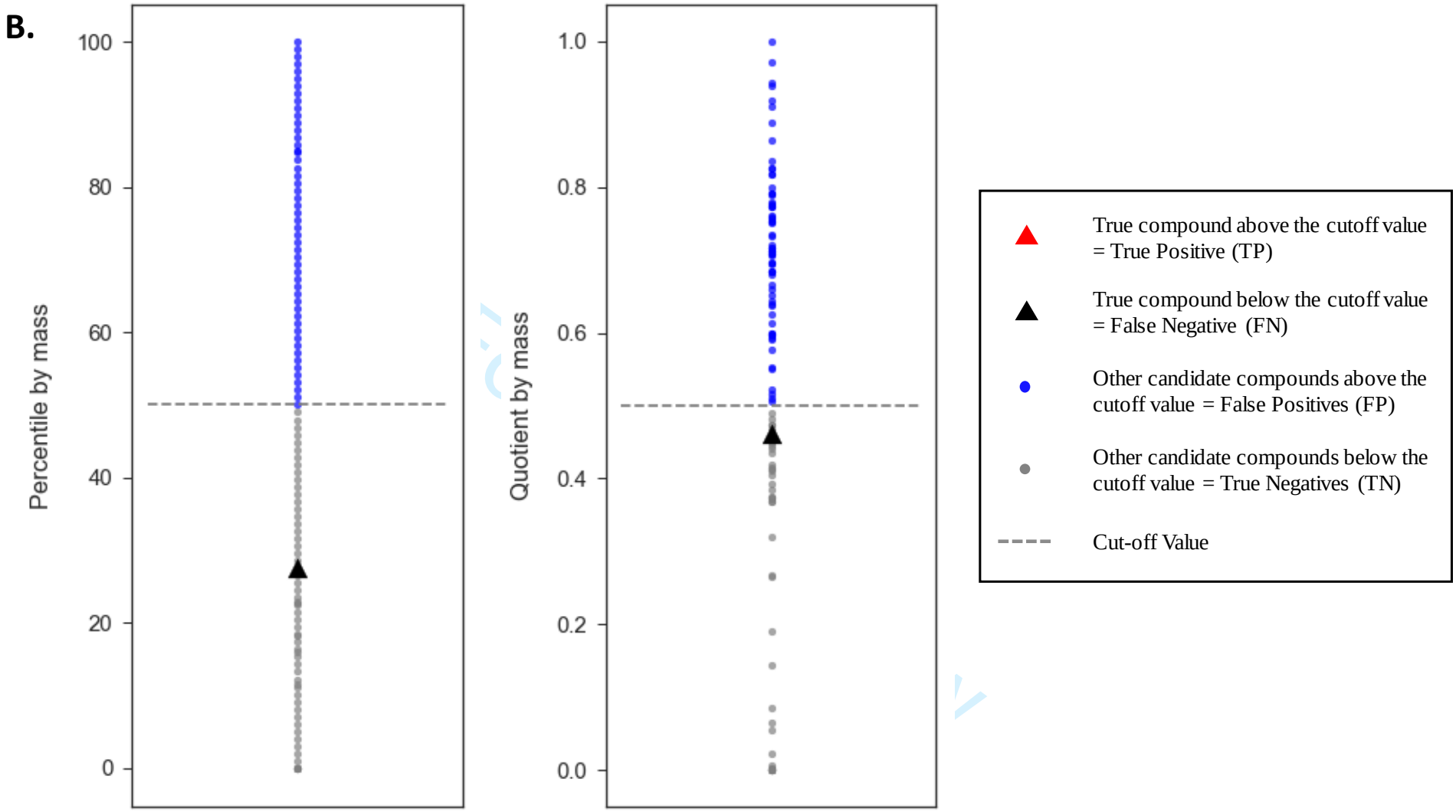


Figure S2B. Distributions of candidate compounds (n=111) for a single CFM-ID query (by mass) based on calculated percentile (left) and quotient (right) values. Here, the "pass" compound was not kept above the selected cut-off values, resulting in a "False Negative" (FN) assignment.

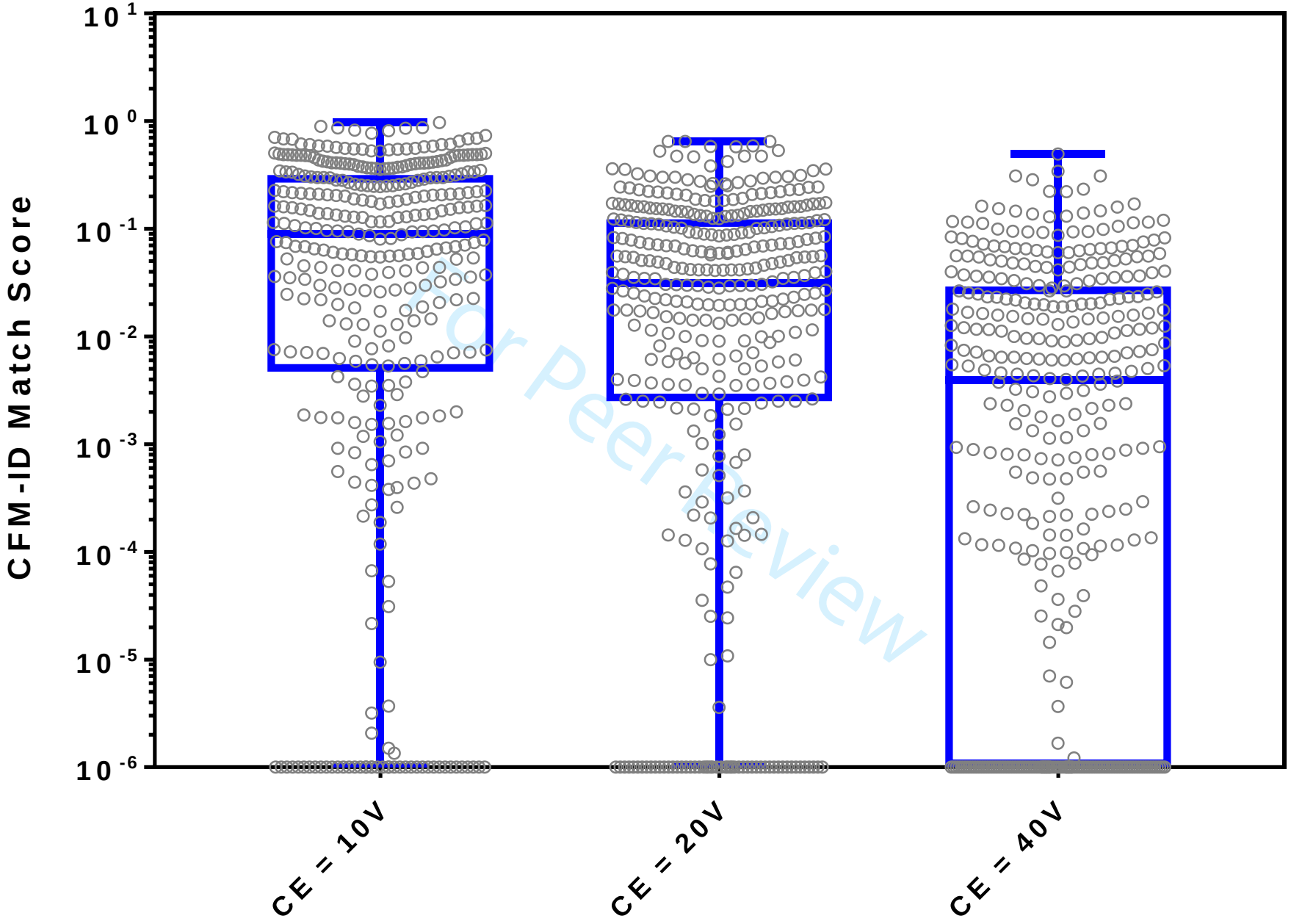


Figure S3. Distributions of CFM-ID match scores at CE=10, 20 and 40V, where $CE_{experimental} = CE_{in\ silico}$.

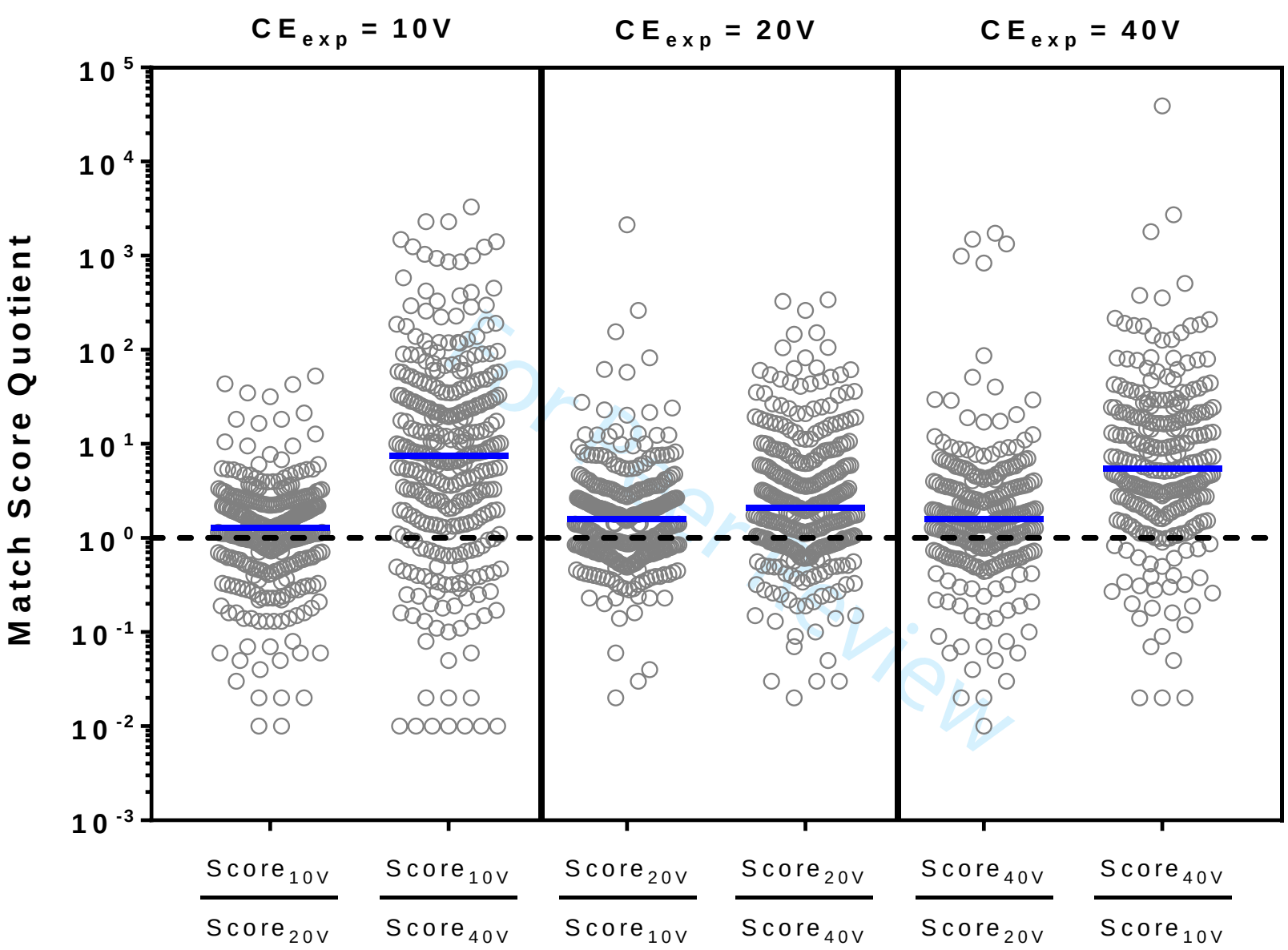


Figure S4. CFM-ID match score quotients for "pass" compounds with MS2 spectra acquired at CE=10, 20 or 40V. Each open circle represents, for a given "pass" compound, the quotient of the CFM-ID score when $CE_{experimental} = CE_{in\ silico}$ vs. the CFM-ID score when $CE_{experimental} \neq CE_{in\ silico}$. The blue horizontal lines represent the median match score quotients for the individual comparison groups (n=6). For each group, the median match score quotient was significantly greater than 1 (p<0.0001).

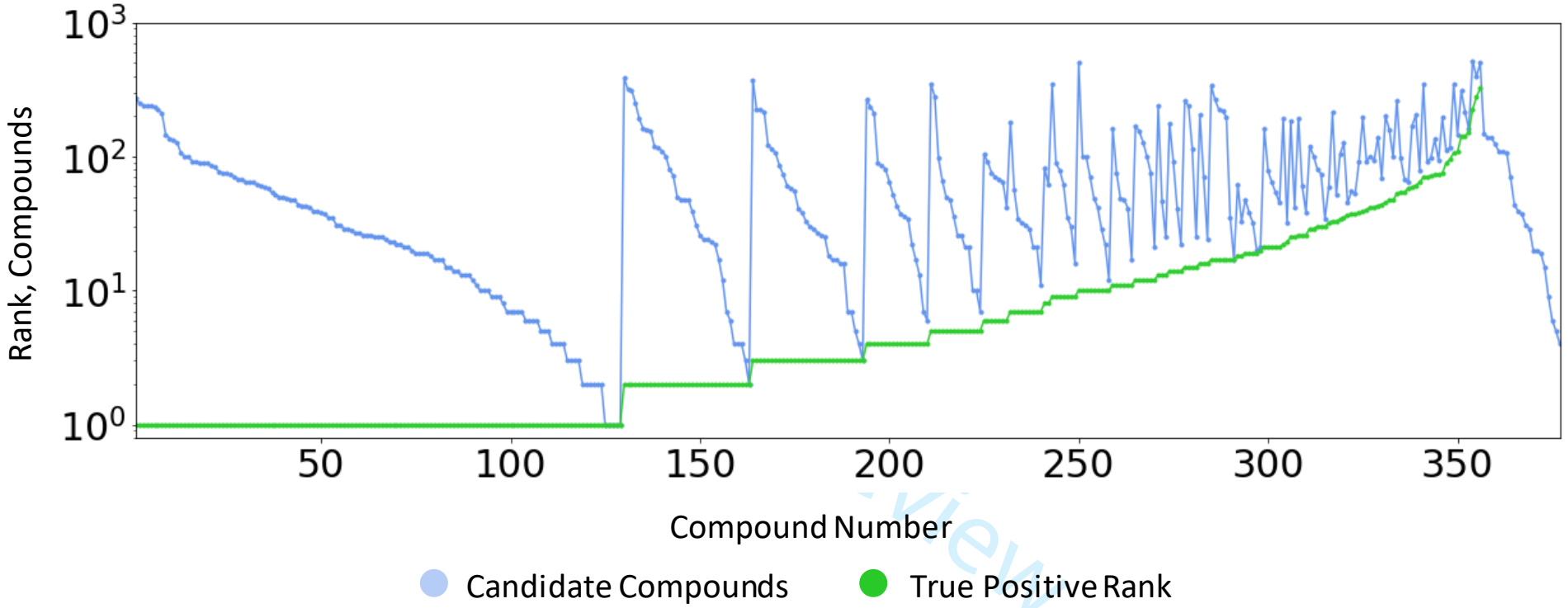


Figure S5A: Results of CFM-ID scoring for all "pass" compounds (n=377) when queried by mass. For each vertical pair, the blue point represents the number of retrieved candidate compounds for a given mass match, and the green point represents the rank of the True Positive (TP) compound. Data are sorted by rank (increasing) and then number of candidate compounds (decreasing). "Pass" compounds without an associated scoring rank had insufficient fragment matches between the experimental and predicted spectra, and therefore, no match score.

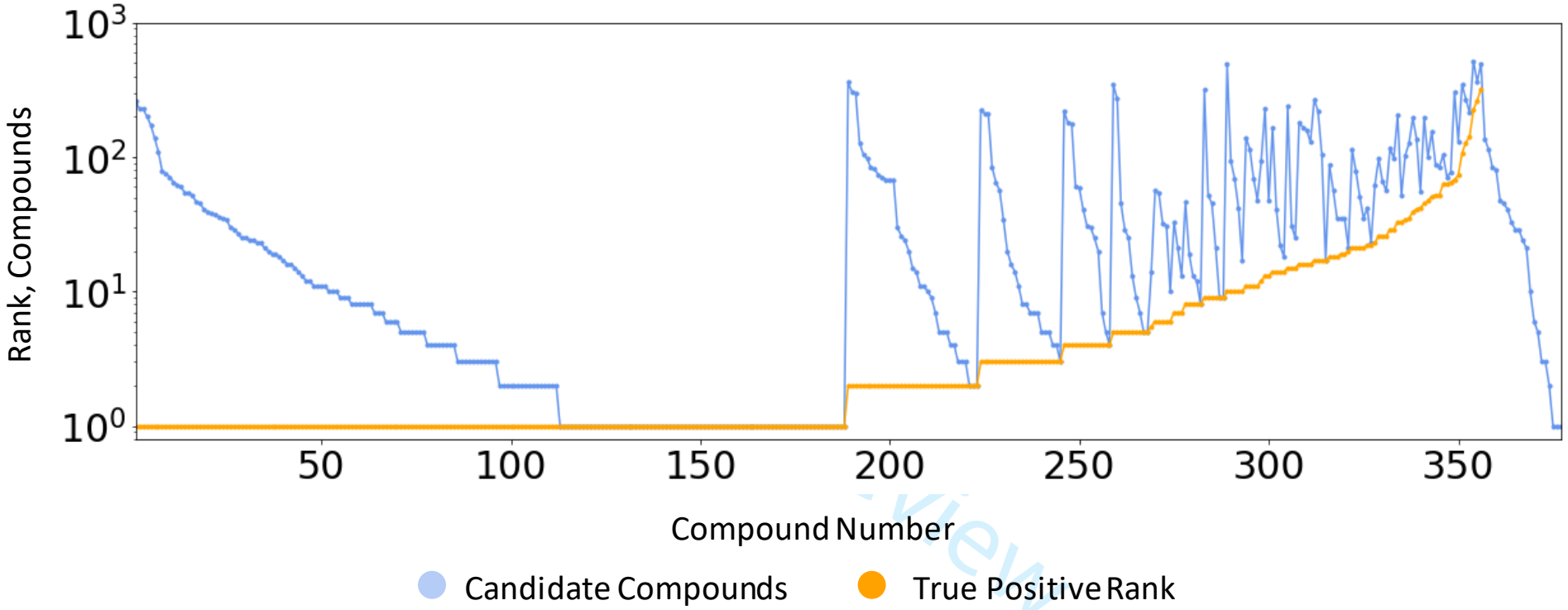


Figure S5B: Results of CFM-ID scoring for all "pass" compounds (n=377) when queried by mass and filtered by molecular formula. For each vertical pair, the blue point represents the number of retrieved candidate compounds for a given formula match, and the orange point represents the rank of the True Positive (TP) compound. Data are sorted by rank (increasing) and then number of candidate compounds (decreasing). "Pass" compounds without an associated scoring rank had insufficient fragment matches between the experimental and predicted spectra, and therefore, no match score.

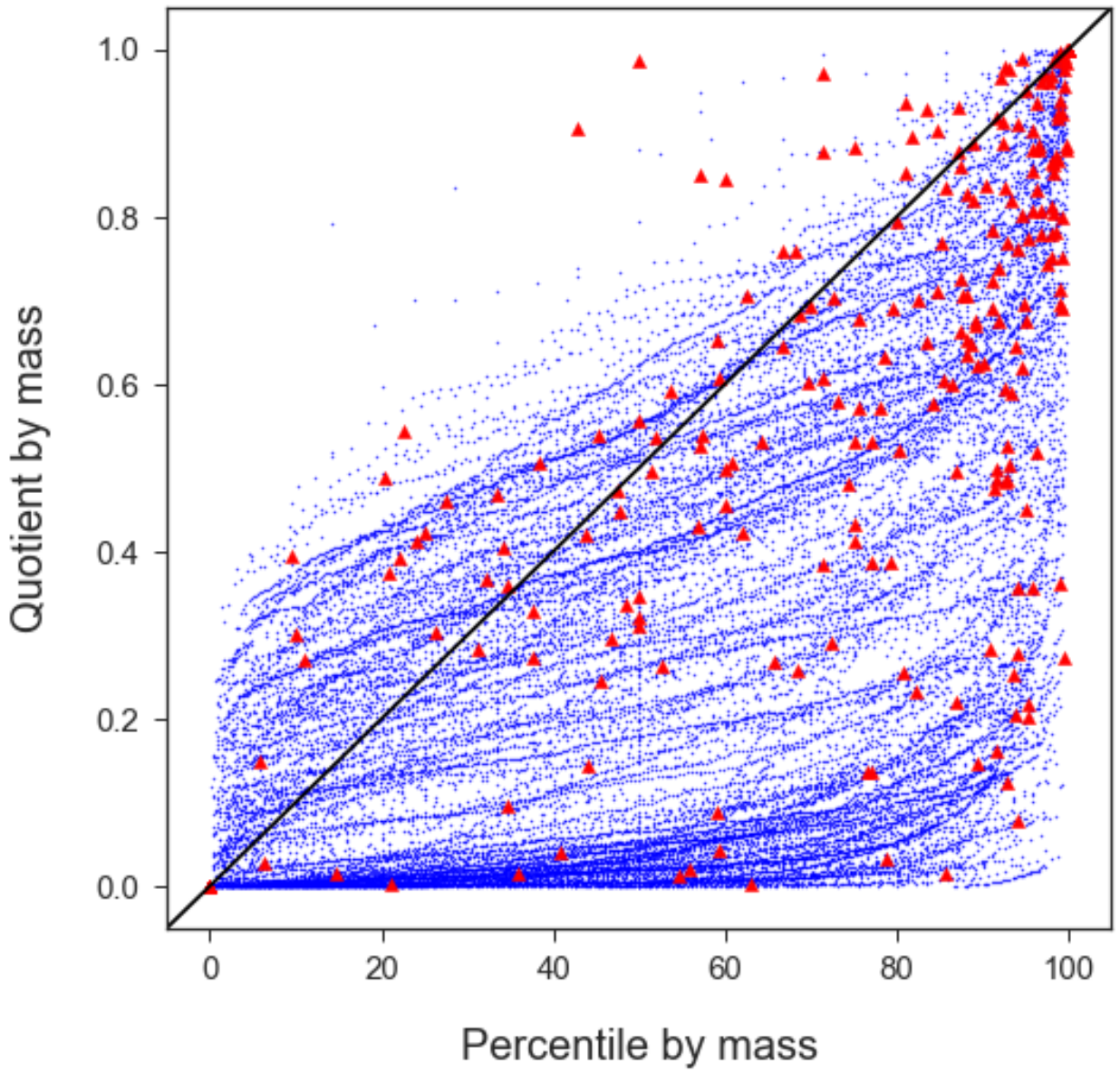


Figure S6. Scatterplot of quotient values vs. percentile values for all candidate compounds with a CFM-ID match score. The majority of values, for both "True Positive" and "Other Candidate" compounds, are below the diagonal line, indicating a less-than proportional increase in quotient values with rising percentile values. This trend reflects the uniform distribution of percentile values vs. right-skewed distribution of quotient values.

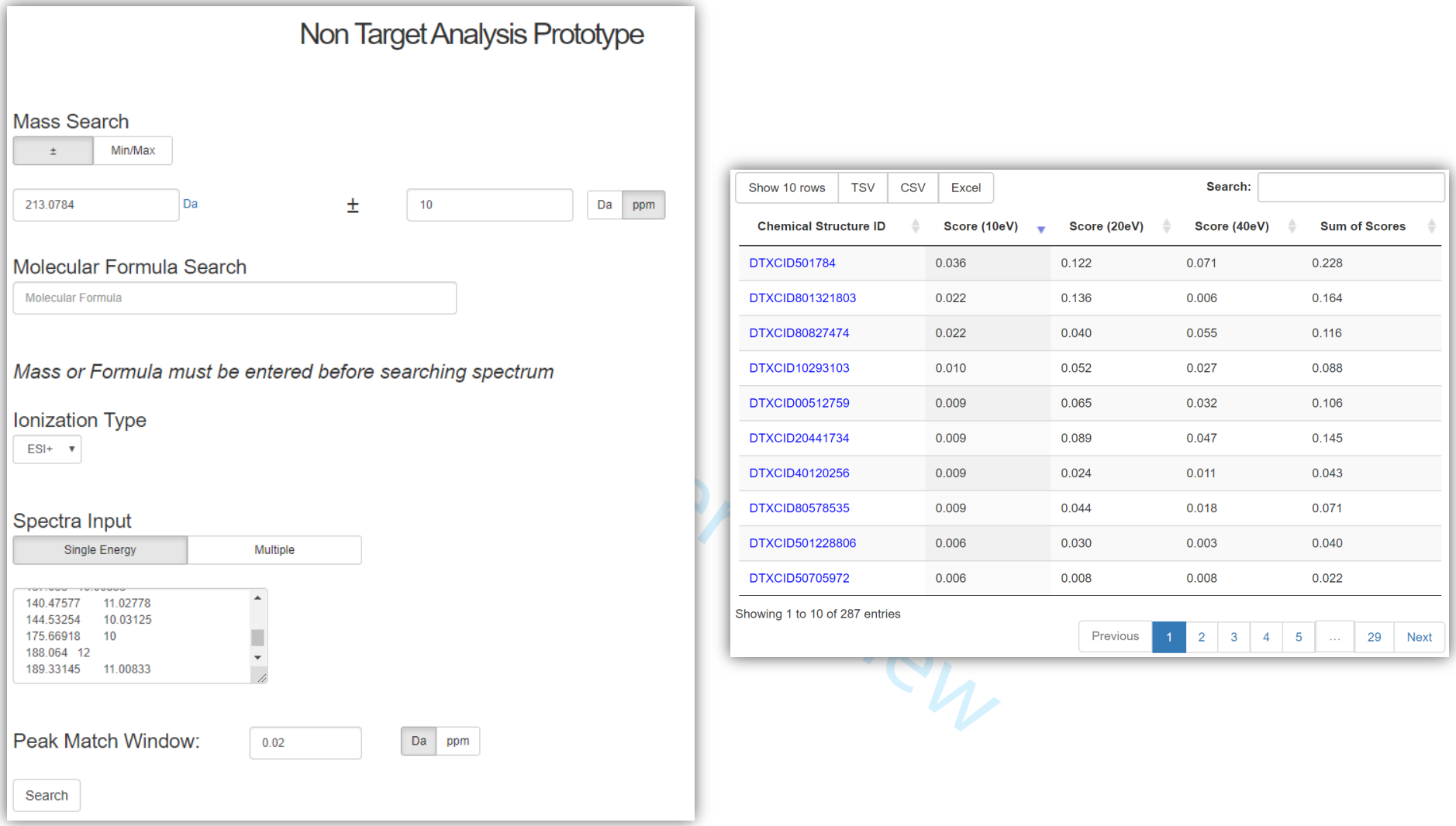
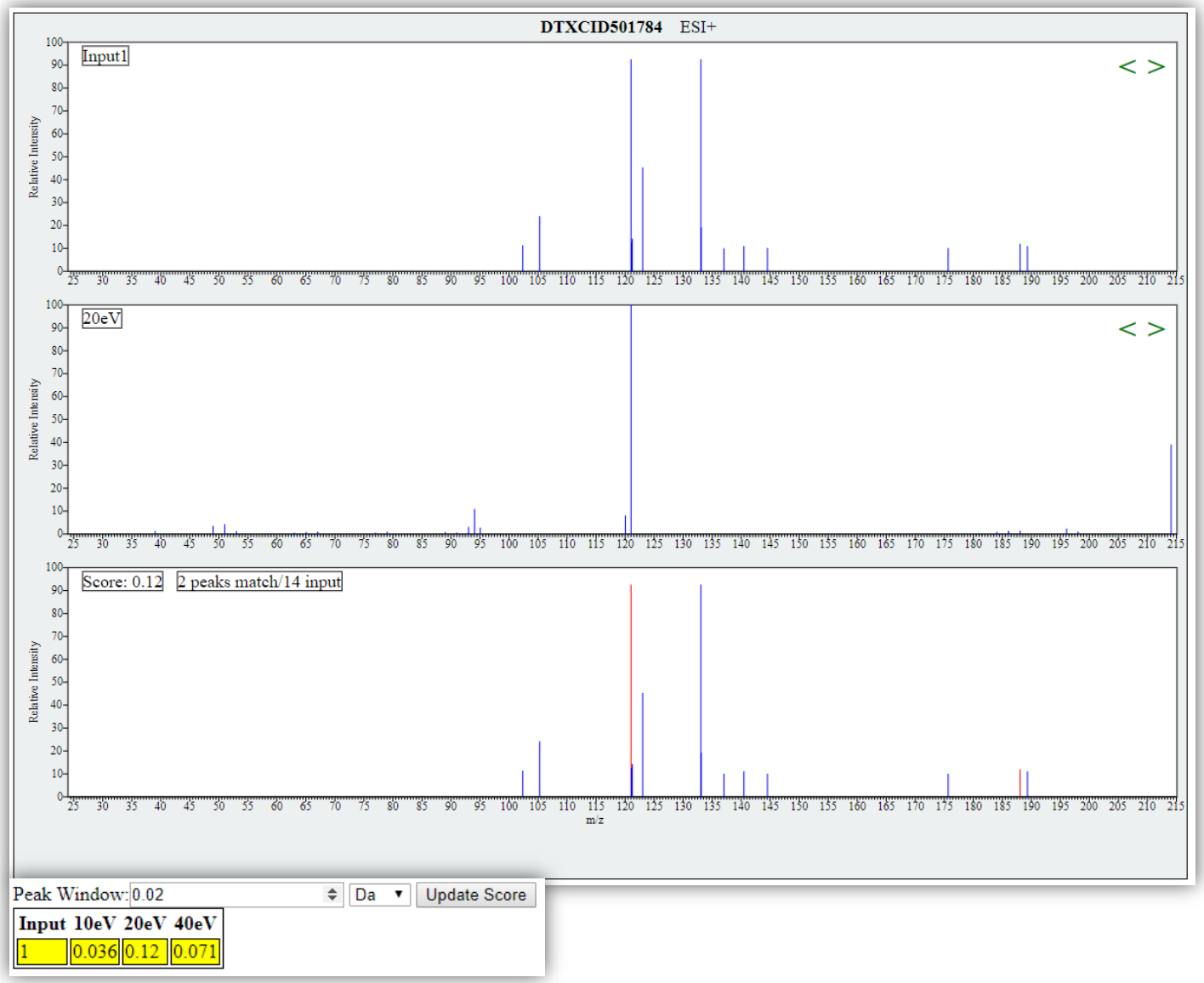
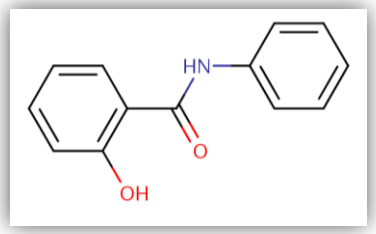


Figure S7A: Input page (left) and search results page (right) for a prototype web-based tool for searching experimental data against the CFM-ID database. The search requires: 1) an input neutral monoisotopic mass or molecular formula; and 2) experimental MS2 data in the format [fragment m/z, fragment intensity] for each line in the input field. One or multiple energies of experimental MS2 data may be entered. The result of the search is a table with all candidate compounds matched to the input mass or formula (identified by MS-Ready DTXCID) along with CFM-ID scores for each $CE_{in\ silico}$. Each candidate result is linked to a visualization page for the experimental and predicted spectra.





Salicylanilide

Figure S7B: Visualization page of experimental and predicted spectra for a candidate compound using the prototype web-based tool. Displayed on the page are the input experimental spectrum (top), predicted spectrum (middle), and spectrum comparison (bottom). The spectrum comparison is a copy of the input experimental spectrum with regards to spectral peaks and intensities; ions appearing as red are those which have been matched to an ion in the predicted spectrum. Each spectrum is scalable and scrollable through CE levels.

Note to the Editor: Addendum to Supplemental Table

Supplemental_Table_v4.xlsx should be included as an attached excel file and not incorporated into the manuscript PDF file for legibility reasons.

Supplemental Table 1. CFM-ID results for "pass" substances (n=377) from ENTACT mixtures (n=10). Data are

"Pass" Substance from ENTACT Mixture(s)	DTXSID
Pioglitazone hydrochloride	DTXSID3044203
Tomelukast	DTXSID7020344
Flutolanil	DTXSID8024109
Piperine	DTXSID3021805
Cybutryne	DTXSID3032416
Roxithromycin	DTXSID8041117
Triflumizole	DTXSID2032500
Atorvastatin calcium	DTXSID6044303
TDCPP	DTXSID9026261
Acetochlor	DTXSID8023848
Bispyribac-sodium	DTXSID7034383
Timolol maleate salt	DTXSID3047504
Anastrozole	DTXSID9022607
Fluroxypyr	DTXSID2034627
Melatonin	DTXSID1022421
Cyanazine	DTXSID1023990
Ethiofencarb	DTXSID3037545
Metolachlor	DTXSID4022448
Oxycarboxin	DTXSID8034792
Isazofos	DTXSID7034676
Diphenyl phthalate	DTXSID3021778
Terbutylazine	DTXSID4027608
Scopolamine hydrochloride	DTXSID6044692
Thiacloprid	DTXSID7034961
Pantoprazole sodium	DTXSID7044215
Triazophos	DTXSID9037612
Fenamiphos	DTXSID3024102
Trifloxystrobin	DTXSID4032580
Salicylanilide	DTXSID7021784
Famotidine	DTXSID5023039
Carbamazepine	DTXSID4022731
Tebufenozide	DTXSID4034948
Fipexide hydrochloride	DTXSID6047446
Bithionol	DTXSID9021342
Fluoxastrobin	DTXSID2034625
Azoxystrobin	DTXSID0032520
Glybenclamide	DTXSID0037237
Tebupirimfos	DTXSID1032482
Indoxacarb	DTXSID1032690
Propyzamide	DTXSID2020420
Thiamethoxam	DTXSID2034962
Thiophanate	DTXSID3034531
Fluazifop-butyl	DTXSID3034612
Deethylatrazine	DTXSID5037494
Clodinafop-propargyl	DTXSID6032354
Chlorbromuron	DTXSID6040290

Cladribine	DTXSID8022828
Dimetilan	DTXSID2041880
Imidacloprid	DTXSID5032442
Ifosfamide	DTXSID7020760
Doxylamine succinate	DTXSID7020552
Chlorpheniramine maleate	DTXSID4020321
Pendimethalin	DTXSID7024245
Buprofezin	DTXSID8034401
Amoxapine	DTXSID7022598
Bupirimate	DTXSID6041688
Pyriproxyfen	DTXSID1032640
Cloquintocet-mexyl	DTXSID3041794
Flamprop-isopropyl	DTXSID6041977
Methoxyfenozide	DTXSID3032628
Quizalofop-ethyl	DTXSID9023889
Quetiapine fumarate	DTXSID3044201
Pyraclostrobin	DTXSID7032638
Colchicine	DTXSID5024845
Fenpyroximate (Z,E)	DTXSID2032550
Nicardipine hydrochloride	DTXSID9046992
Hydramethylnon	DTXSID6023868
Ketoconazole	DTXSID7029879
Amiodarone hydrochloride	DTXSID7037185
Albendazole	DTXSID0022563
Flavone	DTXSID2022048
Cetirizine dihydrochloride	DTXSID2044268
Finasteride	DTXSID3020625
Tebuthiuron	DTXSID3024316
Azaconazole	DTXSID3041613
Benalaxyl	DTXSID3041619
Fipronil	DTXSID4034609
Thiobencarb	DTXSID6024337
Strychnine hemisulphate salt	DTXSID8021721
Bicalutamide	DTXSID2022678
Prometon	DTXSID6022341
Ioxynil	DTXSID8022161
Iprovalicarb	DTXSID0047662
Triphenyl phosphate	DTXSID1021952
Z-Tetrachlorvinphos	DTXSID1032648
Buspirone	DTXSID2022707
Acetyl tributyl citrate	DTXSID2026446
Hexazinone	DTXSID4024145
Prochloraz	DTXSID4024270
Dimethenamid	DTXSID4032376
Tris(2-butoxyethyl) phosphate	DTXSID5021758
Triflumuron	DTXSID5034355
Azinphos-ethyl	DTXSID5037498
Crotamiton	DTXSID6040664
Fenchlorazole-ethyl	DTXSID6041268
Fomesafen	DTXSID7024112

Spirotetramat	DTXSID7044342
Indapamide	DTXSID7044633
Bensulide	DTXSID9032329
Megestrol acetate	DTXSID9040683
Penoxsulam	DTXSID0034803
Meloxicam	DTXSID1020803
Methadone hydrochloride	DTXSID2020501
Fenbuconazole	DTXSID8032548
Chloramphenicol	DTXSID7020265
Simazine	DTXSID4021268
Praziquantel	DTXSID9021182
Dimethyl phthalate	DTXSID3022455
Etoxazole	DTXSID8034586
Pyraflufen-ethyl	DTXSID8034871
Pyrimethamine	DTXSID9021217
Kinetin	DTXSID9035175
Brucine	DTXSID2024662
Carbaryl	DTXSID9020247
Suxibuzone	DTXSID6021296
Tebufenpyrad	DTXSID0034223
Nuarimol	DTXSID2042220
Furalaxyl	DTXSID4047543
Propoxur	DTXSID7021948
Michler's ketone	DTXSID2020894
Venlafaxine hydrochloride	DTXSID8047397
Octabenzone	DTXSID9027441
Clomazone	DTXSID1032355
Triamterene	DTXSID6021373
Flusilazole	DTXSID3024235
Bezafibrate	DTXSID3029869
Napropamide	DTXSID5024211
Triamcinolone acetonide	DTXSID6021371
2-Methyl-4'-(methylthio)-2-morpholinopropiophenone	DTXSID8038857
4,4'-Sulfonyldiphenol	DTXSID3022409
Sulfasalazine	DTXSID0021256
Secbumeton	DTXSID8037594
3-Hydroxy-4-butyrophenetidide	DTXSID6020721
Rosiglitazone maleate	DTXSID2023569
Fluorescein	DTXSID0038887
Apigenin	DTXSID6022391
8-Methoxypsoralen	DTXSID8020830
Isoxaben	DTXSID8024159
Hexachlorophene	DTXSID6020690
Metamitron	DTXSID7047568
Ametryn	DTXSID1023869
Tamoxifen	DTXSID1034187
Acifluorfen	DTXSID0020022
Diclofenac sodium	DTXSID3037208
Probenecid	DTXSID9021188
Diethofencarb	DTXSID5037527

Propiconazole	DTXSID8024280
Diisopropyl phthalate	DTXSID2040731
Carboxin	DTXSID0023951
Aminocarb	DTXSID7022172
Norflurazon	DTXSID8024234
Monolinuron	DTXSID0037576
Leflunomide	DTXSID9023201
Sulfamethazine	DTXSID6021290
Boscalid	DTXSID6034392
2,4-Dinitrophenol	DTXSID0020523
5-Methoxypsoralen	DTXSID1025560
Thiabendazole	DTXSID0021337
Cyprodinil	DTXSID1032359
Oxyphenisatin	DTXSID5044528
Pirimicarb	DTXSID1032569
Dinoseb	DTXSID3020207
Fenarimol	DTXSID2032390
Myclobutanil	DTXSID8024315
17-Methyltestosterone	DTXSID1033664
Penconazole	DTXSID8042260
Phenolphthalein	DTXSID0021125
Vincamine	DTXSID9040134
Cortisone	DTXSID5022857
Icaridin	DTXSID0034227
4-Dimethylaminoantipyrine	DTXSID7020504
Ethylparaben	DTXSID9022528
Imazaquin	DTXSID3024152
Diphenamid	DTXSID8024072
Ethionamide	DTXSID0020577
Diethyl phthalate	DTXSID7021780
Chrysin	DTXSID1022396
Physostigmine	DTXSID3023471
Warfarin	DTXSID5023742
Chloridazon	DTXSID3034872
Fuberidazole	DTXSID4041995
4-Chlorophenoxyacetic acid	DTXSID9034282
E-Cinnamic acid	DTXSID5022489
Tacrine	DTXSID1037272
Butam	DTXSID5041691
5,7-Dimethoxy-2H-chromen-2-one	DTXSID1041421
3,3',5,5'-Tetramethylbenzidine	DTXSID5026120
Cinchophen	DTXSID0040705
Phenazone	DTXSID6021117
Benzylparaben	DTXSID9022526
Phenacetin	DTXSID1021116
2,4-Dichlorophenoxyacetic acid	DTXSID0020442
Oxazepam	DTXSID1021087
Sucralose	DTXSID1040245
Benodanil	DTXSID7041623
Dinoterb	DTXSID7041883

Fluazinam	DTXSID7032551
Thiodicarb	DTXSID0032578
Dinitramine	DTXSID9040265
Indinavir sulfate	DTXSID1044221
Budesonide	DTXSID8020202
Oryzalin	DTXSID8024238
Tetraconazole	DTXSID8034956
Carminic acid	DTXSID9022817
Sulfaquinoxaline	DTXSID8042424
Labetalol hydrochloride	DTXSID0044654
Thiophanate-methyl	DTXSID1024338
Chlorfenvinphos	DTXSID7034250
Tolclofos-methyl	DTXSID0034776
Prednisone	DTXSID4021185
Diclosulam	DTXSID4034528
Difenoconazole	DTXSID4032372
Fluroxypyr-meptyl	DTXSID5034303
Methylprednisolone	DTXSID7023300
Ethyl 3-(N-butylacetamido)propionate	DTXSID9035753
3,3',5,5'-Tetrabromobisphenol A	DTXSID1026081
Arbutin	DTXSID7040152
Halofenozide	DTXSID4032619
Hexaconazole	DTXSID4034653
Corticosterone	DTXSID6022474
Celecoxib	DTXSID0022777
Fluocinolone acetonide	DTXSID0040674
Buturon	DTXSID5041699
Triadimenol	DTXSID0032493
Triclocarban	DTXSID4026214
Bromoxynil	DTXSID3022162
Tebuconazole	DTXSID9032113
Isoprocarb	DTXSID6042072
Dibutyl phthalate	DTXSID2021781
Imazethapyr	DTXSID3024287
2,4-Dihydroxybenzophenone	DTXSID8022406
Dapsone	DTXSID4020371
Norgestrel	DTXSID3036496
17alpha-Hydroxyprogesterone	DTXSID6040747
Monuron	DTXSID0020311
6-Methyl coumarin	DTXSID9025588
4-Methylumbelliferone	DTXSID8025670
Coumatetralyl	DTXSID8041799
Daidzein	DTXSID9022310
Milrinone	DTXSID5023324
Biochanin A	DTXSID1022394
Fenuron	DTXSID7037551
Propylparaben	DTXSID4022527
Dicloran	DTXSID2020426
Diuron	DTXSID0020446
MEHP	DTXSID2025680

Troglitazone	DTXSID8023719
4-Nitrosodiphenylamine	DTXSID1021031
CP-457677	DTXSID2047278
N,N'-Dicyclohexylthiourea	DTXSID9020451
Pirinixic acid	DTXSID4020290
Ingliforib	DTXSID4047252
SB281832	DTXSID5047324
MK-274	DTXSID5047328
2-[4-(Diethylamino)-2-hydroxybenzoyl]benzoic acid	DTXSID4027604
1,3-Dipropan-2-ylurea	DTXSID6044486
Flufenpyr-ethyl	DTXSID3034618
PharmaGSID_47333	DTXSID4047333
3-Hydroxy-N-(3-nitrophenyl)naphthalene-2-carboxamide	DTXSID3044544
PharmaGSID_48172	DTXSID1048172
AVE3295	DTXSID5047372
UK-416244	DTXSID0047290
Thiazopyr	DTXSID1032488
YM218	DTXSID1048176
Candoxatril	DTXSID6047286
N-Benzyl-9-(tetrahydro-2H-pyran-2-yl)adenine	DTXSID8038801
PharmaGSID_48518	DTXSID9048518
Methyl 2-methoxybenzoate	DTXSID5047087
Procyazine	DTXSID1034844
2-Amino-N-cyclohexyl-N-methylbenzenesulfonamide	DTXSID1044982
CP-457920	DTXSID4047254
SAR102608	DTXSID8047391
MK-578	DTXSID0047327
Aplaviroc hydrochloride	DTXSID1047316
GW473178E methyl benzene sulphonic acid	DTXSID6047313
Trimethyl benzene-1,2,4-tricarboxylate	DTXSID8044620
CI-1044	DTXSID5047291
1,3-Diphenylguanidine	DTXSID3025178
Lauryldiethanolamine	DTXSID4042090
Di(2-methoxyethyl) phthalate	DTXSID8025094
CP-122721	DTXSID9047251
CP-114271	DTXSID2047274
PD 0343701	DTXSID7047271
MK-812	DTXSID4047335
Surinabant	DTXSID2047357
PharmaGSID_48511	DTXSID4048511
SAR102779	DTXSID4047387
CP-863187	DTXSID0047294
PharmaGSID_48505	DTXSID0048505
Prodiamine	DTXSID1034210
1,1'-Disulfanediyldiazepan-2-one	DTXSID1044481
CP-728663	DTXSID1047283
Farglitazar	DTXSID1047310
SAR115740	DTXSID1047366
Ethyl phthalyl ethyl glycolate	DTXSID3024100
PFOSA	DTXSID3038939

GSK163929B	DTXSID6047311
FR140423	DTXSID6048175
CP-532623	DTXSID7047279
SAR377142	DTXSID4047385
PharmaGSID_48519	DTXSID4048519
Cyclanilide	DTXSID5032600
SB236057A	DTXSID5047320
SB243213A	DTXSID5047322
PharmaGSID_48516	DTXSID9048516
CI-959	DTXSID8047268
Volinanserin	DTXSID6047363
HMR1171 trifluoroacetate (1:1)	DTXSID3048522
Nelivaptan	DTXSID7047358
PharmaGSID_48506	DTXSID5048506
Carabersat	DTXSID6047319
3-Hydroxy-2-naphthoic acid	DTXSID3026560
N-(4-Ethoxyphenyl)-3-oxobutanamide	DTXSID5044526
5-Chlorosalicylanilide	DTXSID9037749
FR130739	DTXSID1048178
CI-1018	DTXSID0047248
SSR126768	DTXSID0047379
Butylphthalyl butylglycolate	DTXSID7023938
Undecanedioic acid	DTXSID0044862
SSR161421	DTXSID5047374
6-Phenyl-1,3,5-triazine-2,4-diamine	DTXSID1020142
N-(3,4-Dichlorophenyl)-N'-methylurea	DTXSID3042180
Parinol	DTXSID4042256
N-Phenyldiethanolamine	DTXSID5021962
3,3'-Dimethoxybenzidine	DTXSID3025091
Tribromsalan	DTXSID4026181
Ilepatril	DTXSID7047356
Dipropyl 2,5-pyridinedicarboxylate	DTXSID8032544
C.I. Basic Red 9 monohydrochloride	DTXSID1021247
4-Amino-2,5-dimethoxy-N-phenylbenzenesulfonamide	DTXSID2044686
1-(2,6-Dichlorophenyl)-2-indolinone	DTXSID6046979
N-(2,4-Dimethylphenyl)acetamide	DTXSID0044868
Phenethyl anthranilate	DTXSID5025861
2,5-Dichloro-3-nitrobenzoic acid	DTXSID3041372
Bis(2-butoxyethyl) phthalate	DTXSID9047174
2,2'-(3-Chlorophenylimino)diethanol	DTXSID4026557
Imazamox	DTXSID3034664
Methiuron	DTXSID7042140
7-(Dimethylamino)-4-methylcoumarin	DTXSID6041422
Benzo(f)quinoline	DTXSID2024585
2-(3-Phenylpropyl)pyridine	DTXSID7042356
N-Ethyl-4-menthane-3-carboxamide	DTXSID5047039
Ancymidol	DTXSID2034338
4-(3-Phenylpropyl)pyridine	DTXSID5044869
Octahydro-1H-4,7-methanoindene-1,5-diyldimethanediyl bisprop-2-enoate	DTXSID5044940
N-Dodecanoyl-N-methylglycine	DTXSID7042011

2-(N-Ethyl-m-toluidino)ethanol	DTXSID3021801
N,N-Dimethyldecylamine oxide	DTXSID7042190
Ethyl 5-oxo-1-phenyl-4,5-dihydro-1H-pyrazole-3-carboxylate	DTXSID5044738
Dipropyl phthalate	DTXSID5031133
9-Ethyl-3-nitro-9H-carbazole	DTXSID0044995
2-Isopropyl-6-methyl-4-pyrimidone	DTXSID1027502
CP-401387	DTXSID9047253
Fabesetron hydrochloride	DTXSID7048168
Dodecylamine hydrochloride	DTXSID9044322
Diisobutyl phthalate	DTXSID9022522
N-Octyl-2-pyrrolidone	DTXSID4036435
Triallyl trimellitate	DTXSID4044901
Chloranocryl	DTXSID2020424
N-Benzyladenine	DTXSID7032630
4-Chlorophenylurea	DTXSID5041512
7-Diethylamino-4-methylcoumarin	DTXSID9025035
1-Cyclohexylpyrrolidin-2-one	DTXSID7044716
6-Ethoxy-2,3,4-trimethyl-1,2,3,4-tetrahydroquinoline	DTXSID6042365
4-Methoxy-2-methyl-N-phenylaniline	DTXSID3029364
Nootkatone	DTXSID8047050
1-(4-Methoxyphenyl)-1-pentene-3-one	DTXSID4047412
m-Cumenyl methylcarbamate	DTXSID1040324
1-Naphthol	DTXSID6021793
Dichlormid	DTXSID4027997
4-Aminoazobenzene	DTXSID6024460
N-Butyl-p-toluenesulfonamide	DTXSID7042198
N,N'-Dibutylthiourea	DTXSID8042187
Di(ethylene glycol) dimethacrylate	DTXSID8044882
2,2',6,6'-Tetrachlorobisphenol A	DTXSID3021770
Endosulfan sulfate	DTXSID3037541
Neopentyl glycol dibenzoate	DTXSID1038822
Mean	
Median	

Acronym Legend:

CFM-ID: Competitive Fragmentation Modeling-ID
ENTACT: EPA's Non-Targeted Analysis Collaborative Trial
DTXSID: DSSTox Substance Identifier
DTXCID: DSSTox Chemical Identifier
DSSTox: Distributed Structure-Searchable Toxicity Database
PCDL: Personal Compound Database and Library

* - 4-Nitrosodiphenylamine was not "In PCDL" based on automated matching (using CASRN and name) but wa

e sorted by "In PCDL", "PCDL Matched", and "Approach 3 Rank (by Formula)".

MS-Ready DTXCID	In PCDL	PCDL Matched	Approach 1 (CE=10) Score
DTXCID1017129	Y	Y	2.25E-02
DTXCID30344	Y	Y	2.73E-02
DTXCID204109	Y	Y	3.57E-01
DTXCID00809656	Y	Y	4.07E-01
DTXCID1012416	Y	Y	6.06E-01
DTXCID90810337	Y	Y	1.71E-02
DTXCID60810288	Y	Y	6.46E-03
DTXCID80197003	Y	Y	1.76E-03
DTXCID206261	Y	Y	2.88E-03
DTXCID303848	Y	Y	4.40E-02
DTXCID8023977	Y	Y	2.46E-02
DTXCID70197219	Y	Y	5.26E-01
DTXCID202607	Y	Y	5.76E-02
DTXCID0014627	Y	Y	4.24E-03
DTXCID002421	Y	Y	3.19E-01
DTXCID803990	Y	Y	3.50E-01
DTXCID1017545	Y	Y	7.08E-02
DTXCID402448	Y	Y	3.03E-01
DTXCID6014792	Y	Y	2.55E-01
DTXCID5014676	Y	Y	2.22E-01
DTXCID801778	Y	Y	8.03E-02
DTXCID107608	Y	Y	4.02E-01
DTXCID90196688	Y	Y	4.25E-01
DTXCID90810397	Y	Y	0.00E+00
DTXCID803416	Y	Y	8.84E-02
DTXCID7017612	Y	Y	3.38E-01
DTXCID404102	Y	Y	1.59E-01
DTXCID20810441	Y	Y	1.62E-01
DTXCID501784	Y	Y	6.49E-02
DTXCID803039	Y	Y	1.28E-01
DTXCID902731	Y	Y	2.83E-01
DTXCID2014948	Y	Y	1.52E-01
DTXCID3024657	Y	Y	3.66E-01
DTXCID701342	Y	Y	5.32E-05
DTXCID2032057	Y	Y	8.15E-02
DTXCID90810433	Y	Y	5.57E-02
DTXCID8017237	Y	Y	2.07E-02
DTXCID9012482	Y	Y	1.85E-02
DTXCID1020801	Y	Y	8.98E-07
DTXCID80420	Y	Y	1.40E-02
DTXCID0014962	Y	Y	3.18E-06
DTXCID1014531	Y	Y	3.77E-02
DTXCID1014612	Y	Y	4.35E-04
DTXCID3017494	Y	Y	2.11E-01
DTXCID1065215	Y	Y	1.05E-03
DTXCID4020290	Y	Y	9.72E-02

DTXCID60209219	Y	Y	1.17E-01
DTXCID0021880	Y	Y	3.93E-01
DTXCID3012442	Y	Y	1.47E-01
DTXCID90760	Y	Y	2.51E-01
DTXCID102970	Y	Y	3.07E-01
DTXCID402804	Y	Y	3.74E-01
DTXCID004245	Y	Y	9.05E-03
DTXCID6014401	Y	Y	1.80E-01
DTXCID102598	Y	Y	4.85E-01
DTXCID4021688	Y	Y	2.98E-01
DTXCID9012640	Y	Y	9.41E-02
DTXCID1021794	Y	Y	2.70E-01
DTXCID4021977	Y	Y	1.38E-01
DTXCID1012628	Y	Y	1.15E-01
DTXCID903889	Y	Y	2.05E-01
DTXCID903546	Y	Y	6.08E-01
DTXCID5012638	Y	Y	4.07E-02
DTXCID00207037	Y	Y	3.99E-01
DTXCID0065307	Y	Y	4.88E-01
DTXCID90197195	Y	Y	4.23E-01
DTXCID60810263	Y	Y	3.63E-01
DTXCID90196446	Y	Y	0.00E+00
DTXCID702592	Y	Y	4.59E-01
DTXCID202563	Y	Y	4.10E-02
DTXCID602048	Y	Y	2.99E-01
DTXCID802787	Y	Y	2.47E-01
DTXCID60209052	Y	Y	4.75E-01
DTXCID704316	Y	Y	1.84E-01
DTXCID1021613	Y	Y	1.62E-01
DTXCID1021619	Y	Y	3.16E-01
DTXCID2014609	Y	Y	6.85E-02
DTXCID704337	Y	Y	1.02E-01
DTXCID80196697	Y	Y	0.00E+00
DTXCID00209197	Y	Y	2.16E-01
DTXCID402341	Y	Y	5.70E-01
DTXCID602161	Y	Y	1.50E-06
DTXCID50210253	Y	Y	2.69E-02
DTXCID201952	Y	Y	3.87E-01
DTXCID301320	Y	Y	3.21E-02
DTXCID402707	Y	Y	1.70E-01
DTXCID006446	Y	Y	3.72E-02
DTXCID804145	Y	Y	1.12E-01
DTXCID904270	Y	Y	3.60E-03
DTXCID2012376	Y	Y	2.09E-01
DTXCID201758	Y	Y	4.36E-01
DTXCID3014355	Y	Y	7.56E-03
DTXCID3017498	Y	Y	5.87E-03
DTXCID4020664	Y	Y	4.97E-01
DTXCID4021268	Y	Y	8.91E-02
DTXCID704112	Y	Y	2.97E-01

DTXCID40209893	Y	Y	6.50E-02
DTXCID5024633	Y	Y	1.75E-03
DTXCID7012329	Y	Y	5.22E-02
DTXCID90208871	Y	Y	1.38E-01
DTXCID8014803	Y	Y	2.04E-01
DTXCID30208734	Y	Y	1.55E-03
DTXCID003273	Y	Y	1.75E-02
DTXCID6012548	Y	Y	3.45E-01
DTXCID00208665	Y	Y	2.16E-05
DTXCID501268	Y	Y	6.67E-05
DTXCID501182	Y	Y	
DTXCID502455	Y	Y	
DTXCID6014586	Y	Y	
DTXCID6014871	Y	Y	
DTXCID601217	Y	Y	
DTXCID7015175	Y	Y	1.77E-03
DTXCID50197434	Y	Y	8.27E-01
DTXCID10247	Y	Y	1.65E-01
DTXCID601296	Y	Y	1.59E-03
DTXCID8014223	Y	Y	4.09E-01
DTXCID0022220	Y	Y	2.14E-01
DTXCID2027543	Y	Y	1.04E-01
DTXCID301948	Y	Y	1.07E-01
DTXCID80894	Y	Y	3.76E-01
DTXCID403737	Y	Y	4.30E-01
DTXCID107441	Y	Y	4.02E-01
DTXCID9012355	Y	Y	2.89E-01
DTXCID001373	Y	Y	7.34E-01
DTXCID704235	Y	Y	4.08E-01
DTXCID909869	Y	Y	1.90E-01
DTXCID504211	Y	Y	5.33E-03
DTXCID60209113	Y	Y	5.55E-04
DTXCID6018857	Y	Y	1.30E-02
DTXCID602409	Y	Y	2.07E-01
DTXCID80809755	Y	Y	1.16E-01
DTXCID6017594	Y	Y	4.83E-01
DTXCID40721	Y	Y	3.56E-02
DTXCID5017131	Y	Y	5.47E-01
DTXCID8018887	Y	Y	1.53E-03
DTXCID902391	Y	Y	6.84E-01
DTXCID40830	Y	Y	8.71E-01
DTXCID704159	Y	Y	1.02E-01
DTXCID40690	Y	Y	5.94E-03
DTXCID5027568	Y	Y	8.12E-01
DTXCID303869	Y	Y	5.03E-01
DTXCID20809976	Y	Y	5.50E-01
DTXCID6022	Y	Y	5.27E-07
DTXCID802923	Y	Y	2.48E-01
DTXCID901188	Y	Y	7.13E-03
DTXCID3017527	Y	Y	1.61E-01

DTXCID204280	Y	Y	5.46E-02
DTXCID0020731	Y	Y	3.93E-02
DTXCID003951	Y	Y	1.29E-02
DTXCID302172	Y	Y	1.39E-01
DTXCID70810092	Y	Y	5.85E-01
DTXCID8017576	Y	Y	2.60E-01
DTXCID103201	Y	Y	8.36E-04
DTXCID201290	Y	Y	6.27E-03
DTXCID4014392	Y	Y	1.21E-03
DTXCID80523	Y	Y	0.00E+00
DTXCID305560	Y	Y	1.35E-01
DTXCID401337	Y	Y	1.88E-04
DTXCID9012359	Y	Y	5.87E-01
DTXCID3024528	Y	Y	2.73E-04
DTXCID9012569	Y	Y	3.82E-01
DTXCID30207	Y	Y	2.59E-04
DTXCID0012390	Y	Y	3.70E-06
DTXCID304315	Y	Y	2.10E-01
DTXCID20208864	Y	Y	4.81E-01
DTXCID6022260	Y	Y	2.02E-01
DTXCID901125	Y	Y	4.72E-03
DTXCID30199092	Y	Y	2.59E-02
DTXCID90209222	Y	Y	1.13E-01
DTXCID8014227	Y	Y	2.67E-02
DTXCID70504	Y	Y	4.16E-04
DTXCID002528	Y	Y	2.19E-02
DTXCID904152	Y	Y	1.34E-01
DTXCID304072	Y	Y	3.35E-01
DTXCID00577	Y	Y	3.97E-01
DTXCID901780	Y	Y	2.81E-02
DTXCID902396	Y	Y	0.00E+00
DTXCID40209318	Y	Y	2.84E-02
DTXCID20196838	Y	Y	2.83E-01
DTXCID10809809	Y	Y	6.53E-01
DTXCID2021995	Y	Y	5.59E-01
DTXCID7014282	Y	Y	1.58E-01
DTXCID9065316	Y	Y	5.62E-02
DTXCID9017272	Y	Y	8.66E-01
DTXCID3021691	Y	Y	4.16E-01
DTXCID9021421	Y	Y	4.42E-04
DTXCID406120	Y	Y	1.27E-01
DTXCID8020705	Y	Y	3.39E-01
DTXCID401117	Y	Y	8.60E-01
DTXCID202526	Y	Y	2.07E-06
DTXCID001116	Y	Y	3.24E-01
DTXCID80442	Y	Y	0.00E+00
DTXCID301087	Y	Y	0.00E+00
DTXCID00209612	Y	Y	0.00E+00
DTXCID5021623	Y	Y	0.00E+00
DTXCID5021883	Y	Y	0.00E+00

DTXCID5012551	Y	Y	0.00E+00
DTXCID50810232	Y	Y	0.00E+00
DTXCID7020265	Y		2.21E-01
DTXCID80196713	Y		3.46E-03
DTXCID80209019	Y		4.53E-02
DTXCID904238	Y		2.62E-01
DTXCID6014956	Y		3.52E-02
DTXCID00209218	Y		6.65E-02
DTXCID6022424	Y		4.58E-01
DTXCID603191	Y		3.68E-01
DTXCID104338	Y		9.58E-02
DTXCID5014250	Y		3.39E-02
DTXCID8014776	Y		9.47E-06
DTXCID40208866	Y		1.62E-03
DTXCID2014528	Y		9.75E-03
DTXCID2012372	Y		1.12E-01
DTXCID3014303	Y		1.12E-02
DTXCID70209285	Y		3.80E-04
DTXCID7015753	Y		2.20E-02
DTXCID406081	Y		0.00E+00
DTXCID80209610	Y		3.94E-04
DTXCID2012619	Y		0.00E+00
DTXCID2014653	Y		2.01E-01
DTXCID50209162	Y		5.50E-02
DTXCID502777	Y		0.00E+00
DTXCID50209667	Y		3.52E-03
DTXCID3021699	Y		0.00E+00
DTXCID8012493	Y		0.00E+00
DTXCID906214	Y		0.00E+00
DTXCID002162	Y		5.48E-01
DTXCID7012113	Y		2.27E-01
DTXCID4022072	Y		9.40E-02
DTXCID301781	Y		1.88E-02
DTXCID004287	Y		4.88E-01
DTXCID402406	Y		4.77E-01
DTXCID00371	Y		8.16E-03
DTXCID90208770	Y		4.39E-02
DTXCID80209675	Y		1.99E-02
DTXCID80311	Y		1.35E-06
DTXCID105588	Y		9.14E-04
DTXCID805670	Y		0.00E+00
DTXCID00209798	Y		
DTXCID102310	Y		4.84E-07
DTXCID00209293	Y		3.61E-01
DTXCID102394	Y		
DTXCID5017551	Y		5.87E-02
DTXCID602527	Y		0.00E+00
DTXCID60426	Y		0.00E+00
DTXCID00446	Y		0.00E+00
DTXCID105680	Y		0.00E+00

DTXCID603719	Y		0.00E+00
DTXCID401031		Y*	0.00E+00
DTXCID0027278			2.16E-01
DTXCID70451			9.60E-02
DTXCID00290			9.53E-02
DTXCID50210197			2.24E-02
DTXCID3027324			3.00E-01
DTXCID8027329			1.31E-02
DTXCID507604			2.14E-04
DTXCID4024486			6.07E-01
DTXCID1014618			1.57E-01
DTXCID40210222			6.94E-01
DTXCID1024544			3.78E-03
DTXCID2028147			6.13E-01
DTXCID50197212			1.96E-01
DTXCID8027290			1.76E-01
DTXCID9012488			4.83E-01
DTXCID40810483			1.80E-01
DTXCID30210211			7.48E-03
DTXCID6018801			5.77E-01
DTXCID301065038			3.39E-02
DTXCID3027087			1.35E-01
DTXCID9014844			6.46E-04
DTXCID9024982			1.27E-01
DTXCID70210199			2.07E-01
DTXCID6027391			4.92E-01
DTXCID80210221			5.62E-02
DTXCID20197204			5.90E-02
DTXCID60197203			5.51E-01
DTXCID6024620			1.16E-01
DTXCID90210212			2.26E-01
DTXCID005178			9.66E-01
DTXCID2022090			7.07E-01
DTXCID805094			4.35E-02
DTXCID90210196			4.36E-01
DTXCID0027274			4.11E-01
DTXCID1027392			7.69E-01
DTXCID40197206			4.82E-01
DTXCID0027357			4.89E-01
DTXCID7028485			5.31E-01
DTXCID30197215			5.58E-01
DTXCID8027294			6.84E-01
DTXCID3028479			5.27E-02
DTXCID9014210			8.64E-02
DTXCID9024481			2.49E-01
DTXCID00210208			2.58E-01
DTXCID50210218			7.63E-02
DTXCID9027366			1.29E-01
DTXCID604100			7.85E-02
DTXCID1018939			1.53E-01

DTXCID4027311	9.16E-04
DTXCID2028149	3.76E-01
DTXCID80210206	1.83E-03
DTXCID7027386	5.36E-02
DTXCID1028493	0.00E+00
DTXCID3012600	1.89E-01
DTXCID8027321	4.78E-04
DTXCID8027323	1.01E-01
DTXCID601065037	1.48E-01
DTXCID1027269	7.47E-02
DTXCID70210230	6.03E-02
DTXCID4027367	
DTXCID00210228	
DTXCID30210276	
DTXCID20210220	0.00E+00
DTXCID406560	3.10E-01
DTXCID3024526	4.79E-01
DTXCID7017749	7.12E-03
DTXCID6028153	8.45E-04
DTXCID30210195	2.82E-01
DTXCID70197214	2.52E-01
DTXCID203938	1.40E-02
DTXCID8024862	5.61E-03
DTXCID3027374	3.66E-01
DTXCID50142	4.86E-01
DTXCID1022180	6.36E-02
DTXCID2022256	1.31E-01
DTXCID501962	6.88E-02
DTXCID605091	6.95E-03
DTXCID606181	
DTXCID40210227	2.99E-02
DTXCID6012544	2.68E-01
DTXCID3023845	1.46E-02
DTXCID0024686	7.20E-03
DTXCID4026979	2.09E-01
DTXCID8024868	7.47E-02
DTXCID305861	6.18E-02
DTXCID1021372	0.00E+00
DTXCID7027174	
DTXCID906557	1.18E-04
DTXCID1014664	4.11E-01
DTXCID5022140	4.88E-01
DTXCID4021422	1.87E-03
DTXCID804585	5.41E-01
DTXCID5022356	7.08E-03
DTXCID3027039	7.72E-03
DTXCID0014338	2.80E-03
DTXCID3024869	5.05E-01
DTXCID3024940	1.18E-03
DTXCID5022011	

DTXCID101801	4.20E-01
DTXCID5022190	1.99E-03
DTXCID3024738	6.81E-02
DTXCID3011133	5.48E-03
DTXCID8024995	3.62E-02
DTXCID60209431	8.94E-01
DTXCID10210198	4.08E-02
DTXCID80197245	5.90E-01
DTXCID901984	3.64E-01
DTXCID602522	6.98E-04
DTXCID2016435	6.78E-01
DTXCID2024901	2.30E-03
DTXCID00424	
DTXCID5012630	3.11E-05
DTXCID3021512	9.38E-02
DTXCID405035	3.38E-01
DTXCID5024716	3.00E-01
DTXCID4022365	5.80E-02
DTXCID909364	4.09E-01
DTXCID20210184	3.00E-02
DTXCID2027412	2.97E-01
DTXCID9020324	0.00E+00
DTXCID401793	0.00E+00
DTXCID007997	0.00E+00
DTXCID40809630	0.00E+00
DTXCID5022198	0.00E+00
DTXCID6022187	0.00E+00
DTXCID6024882	0.00E+00
DTXCID601770	0.00E+00
DTXCID1017541	0.00E+00
DTXCID9018822	0.00E+00
	1.83E-01
	8.91E-02

as "PCDL Matched" based on manual evaluation.

Approach 1 (CE=20) Score	Approach 1 (CE=40) Score	Approach 2 (CE=10) Score
1.06E-01	8.24E-02	1.08E-01
1.14E-02	0.00E+00	6.26E-02
5.53E-02	1.31E-01	6.90E-01
1.70E-01	4.15E-02	6.35E-01
2.48E-01	1.17E-01	1.09E+00
5.49E-02	6.42E-03	3.12E-02
1.78E-02	0.00E+00	3.93E-02
1.33E-03	2.01E-02	6.20E-03
0.00E+00	0.00E+00	1.99E-02
0.00E+00	0.00E+00	1.42E-01
2.82E-02	6.44E-03	7.68E-02
6.79E-04	0.00E+00	7.25E-01
6.93E-02	1.54E-02	5.90E-02
0.00E+00	0.00E+00	2.11E-02
3.83E-01	2.33E-01	1.09E+00
1.14E-01	6.43E-02	4.65E-01
8.94E-02	4.73E-02	2.21E-01
1.65E-01	1.60E-01	5.48E-01
2.12E-02	6.19E-03	2.96E-01
5.60E-02	9.13E-03	3.39E-01
6.23E-02	1.46E-02	2.58E-01
2.70E-01	7.02E-02	7.35E-01
2.45E-01	6.29E-02	7.02E-01
2.21E-01	9.47E-02	0.00E+00
2.30E-01	1.20E-01	2.87E-01
2.95E-02	5.59E-02	4.47E-01
1.60E-01	8.30E-03	3.31E-01
5.08E-02	6.24E-02	3.47E-01
1.22E-01	0.00E+00	9.43E-02
9.18E-02	5.46E-02	1.99E-01
1.61E-01	7.22E-02	8.26E-01
2.45E-01	3.10E-01	5.17E-01
1.59E-01	1.70E-01	8.05E-01
3.55E-03	2.64E-02	5.47E-02
3.93E-02	2.38E-03	1.63E-01
9.37E-02	6.41E-03	1.21E-01
1.02E-01	1.92E-02	5.00E-02
7.12E-02	1.43E-04	3.41E-02
2.19E-04	2.63E-04	2.61E-05
6.25E-02	0.00E+00	8.49E-02
7.74E-05	1.99E-02	7.18E-05
3.67E-04	4.08E-03	5.20E-02
1.98E-02	1.46E-02	1.94E-03
1.24E-01	3.29E-02	4.64E-01
1.23E-03	2.38E-04	4.20E-03
1.45E-04	0.00E+00	1.63E-01

1.53E-01	6.17E-03	3.62E-01
1.95E-02	1.08E-04	5.21E-01
2.31E-02	3.91E-02	2.27E-01
3.05E-02	2.16E-03	2.74E-01
1.71E-01	2.25E-02	7.84E-01
2.19E-01	7.57E-02	8.24E-01
5.81E-03	5.46E-03	4.73E-02
4.59E-02	1.36E-02	2.97E-01
3.09E-01	4.39E-02	9.40E-01
8.12E-02	7.17E-03	4.14E-01
2.20E-02	1.25E-02	1.77E-01
1.01E-01	2.18E-02	4.36E-01
6.41E-02	5.60E-02	4.28E-01
1.43E-01		3.73E-01
6.90E-02	1.27E-02	3.17E-01
1.81E-01	2.34E-02	9.00E-01
1.07E-02	3.07E-03	1.05E-01
2.85E-01	9.23E-02	5.62E-01
1.92E-01	5.01E-02	7.08E-01
1.23E-01	5.51E-02	5.07E-01
1.73E-01	1.15E-01	5.15E-01
0.00E+00	2.05E-02	0.00E+00
1.10E-01	1.97E-02	5.58E-01
8.32E-02	2.98E-02	2.09E-01
1.46E-01	7.47E-03	6.27E-01
5.41E-02	1.17E-02	4.41E-01
1.08E-01	6.34E-03	6.02E-01
2.78E-01	1.12E-02	4.98E-01
9.96E-03	6.63E-05	1.97E-01
1.13E-01	3.64E-02	6.11E-01
2.12E-03	6.17E-06	1.18E-01
2.79E-02	2.54E-05	2.18E-01
1.83E-01	1.84E-04	0.00E+00
2.63E-01	3.42E-01	6.50E-01
3.60E-01	4.82E-02	9.67E-01
5.88E-01	7.15E-04	7.35E-05
2.22E-02	3.57E-02	9.24E-02
0.00E+00	0.00E+00	7.25E-01
7.95E-04	8.19E-04	4.53E-02
7.48E-02	4.48E-03	2.36E-01
2.06E-04	2.11E-05	6.53E-02
8.78E-03	9.15E-04	1.95E-01
6.03E-03	0.00E+00	1.69E-02
1.52E-01	6.86E-02	4.24E-01
2.46E-03	0.00E+00	8.45E-01
2.68E-02	1.00E-02	2.51E-02
7.74E-04	1.14E-02	5.36E-02
1.61E-01	5.88E-02	8.76E-01
4.14E-02	7.48E-04	1.29E-01
0.00E+00	0.00E+00	5.84E-01

2.97E-02	2.27E-04	1.34E-01
9.17E-02	9.21E-03	2.12E-02
3.06E-02	9.81E-03	1.30E-01
4.11E-02	6.06E-03	2.60E-01
2.10E-02	1.79E-03	2.68E-01
1.01E-02	1.33E-04	1.61E-02
8.18E-03	3.63E-05	7.33E-02
1.15E-02	2.48E-02	3.92E-01
5.10E-04	9.42E-05	4.45E-03
8.76E-02	0.00E+00	8.95E-04
7.98E-02	1.08E-02	
3.13E-01	2.22E-01	
4.02E-02	4.74E-03	
4.15E-02		
	3.10E-02	
1.32E-01	1.54E-01	1.64E-01
5.33E-01	3.31E-02	1.39E+00
3.02E-01	9.43E-02	5.38E-01
	1.03E-04	1.50E-02
1.33E-01	2.32E-02	4.90E-01
1.45E-01	1.08E-04	4.67E-01
6.50E-02	0.00E+00	2.34E-01
1.12E-01	1.58E-02	5.06E-01
1.49E-01		6.85E-01
5.90E-02	1.54E-03	6.69E-01
1.64E-01	4.82E-02	5.74E-01
1.25E-01	1.06E-01	6.56E-01
4.72E-01	9.35E-02	1.55E+00
2.52E-02	2.93E-04	5.12E-01
5.74E-02	0.00E+00	3.01E-01
1.70E-02	2.45E-04	4.64E-02
3.57E-03	5.63E-04	2.66E-03
7.23E-02	2.76E-03	6.31E-02
0.00E+00	0.00E+00	4.23E-01
2.65E-02	2.81E-05	1.99E-01
2.31E-01	2.47E-02	7.53E-01
3.46E-02	6.00E-02	2.45E-01
1.37E-01	0.00E+00	6.38E-01
4.75E-02	1.14E-01	4.54E-02
6.48E-01	7.56E-03	1.20E+00
3.02E-01	9.97E-02	1.50E+00
2.06E-01	1.41E-01	4.48E-01
4.22E-03	0.00E+00	6.67E-02
3.48E-01	2.82E-02	1.56E+00
2.76E-01	1.54E-02	6.96E-01
1.10E-01	1.67E-02	7.04E-01
0.00E+00	1.17E-04	1.59E-04
5.04E-02	4.67E-02	7.66E-01
3.90E-02	0.00E+00	2.33E-02
1.42E-02	1.33E-03	4.27E-01

2.50E-03	3.94E-05	5.59E-02
1.76E-02	1.20E-02	1.36E-01
2.98E-02	0.00E+00	4.21E-02
4.29E-02	3.72E-02	3.27E-01
8.46E-02	7.82E-05	6.19E-01
4.92E-02	4.83E-02	4.49E-01
1.76E-02	1.12E-01	4.97E-03
6.14E-03	1.79E-02	2.94E-02
1.48E-07	0.00E+00	1.09E-02
	7.93E-04	0.00E+00
2.00E-02	0.00E+00	2.53E-01
6.49E-01	8.41E-02	1.97E-02
4.20E-01	2.55E-02	1.04E+00
5.02E-03	5.03E-03	5.29E-03
1.76E-02	8.31E-04	5.00E-01
2.99E-02	1.76E-02	5.05E-02
3.61E-04	0.00E+00	1.17E-04
2.61E-03	2.19E-04	2.15E-01
3.67E-02	6.26E-03	6.91E-01
3.56E-05	0.00E+00	2.13E-01
4.76E-02	7.66E-02	6.83E-02
4.32E-02	0.00E+00	6.23E-02
3.68E-03	7.68E-05	1.36E-01
3.53E-03	0.00E+00	1.00E-01
4.25E-03	0.00E+00	3.30E-03
6.80E-02		1.22E-01
1.30E-01	1.16E-02	1.72E-01
4.14E-02	2.65E-02	5.87E-01
9.10E-03	3.23E-03	6.21E-01
5.31E-03	2.94E-02	1.17E-01
0.00E+00	4.58E-03	0.00E+00
2.63E-03	2.35E-02	7.05E-02
5.52E-02	4.28E-03	5.68E-01
5.25E-01	2.30E-03	1.22E+00
2.44E-01	0.00E+00	1.34E+00
		2.90E-01
0.00E+00		2.85E-01
3.56E-01	3.99E-02	1.63E+00
3.04E-02	0.00E+00	5.63E-01
7.63E-02	2.31E-02	6.58E-03
		2.39E-01
2.52E-03	1.65E-03	5.14E-01
4.72E-01	4.47E-02	1.59E+00
1.65E-04		3.16E-04
0.00E+00	0.00E+00	5.29E-01
0.00E+00	0.00E+00	0.00E+00
0.00E+00	0.00E+00	0.00E+00
	0.00E+00	0.00E+00
0.00E+00	0.00E+00	0.00E+00
0.00E+00	0.00E+00	0.00E+00

0.00E+00	0.00E+00	0.00E+00
0.00E+00	0.00E+00	0.00E+00
1.45E-01	0.00E+00	4.45E-01
6.15E-03	4.88E-03	2.08E-02
1.46E-02	2.49E-04	1.04E-01
4.15E-02	0.00E+00	4.34E-01
1.28E-04	0.00E+00	3.86E-02
1.74E-02	8.91E-04	8.59E-02
2.62E-01	1.29E-02	8.67E-01
1.67E-01	2.21E-02	5.96E-01
1.47E-02	3.15E-03	1.84E-01
3.92E-03	4.31E-03	5.96E-02
0.00E+00	9.68E-05	6.74E-04
5.64E-03	1.56E-03	9.33E-03
1.84E-03	1.43E-04	1.15E-02
3.70E-03	9.49E-04	1.18E-01
1.98E-02	0.00E+00	5.82E-02
1.41E-02	0.00E+00	3.98E-03
1.01E-03	4.85E-05	7.51E-02
0.00E+00	7.05E-06	0.00E+00
0.00E+00	1.45E-05	8.11E-04
0.00E+00	1.64E-02	0.00E+00
6.59E-03	0.00E+00	2.08E-01
5.59E-02	5.51E-04	1.07E-01
3.54E-02		0.00E+00
6.46E-05	3.66E-06	1.03E-02
6.04E-02	0.00E+00	0.00E+00
0.00E+00	1.22E-06	0.00E+00
2.46E-02	0.00E+00	0.00E+00
0.00E+00	0.00E+00	1.47E+00
0.00E+00	0.00E+00	2.42E-01
0.00E+00	0.00E+00	2.13E-01
3.93E-03	9.64E-03	8.08E-02
2.17E-01	3.44E-02	6.75E-01
3.82E-03	0.00E+00	9.06E-01
0.00E+00	0.00E+00	3.18E-02
1.66E-02	1.64E-02	6.84E-02
5.83E-03	1.13E-04	3.00E-02
0.00E+00	0.00E+00	1.66E-03
6.33E-03	5.14E-02	1.34E-02
2.12E-02	7.87E-02	0.00E+00
1.12E-01	9.34E-02	
9.94E-06	9.57E-03	2.42E-03
7.27E-02	2.21E-04	5.33E-01
	6.56E-02	
0.00E+00	0.00E+00	7.81E-02
1.38E-01	0.00E+00	0.00E+00
0.00E+00		0.00E+00
0.00E+00	0.00E+00	0.00E+00
0.00E+00	0.00E+00	0.00E+00

0.00E+00	0.00E+00	0.00E+00
0.00E+00	0.00E+00	0.00E+00
2.94E-01	1.16E-01	5.68E-01
0.00E+00	0.00E+00	1.70E-01
9.27E-02	1.89E-02	1.94E-01
6.93E-03	9.55E-03	3.59E-02
1.01E-01	6.56E-02	4.53E-01
1.12E-01	1.15E-03	2.33E-02
2.40E-01	4.55E-03	1.80E-03
0.00E+00	0.00E+00	9.78E-01
1.64E-02	0.00E+00	1.86E-01
1.70E-01	1.46E-01	1.15E+00
4.25E-02	3.10E-01	2.07E-01
1.53E-01	6.02E-02	8.49E-01
3.43E-02	2.60E-02	2.43E-01
3.18E-04	8.82E-04	2.78E-01
1.83E-01	3.09E-02	6.82E-01
5.43E-02	8.91E-03	2.22E-01
1.72E-02	6.06E-03	1.32E-02
3.61E-01	1.14E-03	1.02E+00
5.01E-03	0.00E+00	3.30E-01
1.19E-01	0.00E+00	3.77E-01
6.05E-02	5.31E-03	1.07E-02
2.00E-02	1.98E-05	2.47E-01
1.96E-02	5.50E-04	2.95E-01
8.60E-02	8.72E-03	7.32E-01
1.42E-02	2.29E-03	8.11E-02
2.87E-02	1.88E-02	9.60E-02
1.74E-01	6.90E-02	7.23E-01
2.32E-01	3.74E-03	2.27E-01
8.20E-02	1.58E-02	3.79E-01
4.73E-01	1.63E-01	2.02E+00
2.09E-01	5.37E-03	9.56E-01
9.03E-03	1.89E-03	1.02E-01
3.06E-01	8.68E-02	6.28E-01
2.09E-01	6.08E-02	5.90E-01
2.15E-01	1.45E-01	9.42E-01
1.87E-01	2.65E-02	6.27E-01
3.01E-01	6.65E-02	7.14E-01
1.88E-01	3.44E-02	7.04E-01
1.05E-01	2.22E-02	6.78E-01
2.52E-01	1.16E-04	1.04E+00
4.93E-02	0.00E+00	7.37E-02
1.55E-01	0.00E+00	1.65E-01
2.93E-03	8.05E-04	4.70E-01
1.33E-01	6.50E-02	4.19E-01
2.15E-03	4.76E-04	1.05E-01
2.37E-02	4.41E-02	1.80E-01
1.00E-02	6.54E-03	2.42E-01
6.47E-01	3.60E-03	8.22E-01

0.00E+00	1.63E-04	1.20E-03
2.25E-02	7.34E-04	6.52E-01
4.72E-05	0.00E+00	3.67E-03
2.07E-04	1.33E-03	5.92E-02
2.98E-02	7.99E-04	0.00E+00
2.97E-03	5.09E-02	4.57E-01
5.00E-08	2.13E-04	2.01E-03
7.31E-02	1.15E-04	1.33E-01
6.98E-02	0.00E+00	2.07E-01
4.14E-02	1.99E-02	1.80E-01
1.51E-01	2.85E-01	8.87E-02
2.17E-03	1.29E-04	
1.53E-02		
	3.86E-03	
2.91E-04	0.00E+00	0.00E+00
5.81E-01	4.96E-01	1.66E+00
3.24E-01	1.30E-01	9.95E-01
2.11E-03	0.00E+00	4.70E-02
5.94E-02	5.61E-02	1.30E-02
3.77E-02	2.38E-03	3.69E-01
7.88E-02	1.44E-02	3.17E-01
1.07E-04	1.16E-02	3.38E-02
5.09E-02		5.53E-02
2.14E-02	4.05E-02	4.54E-01
1.27E-02	0.00E+00	9.75E-01
2.13E-01	0.00E+00	3.43E-01
4.39E-02	4.46E-03	2.14E-01
1.17E-01	8.13E-02	5.36E-01
1.21E-01	7.24E-03	5.46E-02
3.54E-02		
1.27E-04	9.85E-05	3.93E-02
1.62E-01	1.67E-06	5.98E-01
7.06E-02	2.22E-01	1.39E-01
1.09E-02	7.05E-03	2.51E-02
1.46E-01	4.75E-04	3.48E-01
2.53E-02	2.05E-02	4.02E-01
1.31E-01	1.35E-04	3.17E-01
3.22E-02	0.00E+00	0.00E+00
2.40E-03		
2.52E-05	4.89E-04	5.62E-04
		5.70E-01
4.17E-02	0.00E+00	7.84E-01
1.48E-02	1.37E-01	1.60E-01
5.79E-01		1.23E+00
6.15E-02	0.00E+00	3.87E-02
2.51E-03	0.00E+00	1.61E-02
9.17E-03	2.23E-04	7.04E-03
0.00E+00	9.87E-02	7.98E-01
1.53E-03	8.54E-05	4.75E-03
1.42E-04	0.00E+00	

8.89E-02	6.83E-02	9.46E-01
0.00E+00	0.00E+00	1.19E-02
2.22E-01	5.24E-02	3.40E-01
4.00E-03	1.49E-02	4.86E-02
1.08E-05	3.54E-02	7.16E-02
4.66E-01	1.20E-02	1.48E+00
5.77E-04	9.35E-04	5.12E-02
1.04E-01	0.00E+00	7.79E-01
1.15E-01	2.96E-03	5.88E-01
7.03E-03	3.98E-03	1.26E-02
1.53E-01	0.00E+00	9.30E-01
3.61E-06	0.00E+00	4.78E-03
2.44E-05		
8.82E-02	3.67E-02	7.33E-03
0.00E+00	0.00E+00	1.39E-01
1.54E-01	3.61E-02	5.21E-01
0.00E+00	3.14E-04	4.36E-01
1.33E-02	2.05E-03	1.50E-01
3.46E-02	6.60E-03	7.33E-01
0.00E+00	0.00E+00	5.45E-02
3.50E-03	0.00E+00	4.60E-01
1.43E-04	0.00E+00	0.00E+00
0.00E+00	0.00E+00	0.00E+00
0.00E+00	0.00E+00	0.00E+00
0.00E+00	0.00E+00	0.00E+00
0.00E+00	0.00E+00	0.00E+00
0.00E+00	0.00E+00	0.00E+00
0.00E+00	0.00E+00	0.00E+00
0.00E+00	0.00E+00	0.00E+00
0.00E+00	0.00E+00	0.00E+00
0.00E+00	0.00E+00	0.00E+00

8.52E-02	2.70E-02	3.30E-01
3.14E-02	3.92E-03	1.97E-01

Approach 2 (CE=20) Score	Approach 2 (CE=40) Score	Approach 3 Score	Approach 1 (CE=10) Rank (by Mass)
6.07E-01	1.60E-01	8.76E-01	1
2.83E-02	0.00E+00	9.09E-02	1
1.09E-01	2.71E-01	1.07E+00	1
4.00E-01	6.80E-02	1.10E+00	3
4.30E-01	3.51E-01	1.87E+00	1
1.22E-01	3.11E-02	1.84E-01	1
2.49E-02	0.00E+00	6.43E-02	1
2.39E-03	2.52E-02	3.38E-02	1
0.00E+00	0.00E+00	1.99E-02	2
0.00E+00	0.00E+00	1.42E-01	2
5.79E-02	1.77E-02	1.52E-01	1
1.82E-03	0.00E+00	7.26E-01	9
1.77E-01	1.56E-02	2.51E-01	5
0.00E+00	0.00E+00	2.11E-02	2
8.18E-01	3.00E-01	2.21E+00	2
2.59E-01	1.19E-01	8.43E-01	94
2.26E-01	9.69E-02	5.44E-01	1
3.85E-01	2.71E-01	1.20E+00	1
7.41E-02	3.20E-02	4.02E-01	1
1.44E-01	2.67E-02	5.10E-01	1
1.17E-01	3.20E-02	4.08E-01	2
4.83E-01	1.53E-01	1.37E+00	3
6.68E-01	2.17E-01	1.59E+00	13
5.35E-01	1.99E-01	7.34E-01	
4.74E-01	1.82E-01	9.43E-01	1
1.29E-01	1.03E-01	6.78E-01	1
2.39E-01	1.94E-02	5.90E-01	19
2.41E-01	8.78E-02	6.76E-01	16
2.28E-01	0.00E+00	3.23E-01	76
2.82E-01	8.66E-02	5.67E-01	1
3.82E-01	1.06E-01	1.31E+00	3
6.41E-01	5.80E-01	1.74E+00	1
4.21E-01	3.68E-01	1.59E+00	1
4.46E-02	2.90E-02	1.28E-01	1
6.95E-02	5.52E-03	2.38E-01	8
2.58E-01	1.26E-02	3.92E-01	1
2.42E-01	3.88E-02	3.31E-01	1
1.09E-01	6.02E-04	1.44E-01	1
4.46E-04	7.21E-04	1.19E-03	4
1.57E-01	0.00E+00	2.42E-01	2
1.15E-04	2.45E-02	2.47E-02	9
2.38E-03	4.96E-03	5.94E-02	6
3.38E-02	4.44E-02	8.02E-02	1
1.99E-01	4.57E-02	7.08E-01	1
2.19E-03	9.44E-04	7.33E-03	2
4.10E-04	0.00E+00	1.63E-01	1

3.51E-01	8.51E-03	7.21E-01	1
6.12E-02	1.30E-04	5.82E-01	1
7.28E-02	9.36E-02	3.94E-01	2
1.71E-01	7.75E-03	4.53E-01	60
3.22E-01	4.30E-02	1.15E+00	1
3.98E-01	1.75E-01	1.40E+00	1
3.05E-02	9.05E-03	8.69E-02	5
9.26E-02	4.31E-02	4.33E-01	2
5.99E-01	7.40E-02	1.61E+00	1
2.40E-01	1.41E-02	6.68E-01	24
4.30E-02	2.27E-02	2.43E-01	1
1.65E-01	6.23E-02	6.64E-01	8
1.91E-01	9.38E-02	7.13E-01	1
2.99E-01		6.73E-01	1
1.61E-01	2.69E-02	5.05E-01	5
4.72E-01	3.31E-02	1.41E+00	5
1.77E-02	7.56E-03	1.30E-01	8
8.15E-01	1.49E-01	1.53E+00	8
4.96E-01	1.28E-01	1.33E+00	1
4.50E-01	2.46E-01	1.20E+00	4
5.55E-01	6.85E-01	1.75E+00	3
0.00E+00	2.50E-01	2.50E-01	
5.80E-01	4.59E-01	1.60E+00	1
1.22E-01	3.89E-02	3.70E-01	3
3.63E-01	7.84E-03	9.98E-01	3
1.32E-01	2.42E-02	5.97E-01	11
3.88E-01	3.36E-02	1.02E+00	1
5.29E-01	2.39E-02	1.05E+00	1
1.46E-02	1.42E-04	2.12E-01	1
2.28E-01	6.19E-02	9.00E-01	1
3.89E-03	6.17E-06	1.21E-01	3
7.99E-02	1.47E-03	2.99E-01	1
4.24E-01	1.85E-04	4.25E-01	
5.51E-01	6.39E-01	1.84E+00	1
7.53E-01	8.54E-02	1.81E+00	1
1.62E+00	7.89E-04	1.62E+00	1
7.05E-02	6.33E-02	2.26E-01	1
0.00E+00	0.00E+00	7.25E-01	5
1.43E-02	3.96E-03	6.35E-02	1
2.32E-01	6.07E-03	4.74E-01	1
3.18E-04	4.52E-05	6.56E-02	1
2.49E-02	1.04E-03	2.21E-01	1
8.83E-03	0.00E+00	2.57E-02	1
3.34E-01	1.06E-01	8.63E-01	1
6.44E-03	0.00E+00	8.51E-01	2
6.35E-02	1.82E-02	1.07E-01	1
1.01E-03	2.15E-02	7.61E-02	1
3.48E-01	8.83E-02	1.31E+00	1
9.39E-02	7.48E-04	2.24E-01	1
0.00E+00	0.00E+00	5.84E-01	4

6.09E-02	3.48E-04	1.96E-01	1
1.66E-01	1.06E-02	1.98E-01	1
1.29E-01	3.95E-02	2.98E-01	1
1.16E-01	1.43E-02	3.91E-01	2
9.72E-02	5.12E-03	3.70E-01	2
1.72E-02	5.10E-04	3.38E-02	1
2.01E-02	2.49E-04	9.37E-02	1
3.99E-02	3.95E-02	4.71E-01	31
1.60E-03	1.16E-04	6.17E-03	6
1.84E-01	0.00E+00	1.85E-01	240
1.73E-01	1.70E-02	1.90E-01	
7.03E-01	3.30E-01	1.03E+00	
1.48E-01	6.18E-03	1.54E-01	
1.00E-01		1.00E-01	
	3.72E-02	3.72E-02	
3.35E-01	2.00E-01	6.99E-01	29
1.33E+00	2.63E-01	2.98E+00	6
6.39E-01	1.44E-01	1.32E+00	2
	1.33E-04	1.51E-02	2
3.94E-01	3.49E-02	9.18E-01	9
3.04E-01	1.18E-04	7.71E-01	15
1.05E-01	0.00E+00	3.39E-01	2
1.35E-01	4.13E-02	6.83E-01	2
3.05E-01		9.90E-01	3
1.39E-01	2.88E-03	8.11E-01	4
4.20E-01	8.17E-02	1.08E+00	14
5.57E-01	2.74E-01	1.49E+00	2
1.09E+00	9.95E-02	2.74E+00	5
1.19E-01	4.45E-03	6.36E-01	1
1.04E-01	0.00E+00	4.05E-01	11
2.64E-02	6.81E-03	7.96E-02	2
8.37E-03	2.73E-03	1.38E-02	3
1.93E-01	4.32E-03	2.60E-01	2
0.00E+00	0.00E+00	4.23E-01	4
4.69E-02	5.87E-04	2.46E-01	13
6.26E-01	4.34E-02	1.42E+00	3
1.12E-01	7.24E-02	4.29E-01	6
2.23E-01	0.00E+00	8.61E-01	17
1.15E-01	1.75E-01	3.36E-01	27
1.55E+00	9.18E-03	2.76E+00	7
6.38E-01	1.11E-01	2.25E+00	2
4.28E-01	3.40E-01	1.22E+00	4
1.44E-02	0.00E+00	8.11E-02	4
6.56E-01	3.36E-02	2.25E+00	3
6.13E-01	3.02E-02	1.34E+00	3
4.45E-01	3.11E-02	1.18E+00	4
0.00E+00	1.20E-04	2.79E-04	7
1.36E-01	6.16E-02	9.64E-01	4
5.97E-02	0.00E+00	8.30E-02	20
6.60E-02	2.34E-03	4.96E-01	2

6.42E-02	4.54E-05	1.20E-01	15
5.10E-02	2.15E-02	2.08E-01	4
4.47E-02	0.00E+00	8.68E-02	5
1.04E-01	5.50E-02	4.86E-01	2
1.44E-01	3.20E-03	7.67E-01	19
1.32E-01	7.22E-02	6.53E-01	2
9.86E-02	1.55E-01	2.59E-01	5
1.67E-02	3.15E-02	7.77E-02	6
3.54E-05	0.00E+00	1.10E-02	6
	1.73E-03	1.73E-03	
5.82E-02	0.00E+00	3.11E-01	6
1.36E+00	1.16E-01	1.50E+00	34
9.63E-01	6.94E-02	2.08E+00	8
6.14E-03	9.90E-03	2.13E-02	12
3.80E-02	1.21E-03	5.39E-01	7
1.07E-01	2.11E-02	1.79E-01	48
5.28E-04	0.00E+00	6.44E-04	12
1.48E-02	5.40E-04	2.30E-01	89
1.16E-01	1.32E-02	8.20E-01	5
1.11E-03	0.00E+00	2.14E-01	30
1.01E-01	1.47E-01	3.16E-01	11
1.31E-01	0.00E+00	1.93E-01	20
1.00E-02	9.42E-05	1.46E-01	19
1.12E-02	0.00E+00	1.11E-01	11
1.50E-02	0.00E+00	1.83E-02	14
2.11E-01		3.33E-01	16
2.47E-01	1.17E-02	4.31E-01	39
1.08E-01	3.49E-02	7.30E-01	51
2.54E-02	5.20E-03	6.52E-01	79
2.90E-02	4.55E-02	1.92E-01	9
0.00E+00	4.94E-03	4.94E-03	
6.09E-02	2.42E-02	1.56E-01	26
7.53E-02	9.97E-02	7.43E-01	69
1.14E+00	9.15E-03	2.37E+00	31
4.67E-01	0.00E+00	1.80E+00	37
		2.90E-01	41
0.00E+00		2.85E-01	16
8.10E-01	3.24E-01	2.77E+00	20
1.05E-01	0.00E+00	6.69E-01	52
1.91E-01	3.69E-02	2.35E-01	91
		2.39E-01	69
6.49E-03	1.72E-03	5.23E-01	85
1.02E+00	5.32E-02	2.66E+00	68
1.04E-03		1.36E-03	144
0.00E+00	0.00E+00	5.29E-01	303
0.00E+00	0.00E+00	0.00E+00	
0.00E+00	0.00E+00	0.00E+00	
	0.00E+00	0.00E+00	
0.00E+00	0.00E+00	0.00E+00	
0.00E+00	0.00E+00	0.00E+00	

0.00E+00	0.00E+00	0.00E+00	
0.00E+00	0.00E+00	0.00E+00	
3.02E-01	0.00E+00	7.47E-01	64
2.23E-02	1.12E-02	5.43E-02	1
3.64E-02	6.69E-04	1.42E-01	1
1.02E-01	0.00E+00	5.36E-01	39
4.20E-03	0.00E+00	4.28E-02	14
3.83E-02	1.88E-03	1.26E-01	3
4.22E-01	3.96E-02	1.33E+00	6
3.42E-01	5.57E-02	9.94E-01	1
4.09E-02	4.08E-03	2.29E-01	1
3.14E-02	7.18E-03	9.83E-02	1
0.00E+00	3.59E-04	1.03E-03	3
2.18E-02	3.29E-03	3.44E-02	5
1.32E-02	7.05E-03	3.18E-02	1
5.18E-03	1.14E-03	1.24E-01	5
2.51E-02	0.00E+00	8.32E-02	3
4.12E-02	0.00E+00	4.51E-02	4
1.57E-03	1.14E-04	7.67E-02	3
0.00E+00	7.95E-06	7.95E-06	
0.00E+00	3.60E-04	1.17E-03	6
0.00E+00	2.59E-02	2.59E-02	
5.74E-02	0.00E+00	2.65E-01	38
2.12E-01	1.46E-03	3.20E-01	3
1.01E-01		1.01E-01	
3.78E-04	5.86E-05	1.08E-02	1
2.12E-01	0.00E+00	2.12E-01	
0.00E+00	2.68E-04	2.68E-04	
7.83E-02	0.00E+00	7.83E-02	
0.00E+00	0.00E+00	1.47E+00	6
0.00E+00	0.00E+00	2.42E-01	14
0.00E+00	0.00E+00	2.13E-01	3
1.62E-02	1.21E-02	1.09E-01	12
4.66E-01	6.13E-02	1.20E+00	74
1.51E-02	0.00E+00	9.21E-01	2
0.00E+00	0.00E+00	3.18E-02	15
5.78E-02	2.02E-02	1.46E-01	27
1.58E-02	1.50E-04	4.59E-02	19
0.00E+00	0.00E+00	1.66E-03	63
2.01E-02	6.66E-02	1.00E-01	33
2.41E-01	8.82E-02	2.17E+00	
2.48E-01	1.21E-01	3.68E-01	
5.39E-04	3.16E-02	3.46E-02	77
2.39E-01	2.69E-04	7.73E-01	64
	1.22E-01	1.22E-01	
0.00E+00	0.00E+00	7.81E-02	35
5.95E-01	0.00E+00	5.95E-01	
0.00E+00		0.00E+00	
0.00E+00	0.00E+00	0.00E+00	
0.00E+00	0.00E+00	0.00E+00	

0.00E+00	0.00E+00	0.00E+00	
0.00E+00	0.00E+00	0.00E+00	
6.97E-01	3.46E-01	1.61E+00	1
0.00E+00	0.00E+00	1.70E-01	1
2.56E-01	2.38E-02	4.75E-01	2
1.78E-02	2.15E-02	7.51E-02	1
2.72E-01	8.51E-02	8.10E-01	16
2.44E-01	1.29E-01	3.96E-01	5
6.38E-01	7.96E-03	6.48E-01	1
0.00E+00	0.00E+00	9.78E-01	1
1.01E-01	0.00E+00	2.87E-01	7
3.41E-01	4.99E-01	4.99E-01	1
3.27E-01	3.61E-01	8.95E-01	3
3.28E-01	2.58E-01	1.44E+00	1
1.22E-01	5.34E-02	4.18E-01	2
3.43E-04	1.59E-03	2.79E-01	1
5.50E-01	5.80E-02	1.29E+00	15
2.95E-01	2.27E-02	5.40E-01	2
4.77E-02	1.29E-02	7.39E-02	1
8.31E-01	1.62E-03	1.85E+00	1
7.10E-03	0.00E+00	3.37E-01	1
2.36E-01	0.00E+00	6.13E-01	11
7.24E-02	6.28E-03	8.94E-02	64
3.08E-02	1.61E-04	2.78E-01	1
4.71E-02	1.71E-03	3.44E-01	1
2.04E-01	5.75E-02	9.93E-01	2
4.12E-02	1.26E-02	1.35E-01	1
9.59E-02	4.42E-02	2.36E-01	1
6.71E-01	1.02E-01	1.50E+00	6
4.86E-01	8.28E-02	7.96E-01	1
2.11E-01	3.13E-02	6.21E-01	8
1.38E+00	2.08E-01	3.61E+00	1
6.84E-01	1.26E-02	1.65E+00	1
2.01E-02	2.62E-03	1.25E-01	3
7.84E-01	2.42E-01	1.65E+00	12
5.34E-01	1.11E-01	1.24E+00	22
9.79E-01	3.01E-01	2.22E+00	1
6.50E-01	8.89E-02	1.37E+00	1
8.03E-01	2.29E-01	1.75E+00	2
5.53E-01	3.01E-01	1.56E+00	1
4.47E-01	6.76E-02	1.19E+00	1
7.48E-01	1.77E-03	1.79E+00	1
1.76E-01	0.00E+00	2.50E-01	7
3.32E-01	0.00E+00	4.97E-01	49
7.96E-03	6.76E-03	4.84E-01	1
3.08E-01	2.20E-01	9.47E-01	1
5.17E-03	1.29E-02	1.23E-01	1
6.91E-02	8.19E-02	3.31E-01	5
5.59E-02	7.89E-03	3.06E-01	1
8.60E-01	1.52E-02	1.70E+00	1

0.00E+00	2.22E-03	3.42E-03	1
5.13E-02	1.36E-03	7.05E-01	10
8.32E-05	0.00E+00	3.75E-03	1
8.30E-04	3.65E-03	6.37E-02	7
1.21E-01	1.51E-03	1.23E-01	
2.64E-02	6.25E-02	5.46E-01	1
1.17E-05	1.13E-03	3.15E-03	2
1.84E-01	1.87E-03	3.19E-01	1
2.63E-01	0.00E+00	4.70E-01	12
4.84E-02	2.71E-02	2.55E-01	2
4.29E-01	5.72E-01	1.00E+00	2
3.03E-03	1.12E-03	4.14E-03	
4.21E-02		4.21E-02	
	1.17E-02	1.17E-02	
5.24E-04	0.00E+00	5.24E-04	
1.49E+00	1.01E+00	4.17E+00	5
6.40E-01	2.48E-01	1.88E+00	2
6.21E-03	0.00E+00	5.32E-02	1
1.23E-01	7.41E-02	2.10E-01	7
1.08E-01	7.81E-03	4.84E-01	5
2.44E-01	4.21E-02	6.03E-01	2
2.56E-04	1.44E-02	4.84E-02	1
1.06E-01		1.62E-01	9
4.80E-02	1.42E-01	6.44E-01	14
7.07E-02	0.00E+00	1.05E+00	1
6.36E-01	0.00E+00	9.79E-01	4
1.17E-01	5.12E-03	3.36E-01	5
4.21E-01	9.12E-02	1.05E+00	45
2.55E-01	9.89E-03	3.19E-01	18
1.32E-01		1.32E-01	
9.86E-03	1.17E-04	4.93E-02	18
3.14E-01	2.36E-06	9.12E-01	6
1.75E-01	5.73E-01	8.87E-01	6
2.04E-02	1.12E-02	5.67E-02	1
4.39E-01	5.20E-04	7.87E-01	3
5.02E-02	2.53E-02	4.78E-01	7
2.29E-01	2.75E-04	5.46E-01	2
9.64E-02	0.00E+00	9.64E-02	
6.21E-03		6.21E-03	
1.06E-04	6.53E-04	1.32E-03	53
		5.70E-01	76
4.97E-02	0.00E+00	8.34E-01	7
1.18E-01	1.61E-01	4.40E-01	137
1.35E+00		2.58E+00	11
1.15E-01	0.00E+00	1.54E-01	21
4.98E-03	0.00E+00	2.11E-02	4
1.94E-02	3.78E-04	2.68E-02	15
0.00E+00	2.02E-01	1.00E+00	26
5.95E-03	2.21E-04	1.09E-02	16
3.55E-04	0.00E+00	3.55E-04	

2.53E-01	7.65E-02	1.28E+00	32
0.00E+00	0.00E+00	1.19E-02	15
3.74E-01	1.40E-01	8.54E-01	31
1.71E-02	2.01E-02	8.59E-02	25
1.79E-04	1.30E-01	2.02E-01	89
1.14E+00	2.48E-02	2.65E+00	21
2.94E-03	1.46E-03	5.56E-02	20
3.88E-01	0.00E+00	1.17E+00	17
2.95E-01	9.50E-03	8.93E-01	20
1.84E-02	6.19E-03	3.71E-02	30
5.82E-01	0.00E+00	1.51E+00	18
3.40E-05	0.00E+00	4.81E-03	18
1.00E-04		1.00E-04	
2.18E-01	4.53E-02	2.71E-01	16
0.00E+00	0.00E+00	1.39E-01	37
4.26E-01	4.84E-02	9.95E-01	30
0.00E+00	7.70E-04	4.36E-01	12
3.10E-02	6.22E-03	1.87E-01	46
7.39E-02	8.86E-03	8.16E-01	41
0.00E+00	0.00E+00	5.45E-02	62
1.22E-02	0.00E+00	4.72E-01	145
7.75E-04	0.00E+00	7.75E-04	
0.00E+00	0.00E+00	0.00E+00	
0.00E+00	0.00E+00	0.00E+00	
0.00E+00	0.00E+00	0.00E+00	
0.00E+00	0.00E+00	0.00E+00	
0.00E+00	0.00E+00	0.00E+00	
0.00E+00	0.00E+00	0.00E+00	
0.00E+00	0.00E+00	0.00E+00	
0.00E+00	0.00E+00	0.00E+00	
0.00E+00	0.00E+00	0.00E+00	
2.11E-01	5.71E-02	5.79E-01	15
9.95E-02	7.83E-03	3.31E-01	4

Approach 1 (CE=20) Rank (by Mass)	Approach 1 (CE=40) Rank (by Mass)	Approach 1 (CE=10) Rank (by Formula)
20	1	1
1		1
1	1	1
1	4	2
1	1	1
1	1	1
1		1
1	1	1
		2
		1
1	2	1
2		1
1	1	5
		1
1	3	2
15	1	1
1	5	1
2	1	1
2	5	1
13	1	1
1	10	2
1	2	1
3	1	1
1	1	
1	1	1
2	1	1
1	3	1
1	1	1
2		74
1	1	1
2	20	2
1	1	1
1	1	1
1	1	1
1	4	1
1	2	1
1	1	1
1	12	1
2	1	1
1		1
6	1	1
5	3	1
1	2	1
1	5	1
1	12	1
8		1

1	4	1
1	198	1
11	1	1
57	8	2
1	1	1
1	1	1
43	109	2
1	1	1
1	1	1
1	1	1
1	9	1
1	1	1
1	1	1
1		1
1	5	1
2	1	2
2	10	2
1	2	1
1	1	1
1	1	1
3	1	1
	1	
1	1	1
5	1	2
1	6	3
4	2	2
1	3	1
1	5	1
1	19	1
1	2	1
1	2	1
2	15	1
3	49	
1	1	1
1	1	1
1	2	1
1	1	1
		1
1	1	1
1	1	1
2	14	1
1	9	1
1		1
1	1	1
3		2
1	2	1
3	1	1
1	4	1
1	2	1
		1

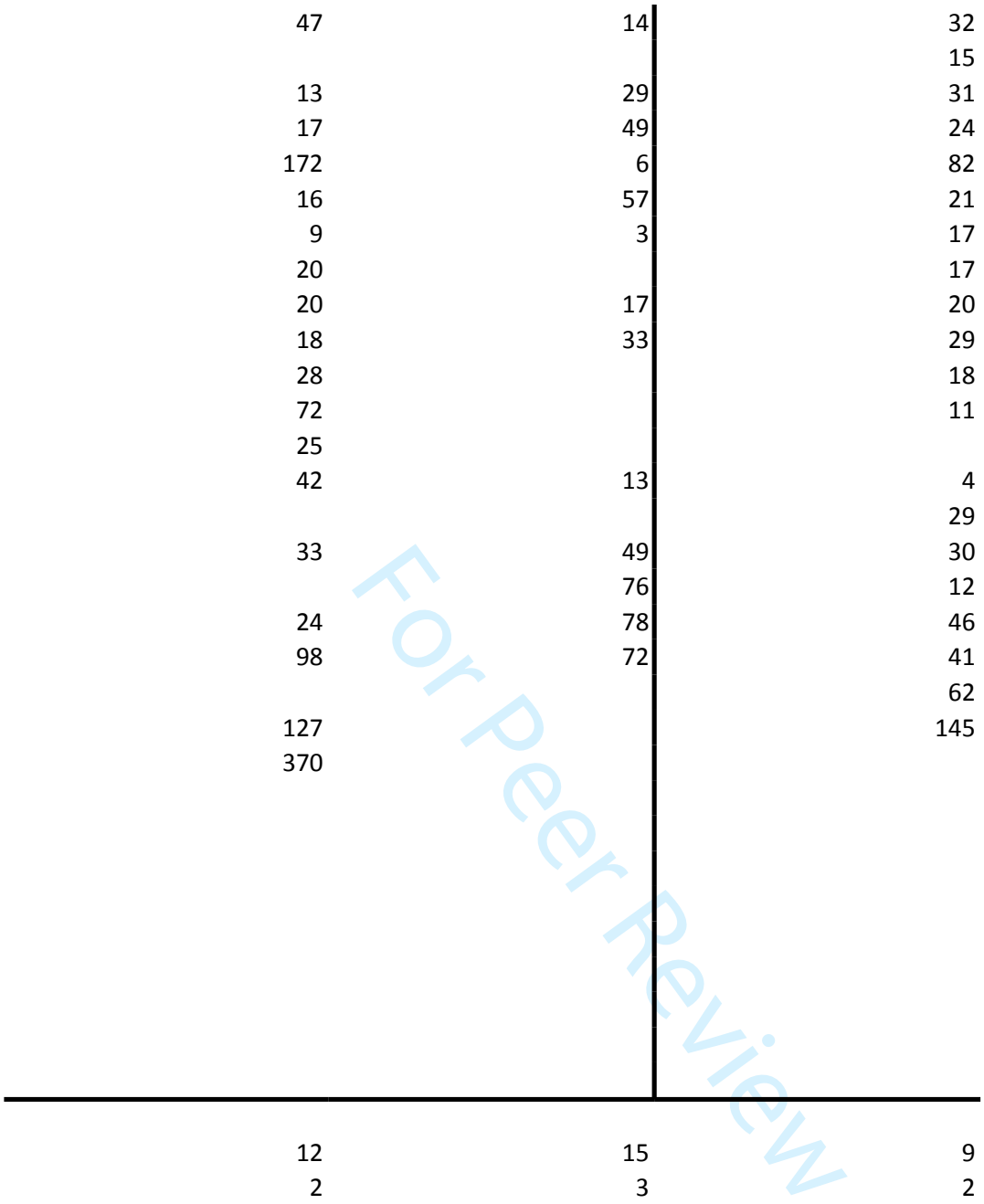
1	16	1
4	1	1
1	1	1
1	1	2
1	1	1
1	6	1
3	36	1
4	1	1
1	8	2
1		8
1	8	
4	2	
2	1	
2		
	39	
1	2	6
5	10	2
1	39	2
	4	2
1	1	3
9	28	1
4		2
2	66	2
1		3
1	58	4
2	4	8
4	4	2
2	5	2
18	50	1
2		2
2	59	2
2	7	2
2	11	2
		2
1	13	5
3	2	3
9	3	6
2		3
5	1	4
3	45	7
1	1	2
2	3	4
4		4
6	13	3
2	1	3
4	2	4
	5	2
3	5	4
3		9
12	25	2

37	26	2
4	17	4
2		5
11	35	2
4	2	5
7	13	2
5	6	4
4	4	6
10		5
	10	
11		6
1	7	23
4	9	8
8	3	5
7	3	6
10	32	24
3		4
13	67	8
20	40	5
6		9
9	10	10
1		17
9	8	9
19		11
10		12
16		15
2	1	27
11	32	51
35	25	21
43	39	8
	17	
24	18	24
3	10	34
14	36	28
61		32
		29
		16
58	39	20
40		52
55	89	69
		64
70	79	61
51	245	64
8		128
		302

41		1
1	1	1
2	1	1
33		1
17		1
1	1	1
1	2	3
1	3	1
1	3	1
2	3	1
	4	1
4	20	4
1	1	1
1	1	1
2		1
3		4
38	41	3
	2	
	15	1
	6	
55		3
9	13	3
5		
11	14	1
3		
	21	
4		
		6
		9
		3
13	10	12
1	3	29
5		2
		15
32	4	26
22	33	19
		45
47	4	33
77	60	
84	1	
129	12	46
112	64	63
	59	
		35
310		

1	1	1
		1
4	1	1
1	1	1
1	1	1
1	1	2
1	5	1
		1
9		1
1	1	1
2	1	3
1	1	1
1	1	1
5	4	1
8	1	1
2	1	1
1	1	1
1	12	1
1		1
1		11
1	41	1
1	31	1
2	10	1
1	3	1
1	1	1
1	1	1
3	2	2
1	27	1
2	1	3
5	1	1
1	1	1
4	31	3
1	1	1
3	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
2	16	1
8		1
22		1
2	10	1
1	1	1
1	2	1
1	1	1
4	4	1
1	1	1

	2	1
3	17	1
4		1
5	1	1
1	1	
1	1	1
2	1	1
1	6	1
12		1
1	1	1
1	1	1
2	3	
1		
	2	
3		
6	10	5
1	4	2
8		1
1	2	2
2	2	3
2	2	2
1	2	1
2		9
1	1	4
24		1
2		4
7	3	2
7	7	45
2	113	18
3		
6	5	4
1	124	6
8	4	6
2	6	1
18	22	1
9	53	7
5	172	2
9		
14		
97	59	7
		12
7		7
83	20	135
10		10
12		21
19		4
4	123	9
	5	26
15	36	9
9		



Approach 1 (CE=20) Rank (by Formula)	Approach 1 (CE=40) Rank (by Formula)	Approach 2 (CE=10) Rank (by Mass)
5	1	1
1		1
1	1	1
1	3	2
1	1	1
1	1	1
1		1
1	1	1
		1
		1
1	1	1
1		13
1	1	8
		9
1	3	1
1	1	135
1	3	1
2	1	1
1	5	1
1	1	2
1	9	1
1	2	5
1	1	9
1	1	
1	1	1
1	1	2
1	1	10
1	1	6
2		86
1	1	1
2	20	1
1	1	1
1	1	1
1	1	1
1	1	6
1	2	1
1	1	1
1	1	1
1	1	4
1		1
1	1	10
1	1	7
1	1	4
1	2	1
1	1	2
2		3

1	4	1
1	17	1
1	1	7
1	1	61
1	1	1
1	1	1
5	15	4
1	1	3
1	1	1
1	1	21
1	5	1
1	1	7
1	1	1
1		1
1	1	5
1	1	4
1	3	1
1	2	7
1	1	1
1	1	4
1	1	3
	1	
1	1	1
2	1	2
1	2	1
1	2	6
1	3	1
1	3	1
1	6	13
1	2	1
1	1	3
1	3	1
1	35	
1	1	1
1	1	1
1	2	2
1	1	3
		3
1	1	1
1	1	1
1	7	1
1	5	1
1		1
1	1	1
2		1
1	1	1
1	1	1
1	4	3
1	1	1
		4

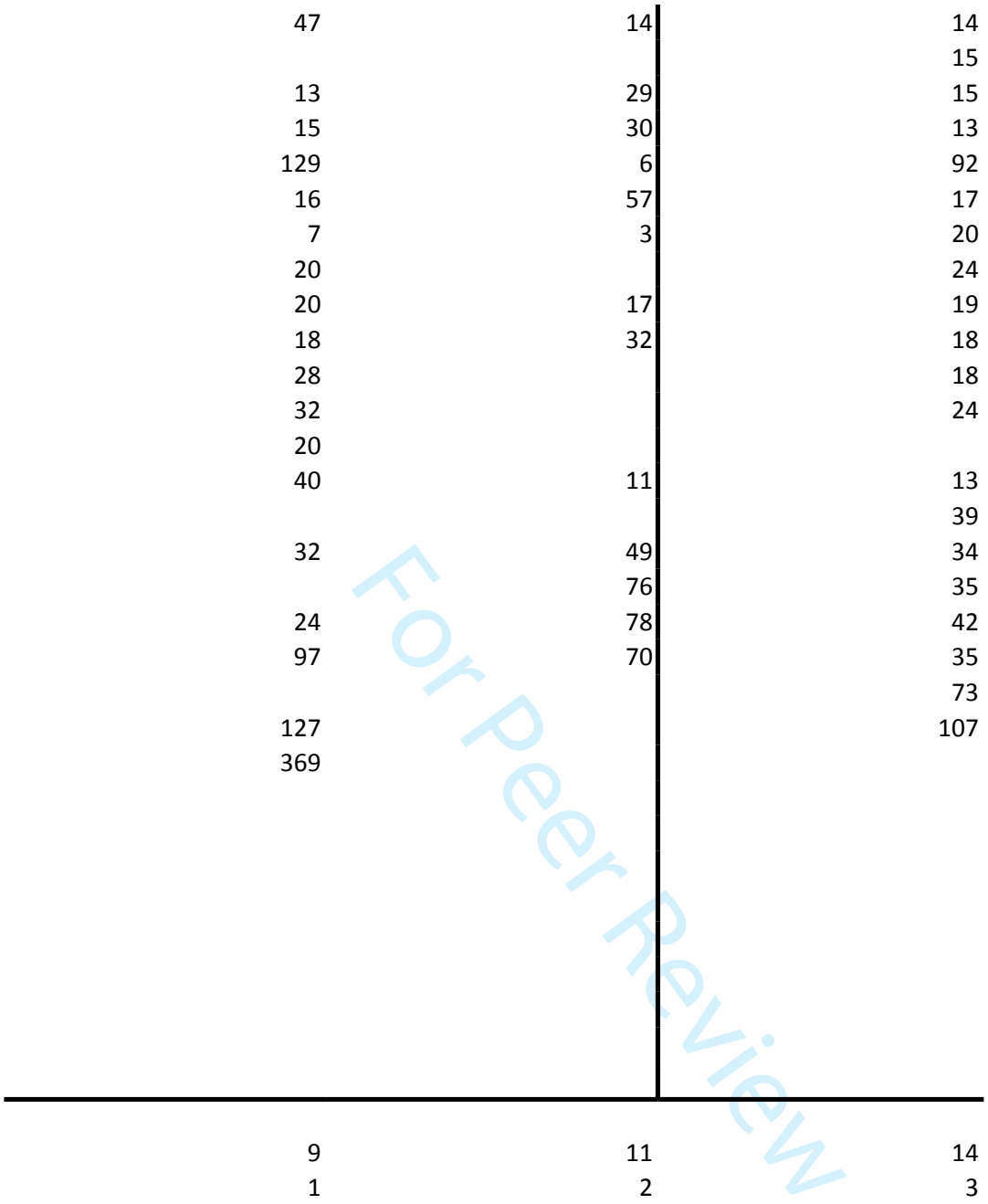
1	10	1
1	1	1
1	1	1
1	1	1
1	1	2
1	2	1
3	28	1
1	1	40
1	1	2
1		290
1	8	
4	2	
1	1	
1		
	1	
1	2	8
2	4	4
1	39	2
	1	2
1	1	10
2	2	6
4		2
2	66	2
1		2
1	54	3
2	4	11
4	4	2
2	2	2
1	2	15
2		10
2	54	1
2	5	3
2	6	2
		3
1	4	9
3	2	3
9	3	2
1		23
3	1	11
3	44	3
1	1	15
2	3	3
4		3
6	13	2
2	1	3
4	2	3
	3	7
2	5	1
3		12
12	21	3

5	5	24
4	17	4
2		2
10	31	2
4	1	23
7	13	16
5	6	6
3	3	7
5		2
	5	
9		6
1	7	24
4	9	6
4	3	8
6	3	10
8	15	8
2		17
4	5	111
20	40	3
2		32
9	10	9
1		14
7	2	19
19		10
9		19
16		13
2	1	38
11	32	21
11	7	75
38	36	10
	11	
23	17	18
3	8	61
14	30	20
61		6
		36
		17
56	39	15
40		48
35	79	85
		53
58	65	80
51	244	32
4		146
		225

1		56
1	1	1
1	1	1
2		37
1		16
1	1	3
1	2	4
1	3	1
1	2	1
2	2	1
	1	4
3	8	5
1	1	1
1	1	5
1		2
3		4
4	38	2
	2	
	3	17
	2	
4		47
5	11	3
3		
3	2	2
3		
	4	
4		
		5
		13
		10
13	10	11
1	3	71
3		24
		15
30	4	29
22	28	19
		38
47	4	51
76	60	
44	1	
74	11	44
59	32	62
	53	
		74
294		

1	1	1
		1
1	1	2
1	1	1
1	1	15
1	1	3
1	3	1
		1
1		8
1	1	1
1	1	1
1	1	2
1	1	2
1	2	3
1	1	14
1	1	2
1	1	1
1	7	1
1		1
1		1
1	1	14
1	12	1
1	6	1
1	1	2
1	1	1
1	1	1
1	2	6
1	9	1
2	1	6
5	1	2
1	1	1
4	5	1
1	1	6
1	1	19
1	1	1
1	1	1
1	1	2
1	1	1
1	1	1
1	1	3
1		7
1		38
1	1	1
1	1	1
1	1	1
1	1	5
3	3	1
1	1	1

	1	1
1	1	2
1		1
1	1	7
1	1	
1	1	1
1	1	2
1	1	2
1		12
1	1	1
1	1	2
1	1	
1		
	1	
2		
5	10	2
1	4	2
7		2
1	2	5
1	1	6
2	2	2
1	2	1
2		8
1	1	12
6		2
2		3
5	2	8
7	7	7
2	111	7
3		
1	1	18
1	104	5
8	4	5
2	5	2
10	12	13
9	53	5
5	165	2
8		
6		
7	4	86
		70
7		8
81	20	3
10		10
12		20
19		9
4	68	11
	5	43
11	13	25
9		



Approach 2 (CE=20) Rank (by Mass)	Approach 2 (CE=40) Rank (by Mass)	Approach 2 (CE=10) Rank (by Formula)
8	1	1
1		1
1	1	1
1	3	1
1	1	1
1	1	1
1		1
1	1	1
		1
		1
1	2	1
1		1
1	1	6
		1
1	4	1
12	1	1
1	3	1
1	1	1
1	1	1
1	1	1
1	9	1
1	1	1
1	1	1
1	1	
1	1	1
1	1	1
1	3	1
1	1	1
1		84
1	1	1
2	16	1
1	1	1
1	1	1
1	1	1
1	4	1
1	1	1
1	1	1
1	10	1
1	1	1
1		1
8	1	1
5	3	1
1	2	3
1	5	1
2	9	1
16		1

1	5	1
1	202	1
6	2	1
30	1	2
1	1	1
1	1	1
12	110	1
1	1	1
1	1	1
1	1	1
2	8	1
1	1	1
1	1	1
1		1
1	2	1
2	1	1
1	7	1
1	2	1
2	1	1
1	1	1
4	1	1
	1	
1	1	1
1	4	2
1	8	1
4	5	1
1	1	1
1	7	1
1	17	2
1	1	1
1	2	1
1	3	1
4	54	
1	1	1
1	1	1
1	2	2
1	1	3
		1
1	1	1
1	1	1
4	14	1
1	11	1
1		1
1	1	1
4		1
1	1	1
6	1	1
1	6	3
1	2	1
		1

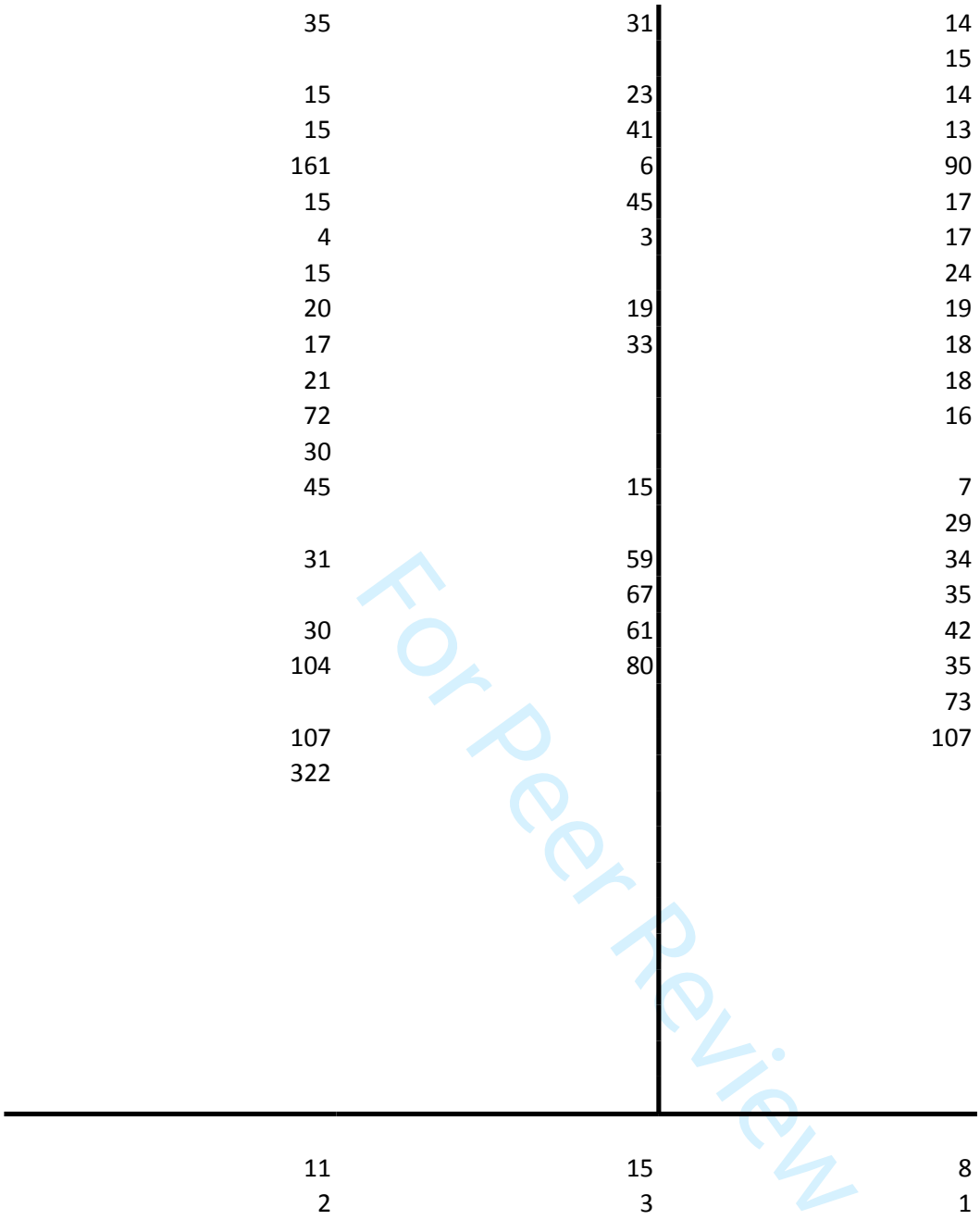
1	17	1
1	4	1
1	1	1
1	1	1
1	1	1
1	5	1
2	22	1
2	1	1
1	8	1
1		7
1	7	
4	2	
1	1	
1		
	49	
1	2	1
5	11	1
2	10	2
	5	2
1	1	3
9	33	1
4		2
4	15	2
1		2
1	62	3
2	3	7
2	2	2
2	21	2
3	28	2
2		1
6	14	1
1	3	3
2	12	2
		2
3	9	4
3	2	3
6	2	2
3		3
4	4	4
3	45	3
2	6	5
3	3	3
6		3
5	13	2
2	1	3
3	1	3
	6	2
6	5	1
4		7
7	33	3

25	26	4
4	23	4
3		2
13	35	2
6	1	5
7	15	4
5	5	4
3	6	7
10		1
	9	
8		6
4	12	23
3	7	6
11	4	5
5	3	8
4	35	5
6		7
8	61	8
16	35	3
3		9
9	10	9
1		12
10	11	9
18		10
7		16
13		13
3	2	27
13	46	21
42	30	18
24	43	9
	20	
9	22	16
3	3	30
19	30	20
65		5
		26
		17
79	18	15
38		48
47	82	69
		52
75	80	59
96	266	28
23		128
		225

54		1
1	1	1
3	1	1
36		1
17		1
1	1	1
1	2	3
1	2	1
1	3	1
1	3	1
	4	1
2	14	4
1	1	1
1	2	1
3		1
2		4
6	41	2
	2	
	11	1
	7	
38		3
13	9	3
4		
7	11	2
2		
	21	
5		
		5
		9
		10
12	12	11
1	3	24
8		18
		15
29	5	28
22	34	19
		26
54	7	51
20	82	
99	1	
118	31	34
121	74	61
	75	
		74
270		

1	1	1
		1
4	1	1
1	1	1
1	1	1
1	1	2
1	8	1
		1
9		1
1	1	1
1	1	1
1	1	1
1	1	1
9	4	1
11	1	1
2	1	1
1	1	1
1	13	1
1		1
1		1
4	50	1
1	28	1
1	9	1
1	1	1
1	1	1
1	1	1
2	2	2
1	3	1
1	1	2
1	3	2
1	1	1
4	48	1
1	1	1
3	1	1
1	1	1
1	1	1
1	1	1
1	1	1
1	1	1
2	14	1
7		1
38		1
1	1	1
1	1	1
1	1	1
1	1	1
2	4	1
1	1	1

	1	1
2	22	1
5		1
5	1	1
1	1	
1	1	1
2	1	1
1	2	1
12		1
1	1	1
1	1	1
2	2	
1		
	1	
8		
2	5	2
2	4	2
12		2
2	2	2
1	2	3
2	2	2
1	2	1
2		8
1	1	3
10		1
3		3
3	2	5
5	13	7
2	114	7
3		
3	6	4
4	124	5
5	4	5
5	6	1
8	22	6
10	68	5
5	184	2
7		
14		
105	58	9
		8
7		8
17	37	3
9		10
9		20
25		9
6	123	9
	9	43
16	36	12
11		



Approach 2 (CE=20) Rank (by Formula)	Approach 2 (CE=40) Rank (by Formula)	Approach 3 Rank (by Mass)	Approach 3 Rank (by Formula)
1	1	4	1
1		1	1
1	1	1	1
1	3	1	1
1	1	1	1
1	1	1	1
1		7	1
1	1	1	1
		1	1
		1	1
1	1	1	1
1		13	1
1	1	1	1
		10	1
1	4	1	1
1	1	53	1
1	3	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	8	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	5	1
1	1	1	1
1		1	1
1	1	1	1
1	15	2	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	5	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1		1	1
1	1	1	1
1	1	4	1
1	1	2	1
1	2	1	1
1	1	3	1
2		4	1

1	4	1	1
1	17	1	1
1	1	6	1
1	1	58	1
1	1	1	1
1	1	1	1
2	16	12	1
1	1	1	1
1	1	1	1
1	1	10	1
2	4	1	1
1	1	3	1
1	1	1	1
1		1	1
1	1	2	1
1	1	3	1
1	2	1	1
1	2	1	1
1	1	1	1
1	1	1	1
1	1	2	1
	1	4	1
1	1	1	1
1	2	1	1
1	2	1	1
1	2	3	1
1	1	1	1
1	5	1	1
1	6	11	1
1	1	1	1
1	1	3	1
1	2	1	1
1	39	4	1
1	1	1	1
1	1	1	1
1	2	1	1
1	1	1	1
		3	1
1	1	1	1
1	1	1	1
2	7	1	1
1	5	1	1
1		1	1
1	1	1	1
2		1	1
1	1	1	1
1	1	1	1
1	6	1	1
1	1	1	1
		4	1

1	10	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
2	18	1	1
1	1	30	1
1	1	2	1
1		3	1
1	7	1	1
4	2	1	1
1	1	1	1
1		1	1
	1	151	1
1	2	2	2
2	3	5	2
2	10	2	2
	1	2	2
1	1	5	2
2	2	7	2
4		2	2
4	15	2	2
1		2	2
1	58	2	2
2	3	6	2
2	2	2	2
2	2	2	2
1	1	12	2
2		6	2
6	12	2	2
1	2	2	2
2	7	2	2
		2	2
2	3	7	3
3	2	3	3
6	2	3	3
1		18	3
3	2	5	3
3	44	3	3
2	6	3	3
3	3	3	3
5		3	3
5	13	3	3
2	1	3	3
3	1	3	3
	3	11	3
5	4	4	4
4		7	4
7	29	4	4

4	5	26	4
4	22	4	4
3		4	4
12	35	4	4
4	1	23	5
7	13	5	5
5	5	5	5
3	5	6	5
5		7	5
	5	37	6
7		6	6
4	12	6	6
3	7	6	6
6	3	8	6
4	3	9	7
4	17	10	7
4		25	8
4	5	108	8
16	35	9	9
2		32	9
9	10	9	9
1		11	10
8	3	19	10
18		10	10
7		11	11
13		13	13
3	1	15	14
13	46	14	14
12	8	72	15
23	41	17	17
	11	47	18
8	21	19	18
3	3	40	19
19	22	19	19
65		26	26
		30	26
		29	29
77	18	43	41
38		47	47
36	73	61	51
		54	52
57	65	89	65
96	256	70	68
15		141	128
		225	225

1			
1		54	1
1	1	1	1
1	1	1	1
1		37	1
1		17	1
1	1	3	1
1	2	1	1
1	2	1	1
1	2	1	1
1	2	1	1
	1	4	1
1	5	3	1
1	1	1	1
1	1	5	1
1		3	1
2		2	2
6	38	2	2
	2	2	2
	3	16	2
	2	14	2
3		44	3
8	9	5	3
3		4	3
3	2	3	3
2		12	4
	4	21	4
5		5	5
		5	5
		10	9
		10	10
12	12	11	11
1	3	42	13
5		22	16
		17	17
29	5	21	21
22	29	21	21
		42	26
54	7	29	29
20	82	34	34
53	1	73	42
68	31	59	52
65	32	64	63
	63	75	63
		142	142
256		278	264

1	1	1	1
		1	1
1	1	3	1
1	1	1	1
1	1	7	1
1	1	1	1
1	6	1	1
		1	1
1		9	1
1	1	1	1
1	1	1	1
1	1	2	1
1	1	1	1
2	2	3	1
1	1	13	1
1	1	2	1
1	1	1	1
1	7	1	1
1		1	1
1		1	1
1	1	15	1
1	12	1	1
1	5	1	1
1	1	2	1
1	1	1	1
1	1	1	1
1	2	3	1
1	2	1	1
1	1	3	1
1	3	1	1
1	1	1	1
4	10	1	1
1	1	1	1
1	1	9	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
1		7	1
1		37	1
1	1	1	1
1	1	1	1
1	1	1	1
1	1	1	1
2	3	1	1
1	1	1	1

	1	2	1
1	1	3	1
1		1	1
1	1	7	1
1	1	1	1
1	1	1	1
1	1	2	1
1	1	1	1
1		12	1
1	1	1	1
1	1	1	1
1	1	2	1
1		1	1
	1	1	1
2		17	2
2	5	2	2
2	4	2	2
9		2	2
1	2	2	2
1	1	4	2
2	2	2	2
1	2	2	2
2		2	2
1	1	10	2
3		3	2
3		3	3
2	1	5	3
5	13	3	3
2	111	3	3
3		3	3
1	1	16	4
4	104	4	4
5	4	4	4
5	5	4	4
5	12	10	5
10	68	5	5
5	178	5	5
6		7	6
7		14	7
8	4	95	8
		70	8
7		8	8
16	37	9	9
9		9	9
9		10	10
25		11	11
6	70	12	11
	9	12	12
10	14	26	14
11		14	14

35	31	15	15
		15	15
14	23	17	16
14	39	17	16
120	6	16	16
15	45	17	17
3	3	19	17
15		18	18
20	19	20	20
17	32	21	21
21		21	21
32		38	22
22		30	22
43	10	25	23
		33	33
31	59	33	33
	67	35	35
30	61	39	39
103	78	45	45
		73	73
107		107	107
321		322	321
8	11	14	9
1	2	3	1

Approach 3 Quotient (by Mass)	Approach 3 Quotient (by Formula)	Approach 3 Percentile (by Mass)	Approach 3 Percentile (by Formula)
0.89	1.00	94	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
0.53	1.00	81	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
0.72	1.00	74	100
1.00	1.00	100	100
0.65	1.00	59	100
1.00	1.00	100	100
0.52	1.00	80	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
0.92	1.00	92	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
0.98	1.00	99	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
0.50	1.00	60	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
0.70	1.00	82	100
0.95	1.00	97	100
1.00	1.00	100	100
0.50	1.00	93	100
0.74	1.00	92	100

1.00	1.00	100	100
1.00	1.00	100	100
0.77	1.00	95	100
0.27	1.00	11	100
1.00	1.00	100	100
1.00	1.00	100	100
0.53	1.00	93	100
1.00	1.00	100	100
1.00	1.00	100	100
0.90	1.00	82	100
1.00	1.00	100	100
0.98	1.00	93	100
1.00	1.00	100	100
1.00	1.00	100	100
0.78	1.00	97	100
0.66	1.00	88	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
0.83	1.00	86	100
0.35	1.00	50	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
0.82	1.00	93	100
1.00	1.00	100	100
1.00	1.00	100	100
0.68	1.00	76	100
1.00	1.00	100	100
0.56	1.00	50	100
1.00	1.00	100	100
0.81	1.00	97	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
0.83	1.00	96	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
0.85	1.00	57	100

1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
0.53	1.00	64	100
0.36	1.00	96	100
0.27	1.00	99	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
0.03	1.00	6	100
0.97	0.97	98	95
0.90	0.99	85	90
1.00	1.00	100	100
0.48	0.48	92	91
0.85	0.93	81	75
0.62	0.98	89	50
0.80	0.80	99	99
0.89	0.89	100	100
0.99	0.99	99	99
0.92	0.92	99	99
0.80	0.95	95	99
0.99	0.99	99	99
0.88	0.88	100	80
0.60	0.99	86	50
0.64	0.87	88	93
1.00	1.00	99	99
0.93	0.93	96	80
0.75	0.75	98	96
0.78	0.78	98	97
0.46	0.54	71	75
0.83	0.83	88	88
0.69	0.69	99	99
0.34	0.80	48	75
0.50	0.69	92	50
0.99	0.99	98	98
0.80	0.80	98	96
0.97	0.97	97	94
0.86	0.86	88	82
0.96	0.96	98	97
0.88	0.88	71	71
0.61	0.61	71	71
0.02	0.31	41	50
0.98	0.98	93	90
0.65	0.68	97	93
0.89	0.89	96	95

0.40	0.84	34	40
0.36	0.36	99	98
0.22	0.22	95	95
0.87	0.87	99	99
0.28	0.52	31	20
0.88	0.88	96	91
0.36	0.36	94	84
0.77	0.77	93	86
0.32	0.40	82	20
0.00	0.10	21	68
0.59	0.59	93	91
0.83	0.83	93	84
0.91	0.91	92	91
0.16	0.16	91	84
0.52	0.52	73	71
0.28	0.28	91	82
0.03	0.06	43	46
0.30	0.37	26	13
0.93	0.93	87	85
0.47	0.47	47	11
0.23	0.23	90	82
0.69	0.69	80	79
0.42	0.55	44	47
0.73	0.73	87	87
0.22	0.22	87	86
0.45	0.45	95	95
0.71	0.71	88	68
0.48	0.48	93	92
0.46	0.52	28	55
0.20	0.20	95	94
0.01	0.33	71	81
0.26	0.26	53	51
0.53	0.68	57	49
0.71	0.71	63	49
0.88	0.88	87	74
0.51	0.67	61	62
0.29	0.29	72	71
0.69	0.69	70	71
0.59	0.59	54	54
0.33	0.33	71	68
0.54	0.54	45	39
0.35	0.35	21	17
0.80	0.80	80	78
0.01	0.01	55	53
0.43	0.43	57	57
0.00	0.00	32	27
0.00		21	51
0.00		47	75
0.00	0.00	34	33
0.00	0.00	8	5

0.00		50	67
0.00		40	100
0.39	1.00	22	100
1.00	1.00	100	100
1.00	1.00	100	100
0.36	1.00	35	100
0.15	1.00	6	100
0.47	1.00	33	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
0.14	1.00	77	100
0.69	1.00	95	100
1.00	1.00	100	100
0.45	1.00	60	100
0.28	1.00	94	100
0.90	0.90	96	89
0.92	0.92	99	99
0.43	0.43	75	67
0.03	0.18	79	80
0.01	0.93	86	91
0.27	0.52	38	60
0.71	0.92	85	86
0.60	0.98	86	60
0.82	0.82	89	33
0.48	0.48	91	90
0.00	0.01	63	57
0.94	0.94	81	56
0.91	0.91	43	43
0.42	0.42	25	11
0.86	0.86	98	98
0.20	0.20	94	91
0.61	0.79	59	74
0.68	0.68	89	88
0.13	0.13	93	85
0.68	0.68	69	61
0.54	0.54	56	43
0.02	0.03	56	55
0.14	0.14	77	76
0.76	0.76	68	68
0.54	0.54	23	25
0.27	0.34	66	42
0.49	0.49	20	13
0.42	0.42	62	40
0.10	0.10	35	34
0.47	0.47	30	28
0.00	0.00	44	45
0.00	0.00	34	45
0.00		48	50

0.00		30	100
0.00	0.00	21	21
1.00	1.00	100	100
1.00	1.00	100	100
0.88	1.00	97	100
1.00	1.00	100	100
0.72	1.00	81	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
0.31	1.00	50	100
1.00	1.00	100	100
1.00	1.00	100	100
0.91	1.00	94	100
1.00	1.00	100	100
0.97	1.00	92	100
0.54	1.00	52	100
0.88	1.00	75	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
0.62	1.00	95	100
1.00	1.00	100	100
1.00	1.00	100	100
0.85	1.00	96	100
1.00	1.00	100	100
1.00	1.00	100	100
0.65	1.00	88	100
1.00	1.00	100	100
0.89	1.00	92	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
0.39	1.00	77	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
0.52	1.00	79	100
0.37	1.00	32	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100
1.00	1.00	100	100

0.76	1.00	67	100
0.96	1.00	97	100
1.00	1.00	100	100
0.25	1.00	45	100
1.00	1.00	100	100
1.00	1.00	100	100
0.93	1.00	83	100
1.00	1.00	100	100
0.45	1.00	48	100
1.00	1.00	100	100
1.00	1.00	100	100
0.41	1.00	75	100
1.00	1.00	100	100
1.00	1.00	100	100
0.03	0.05	54	67
0.99	0.99	99	99
0.98	0.98	100	100
0.95	0.95	99	99
0.94	0.94	95	67
0.69	0.93	91	86
0.99	0.99	50	50
0.81	0.81	98	93
0.75	0.75	99	99
0.44	0.88	69	75
0.85	0.98	98	96
0.95	0.95	95	90
0.89	0.89	89	71
0.92	0.92	99	99
0.94	0.94	99	99
0.85	0.85	60	60
0.33	0.90	38	25
0.78	0.78	99	98
0.50	0.50	92	88
0.52	0.52	96	85
0.63	0.87	79	69
0.70	0.70	99	99
0.87	0.87	99	99
0.97	1.00	71	50
0.26	0.27	68	54
0.00	0.05	20	42
0.41	0.51	24	63
0.65	0.65	89	85
0.61	0.61	98	97
0.78	0.78	91	62
0.72	0.72	91	90
0.39	0.39	79	79
0.25	0.25	93	93
0.67	0.67	89	88
0.09	0.12	59	28
0.04	0.04	41	41

0.76	0.76	94	94
0.14	0.14	44	44
0.68	0.68	92	91
0.08	0.08	94	92
0.48	0.48	93	91
0.60	0.60	93	93
0.30	0.30	10	6
0.70	0.70	73	70
0.39	0.39	10	10
0.06	0.06	88	82
0.48	0.48	74	74
0.04	0.04	59	50
0.02	0.03	15	9
0.50	0.50	87	65
0.51	0.51	38	38
0.77	0.77	85	84
0.58	0.58	73	73
0.26	0.26	81	81
0.57	0.57	78	78
0.30	0.30	47	45
0.60	0.60	70	69
0.01	0.01	36	35
		51	51
0.00	0.00	28	50
0.00	0.00	10	11
0.00	0.00	46	45
0.00	0.00	45	45
0.00	0.00	17	17
0.00	0.00	28	33
		63	100
0.00	0.00	6	4
0.69	0.79	81	84
0.83	1.00	94	100