



Published in final edited form as:

*Anal Chem.* 2021 April 13; 93(14): 5754–5762. doi:10.1021/acs.analchem.0c04895.

## Extension of Diagnostic Fragmentation Filtering for Automated Discovery in DNA Adductomics

Kevin J. Murray<sup>†</sup>, Erik S. Carlson<sup>‡</sup>, Alessia Stornetta<sup>†</sup>, Emily P Balskus<sup>‡</sup>, Peter W. Villalta<sup>†,‡</sup>, Silvia Balbo<sup>†,λ,\*</sup>

<sup>†</sup>Masonic Cancer Center, University of Minnesota, 2231 6th Street SE, Minneapolis, MN 55455, USA.

<sup>‡</sup>Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA.

<sup>‡</sup>Department of Medicinal Chemistry, University of Minnesota, Minneapolis, MN 55455, USA.

<sup>λ</sup>Division of Environmental Health Sciences, School of Public Health, University of Minnesota.

### Abstract

Development of high resolution/accurate mass liquid chromatography-coupled tandem mass spectrometry (LC-MS/MS) methodology enables the characterization of covalently modified DNA induced by interaction with genotoxic agents in complex biological samples. Constant neutral loss monitoring of 2'-deoxyribose or the nucleobases using data-dependent acquisition represents a powerful approach for the unbiased detection of DNA modifications (adducts). The lack of available bioinformatics tools necessitates manual processing of acquired spectral data and hampers high throughput application of these techniques. To address this limitation, we present an automated workflow for the detection and curation of putative DNA adducts by using diagnostic fragmentation filtering of LC-MS/MS experiments within the open-source software MZmine. The workflow utilizes a new feature detection algorithm, DFBuilder, which employs diagnostic fragmentation filtering using a user-defined list of fragmentation patterns to reproducibly generate feature lists for precursor ions of interest. The DFBuilder feature detection approach readily fits into a complete small molecule discovery workflow and drastically reduces the processing time associated with analyzing DNA adductomics results. We validate our workflow using a mixture of authentic DNA adduct standards and demonstrate the effectiveness of our approach by reproducing and expanding the results of a previously published study of colibactin-induced DNA adducts. The reported workflow serves as a technique to assess the diagnostic potential of novel fragmentation pattern combinations for the unbiased detection of chemical classes of interest.

### Graphical Abstract

\*Corresponding Author: balbo006@umn.edu, Fax: (612) 624-3869. Tel.: (612) 624-4240.

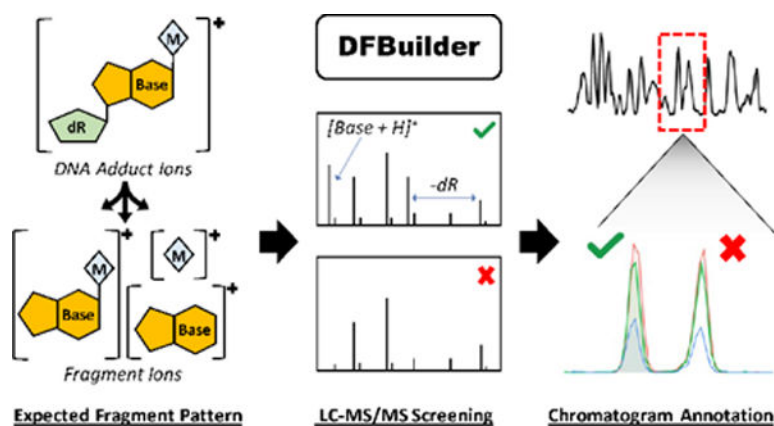
ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.0c04895>.

Figures S-1 – S-3 and Tables S1 – S-3 as described in text (PDF)

The authors declare no competing financial interest.



Exposure to most genotoxic chemicals induces covalent modifications of DNA that, when not repaired, represent a major risk factor for disease pathogenesis and the development of cancer.<sup>1-4</sup> A wide array of distinct adduction products result from alkylation, oxidation, and deamination reactions involving modifying agents.<sup>5</sup> The simultaneous identification of multiple DNA adducts in complex biological matrices represents a major bottleneck for the comprehensive investigation of the genotoxic effects of exposures. The high sensitivity and structural elucidation capability of liquid chromatography-coupled tandem mass spectrometry (LC-MS/MS) enable the rapid characterization of known and unknown modifications.<sup>4,6,7</sup> Application of this technology for the discovery of DNA modifications led to the development of DNA adductomics, a systems biology approach designed to identify and annotate the multitude of known and novel adduction products induced by the effects of the combination of harmful exposures on DNA.<sup>8</sup>

A variety of LC-MS/MS techniques may be applied to characterize DNA adducts present in common biological samples, including tissues, blood, or cell cultures.<sup>8</sup> Prior to analysis, DNA is isolated—typically hydrolyzed to nucleosides—enriched, and purified to ensure sufficient signals for MS analysis. Often, clean-up steps such as solid phase extraction are used to remove unmodified bases and other sample components that confound the results.<sup>9</sup> The most common LC-MS/MS techniques for DNA adducts analysis monitor known adduct precursor-to-product ion transitions using selected reaction monitoring or parallel reaction monitoring experiments.<sup>10</sup> Although offering high sensitivity for quantitation of low abundant ions, these techniques require an *a priori* defined precursor mass, precluding the detection of novel DNA adducts.<sup>4,6</sup> Alternatively, untargeted screening of precursor-product ion transitions is increasingly applied for the unbiased assessment of complex matrices. It has been observed that the neutral loss of the 2'-deoxyribose moiety during collision-induced dissociation (CID) is a highly selective signature of deoxynucleosides.<sup>11</sup> Monitoring this predictable fragmentation pattern via constant-neutral loss (CNL) scans serves as the foundation of many DNA adductomics strategies, including data-independent acquisition (DIA) and data-dependent acquisition (DDA) methodologies. DIA approaches provide a near full coverage of complex matrices by simultaneously assessing multiple compound fragmentation patterns using wide-window precursor isolation.<sup>12</sup> These techniques, however, offer limited structural information of low abundant ions. Additionally, DIA methods are prone to false discovery resulting from the co-fragmentation of multiple ion species in each

successive MS/MS scan.<sup>13</sup> Robust sample-specific spectrum libraries ease the deconvolution and detection of putative DNA adducts but the generation of these libraries requires considerable time and effort. DDA strategies provide more precursor structural information necessary for compound identification. However, insufficient scanning speed of complex sample matrices limits the ability of this approach to fragment low abundance ions and discovery of trace-level adducts.<sup>12</sup>

A DDA-CNL-MS<sup>3</sup> DNA adductomics approach with high-resolution/accurate mass MS<sup>n</sup> fragmentation has recently been developed.<sup>6,14</sup> The structural elucidation potential offered by MS<sup>2</sup> and MS<sup>3</sup> fragmentation provides valuable insight into unknown modifications and simultaneously controls for false-positive detection using accurate mass monitoring. However, the success of this approach has been limited by the traditional data analysis strategy, which uses the appearance of an MS<sup>3</sup> event as an indicator of a putative DNA adduct. This approach limits the flexibility of the methodology because alternative neutral losses and product ions in the MS<sup>2</sup> spectra cannot be identified and is dependent upon the fidelity of the instrument in the triggering of the MS<sup>3</sup> data acquisition upon observation of the criteria neutral loss.

Diagnostic fragmentation filtering is a computational method that closely resembles CNL-MS/MS data acquisition methods.<sup>15</sup> The algorithm searches for the presence of diagnostic pattern in all MS<sup>2</sup> spectra, exporting the corresponding precursors of interest. In contrast to instrument CNL methods, diagnostic fragmentation filtering is applied post-acquisition and facilitates repeated analysis with new fragmentation patterns without the need for sample re-injection. Application of this approach in the field of microbial natural products chemistry demonstrated the potential of fragmentation monitoring to detect known and novel products in complex biological matrices.<sup>15</sup> Similarly, diagnostic fragmentation scanning approaches are utilized in LC-MS/MS proteomics to characterize peptides and proteins with complex post-translational modifications, such as glycopeptides.<sup>16,17</sup> To the best of our knowledge, application of diagnostic fragmentation filtering for the discovery of DNA adducts is yet to be explored. Instrument-independent data analysis for DNA adducts overcomes the limitations of previous screening strategies and enhances the application of DNA adductomics to new experiments.

In this study, we adapted the diagnostic fragmentation filtering approach to enable an automatic data processing workflow for the discovery of DNA adducts using a newly developed MZmine module,<sup>18</sup> called DFBuilder. The module scans all input LC-MS/MS spectra for any number of user-specified fragmentation patterns and exports a feature list of targeted-extracted ion chromatograms (EIC) for precursors of interest. The resulting feature lists are further processed to ensure that only high-quality peaks remain and any detected duplicate, isotope, and in-source fragments are removed. The strength of our workflow is its capacity for fully automated and reproducible data analysis via batch processing relative to manual processing. All workflow components and corresponding parameterization are saved in an XML file, which can be easily reprocessed or shared between experiments.<sup>18</sup> Our approach is only dependent on computer processing resources and enables the scaling up of experimental designs that would otherwise require days of manual data processing time.

Here, we establish the effectiveness of our workflow for the automated discovery of DNA adducts and demonstrate its potential to expand screening strategies in future experiments.

## EXPERIMENTAL SECTION

### Chemical Standard Mixture.

*O*<sup>6</sup>-Methyl-2'-deoxyguanosine (*O*<sup>6</sup>-me-dG) (**1**), 8-oxo-7, 8-dihydro-2'-deoxyguanosine (8-oxo-dG) (**2**), *N*<sup>6</sup>-hydroxymethyldeoxyadenosine (*N*<sup>6</sup>-Me-dA) (**6**), and 1, *N*<sup>6</sup>-etheno-2'-deoxyadenosine (*e*-dA) (**7**) were purchased from Sigma-Aldrich (St. Louis, MO). *N*<sup>2</sup>-Ethyl-2'-deoxyguanosine (*N*<sup>2</sup>-ethyl-dG) (**3**), (6*R/S*)-3-(2'-deoxyribose-1'-yl)-5,6,7,8-tetrahydro-6-hydroxypyrimido[1,2-*a*]-purine-10(3*H*)one (OH-PdG) (**4**) and *O*<sup>2</sup>-[4-(3-pyridyl)-4-oxobut-1-yl]thymidine (*O*<sup>2</sup>-POB-dT) (**8**), D<sub>5</sub>-ethyl-2'-deoxycytidine (D<sub>5</sub>-ethyl-dC) (**9**) were prepared as described.<sup>19–22</sup> 6-(1-Hydroxyhexanyl)-8-hydroxy-1, *N*<sup>2</sup>-propano-2'-deoxyguanosine (HNE-dG) (**5**) was generously donated by Dr. Fung-Lung Chung of Georgetown University Medical Center. The nine standards were dissolved in 20% methanol and combined at a final concentration of 10 fmol/μL. The mixture of the standards was prepared in triplicate for LC-MS analysis. All solvents were LC-MS grade and were purchased from Sigma-Aldrich.

### DNA from HeLa Cells Exposed to *pks*<sup>+</sup> *E. coli*.

The previously acquired data was obtained with permission and a complete research protocol has been previously described.<sup>23</sup> The BAC*pks* island (*pks*<sup>+</sup>) and empty pBeloBAC (*pks*<sup>-</sup>) bacterial artificial chromosomes were used to generate derivative parent strains of *E. coli* harboring colibactin biosynthesis genes. HeLa cells were transiently infected with each strain, and genomic DNA was isolated for DDA-CNL-MS<sup>3</sup> analysis. Results from the DNA adductomics analysis of one replicate of *pks*<sup>+</sup> and *pks*<sup>-</sup> were used in the DFBuilder workflow.

### LC-MS Parameters.

All analyses were conducted using identical chromatographic conditions and MS instrument settings, unless otherwise described. An UltiMate 3000 RSLCnano HPLC system (Thermo Scientific, Waltham, MA) was interfaced to an Orbitrap Fusion Tribrid MS (Thermo Fisher Scientific, San Jose, CA). One microliter of the authentic DNA standard mixture and 5 μL of *E. coli* DNA extracts were injected onto the analytical platform equipped with a 5 μL injection loop. Solvent blanks were analyzed before and after acquisition to assess contamination and sample carryover between injections. Chromatographic separation was performed using a custom-packed capillary column (75 μm ID, 20 cm length, 10 μm orifice) using a commercially available fused-silica emitter (New Objective, Woburn MA) containing a Luna C18 (Phenomenex Corp. Torrance, CA) stationary phase (5 μm, 120 Å). The LC solvents were (A) 0.05% HCO<sub>2</sub>H in H<sub>2</sub>O and (B) CH<sub>3</sub>CN solutions. The flow rate was 1000 nL/min for 5.5 min at 2% B and then decreased to 300 nL/min with a 25 min linear gradient from 2 to 50% B, an increase to 98% B in 1 min, with a 4 min hold, and a 5 min equilibration at 1000 nL/min to the starting conditions. The injection valve was switched at 5.5 min to remove the sample loop from the flow path during the gradient. A

Nanospray Flex ion source (Thermo Fisher Scientific) was used with a source voltage of 2.2 kV and capillary temperature of 300 °C. The S-Lens RF level setting was 60%.

Untargeted DDA-CNL-MS<sup>3</sup> analyses were performed with full-scan detection followed by MS<sup>2</sup> acquisition and constant neutral loss triggering of MS<sup>3</sup> fragmentation. Full-scan detection was performed using the Orbitrap detection at a resolution of 60,000, automatic gain control (AGC) targeted setting of  $2 \times 10^5$ , and a maximum ion injection time setting of 118 ms. Full scan ranges of 300 – 1000 *m/z* and 150 – 1000 *m/z* were used for the *pks*<sup>+</sup> infected cells and the authentic standards, respectively. MS<sup>2</sup> spectra were acquired with quadrupole isolation of 1.5 *m/z*, fragmentation of the top 10 most intense full scan ions with Orbitrap detection at a resolution of 15,000, an AGC setting of  $5 \times 10^4$ , and a maximum ion injection time of 200 ms. The analysis of authentic standards utilized CID fragmentation with a constant collision energy of 30% and maximum ion injection time of 75 ms. The analysis of *pks*<sup>+</sup> infected cells utilized a HCD fragmentation with a stepped collision energy of 5, 15 and 25 % and maximum ion injection time of 200 ms. Data-dependent parameters were as follows: a triggering threshold of  $2.0 \times 10^4$ , repeat count of 1, and exclusion duration of 15 s. An exclusion mass list of the most intense ions observed in ctDNA was excluded from fragmentation in the analysis of HeLa cells treated with *pks*<sup>+</sup> *E. coli* ( $\pm 5$  ppm). No masses were excluded in the analysis of the authentic standards. MS<sup>3</sup> HCD fragmentation scans (2.5 *m/z* isolation width, collision energy of 30%) with Orbitrap detection at a resolution of 15,000 were triggered upon observation of neutral losses of 116.0474, 151.0494, 135.0545, 126.0429 and 111.0433 *m/z*. A minimal product ion signal of  $1.0 \times 10^4$  was used. All spectra were acquired with the EASY-IC lock mass (202.0777 *m/z*) enabled.

### Data Processing.

Raw data files were converted to mzML format and centroid mode using MSConvert (ProteoWizard) and imported into MZmine.<sup>18,24</sup> All MZmine data processing utilized a mass tolerance of 5 ppm. Automated detection of DNA adducts was performed using the DFBuilder module. Data-dependent DFFBuilder parameters were matched to the scan range and chromatographic parameters of each experiment, respectively. Diagnostic ion thresholds were matched to the CNL-MS<sup>3</sup> triggering threshold, detailed above. Extraction ion chromatogram (EIC) retention time (RT) tolerances of 1.5 and 0.5 min were used for the analysis of authentic standards and *pks* infected cells, respectively. An exclusion list of contaminant and background signals was constructed by analyzing blank injections. EIC were deconvoluted using the Local Minimum search algorithm, including a chromatographic threshold of 30% and minimum peak top/edge ratio of 5. Duplicate peaks and isotopes were grouped with a retention time tolerance of 0.1 min. Putative adduct peaks were aligned between raw data files using the Join Aligner algorithm with a tolerance of 0.3 min. Compound cation adducts, neutral losses, fragments, and complexes were annotated. Missing peak values were estimated using Gap-filling with a retention time tolerance of 0.3 min. Features detected in all three DNA adducts standards mixtures were retained in the final feature list, and putative adduct features unique to the *pks*<sup>+</sup> transiently infected HeLa cells were retained in the feature list. The final feature list was exported to a CSV file for manual review. A complete list of processing parameters is summarized in Table S-1.

## RESULTS AND DISCUSSION

### Workflow Overview.

We developed the DFBuilder module to apply the diagnostic fragmentation filtering algorithm as part of a fully automated data processing workflow for the analysis of any data-dependent MS/MS data set. The workflow components and processing design are presented in Figure 1. The DFBuilder module monitors for user-defined product ions and neutral losses in tandem MS/MS spectra in input raw data files and builds EICs for precursors of interest (Table S-2). To minimize false discovery rates, we designed DFBuilder to utilize mass tolerances, signal thresholds, and retention time limits to ensure only high-quality spectra and peaks are detected (Figure S-1, Top Panel). An optional exclusion list removes contaminants detected in blank injections or previously defined signals (Table S-3). The DFBuilder module exports the detected precursor targeted-EIC as feature lists for further downstream processing in the analytical workflow (Figure S-1, Bottom Panel). Alternatively, the module may be used stand-alone to export precursor m/z, retention time, and tandem MS/MS scan information of interest to a CSV file for manual review.

The DFBuilder module serves as the entrance to a complete analytical workflow for DNA adduct discovery and annotation using the MZmine platform. Following diagnostic fragmentation filtering, additional processing curates the resultant feature list to remove low-quality or redundant peaks. Then, chromatographic deconvolution eliminates non-reproducible features from downstream analysis and provides more accurate peak quantitation. Grouping duplicate hits, in-source fragments, and compound complexes simplifies the results to a feature list of monoisotopic precursors. Finally, retention time alignment corrects spectral drift commonly observed between repeated injections, and feature annotations inadvertently excluded during the previous steps may be estimated with Gap-filling. The end product of our DFBuilder workflow is a feature list of high-quality, non-redundant putative DNA adducts. Following data analysis, peak quantitation estimates may be exported for external statistical review and compound identities assessed manually or with assistance from spectral library searching algorithms. The ability to post hoc update diagnostic patterns and processing parameters for optimized analysis is the main strength of our workflow compared to instrument-based approaches. Through the use of open-source data formats, our approach is not limited to vendor-specific raw files and may be utilized with multiple mass detector types, including ion traps, quadrupole time-of-flight, and other hybrid instruments. The DFBuilder module and workflow is compatible with batch-based analysis to ensure computational repeatability in large-scale studies. In the following sections, we demonstrate the capability of diagnostic fragmentation filtering to detect DNA adducts in complex mixtures and the potential of our workflow to serve as part of an automated pipeline for DNA adductomics analysis. Although the following experiments focus on the detection of DNA adducts, diagnostic fragmentation filtering and the DFBuilder workflow may be applied to any class of molecule with diagnostic fragmentation, including lipids, glycoproteins, and other complex modified compounds.



## Automated Detection of Authentic Standards using the DFBuilder Workflow.

For validation of the DFBuilder workflow, we analyzed a mixture of nine DNA adduct standards (Figure 2). The authentic standard mixture contains adducts of all four nucleobases, which possess polarities with enough difference to cover a wide chromatographic range. The loss of the 2'-deoxyribose moiety (116.0474 amu) is the most selective transition for the detection of DNA nucleoside adducts using CNL screening.<sup>25</sup> Recent studies have demonstrated that neutral loss monitoring of the nucleobases, including cytosine (111.0433 amu), thymine (126.0429 amu), adenine (135.0545 amu), and guanine (151.0494 amu), expands the coverage to DNA nucleobase adducts.<sup>14</sup> The application of the DFBuilder module applied here used diagnostic fragmentation filtering of the neutral losses of deoxyribose and the four nucleobases to monitor for the presence of nucleoside and nucleobase DNA adducts in solution. Prior to the analysis of the standard mixture, a solvent blank injection was analyzed to assess the presence of low abundant contaminant signals in our system. In total, 54 MS/MS spectra exhibited a diagnostic fragmentation. These contaminant signals populated an exclusion list for the subsequent analysis.

The application of the DFBuilder module to analyze the authentic standard mixtures detected 48 tandem MS/MS exhibiting a diagnostic neutral loss. Of these spectra, 47 exhibited a neutral loss of 2'-deoxyribose and the final signal displayed a neutral thymine nucleobase loss. Empirical evaluation of the resultant targeted EICs revealed many of the detected spectra represented duplicate hits from repeated MS/MS fragmentation of a single precursor ion. After filtering poor-quality and duplicate peaks, 18 putative DNA adduct signals remained. Precursor annotation indicated that six of these signals were characterized to be in-source fragments or sodium adducts  $[M+Na]^+$  formed during ionization. At the completion of our workflow, 12 monoisotopic precursors were categorized as putative DNA adducts in the untargeted analysis of the authentic standard mixture (Table 1). Diagnostic fragmentation filtering detected all nine authentic standards from the characteristic 2'-deoxyribose neutral loss induced during CID fragmentation. The identity for each standard was confirmed by manual inspection of the acquired MS<sup>2</sup> and MS<sup>3</sup> spectra. No apparent relationship between the authentic standards and the three unidentified signals could be made. Manual evaluation of the acquired spectra confirmed the presence of a 2'-deoxyribose neutral loss among these features. The precursor of  $m/z$  211.1329  $[M+H]^+$  at 27.13 min was determined to be a persistent background contaminant that failed to trigger a fragment spectra in the blank injection. The remaining signals of  $m/z$  296.1353  $[M+H]^+$  and 398.1669  $[M+H]^+$  at 28.51 and 22.32 min, respectively, represent potential impurities or contaminants in our standard mixture based on their reproducible detection and fragmentation patterns.

## Reproducible Detection of Colibactin-Induced DNA Adducts.

Colibactin is a genotoxic secondary metabolite produced by commensal and extra-intestinal pathogenic strains of *E. coli* harboring the *pks* genomic island (*pks*<sup>+</sup>). Infection with colibactin-harboring microbes influences tumorigenesis in multiple mouse models of colitis-associated colorectal cancer.<sup>27,28</sup> Despite its strong relationship with disease, the genotoxic mechanism driving pathogenesis has remained uncertain, although recent research has revealed colibactin alkylates DNA *in vivo*, producing large genomic cross-links and DNA adducts.<sup>14,23</sup> A DDA-CNL-MS<sup>3</sup> adductomics approach identified two diastereomeric

adducts at  $m/z$  540.1765  $[M+H]^+$  exclusively in cells transiently infected with  $pks^+$  *E. coli* compared to negative controls ( $pks^-$  *E. coli*).<sup>23</sup> Here, we have utilized data generated in this study to demonstrate the potential of our workflow to detect DNA modifications in complex matrices.

We applied the automated DFBuilder workflow for the discovery of colibactin-induced DNA adducts in HeLa cells transiently infected with  $pks^+$  and  $pks^-$  *E. coli*. Tandem MS/MS fragmentation filtering utilized all previously mentioned DNA-specific neutral losses. In total, 212 MS/MS spectra exhibited a diagnostic neutral loss. After removing non-reproducible peaks, duplicate hits, isotopes, and ion complexes, 56 features remained. Three putative DNA adduct features were unique to  $pks^+$  infected cells and exhibited no signal among negative controls. In agreement with the original findings, we detected the two colibactin-induced DNA adducts of  $m/z$  540.1765 at 16.92 and 17.50 min, respectively (Figure S-2). Our workflow discovered one additional putative adduct at 7.68 min exhibiting an  $m/z$  of 342.1680  $[M+H]^+$  (Figure 3A). As this putative adduct was not reported in the original publication, we further evaluated its properties. The putative adduct exhibited a neutral loss of adenine (135.0545 amu) in the acquired MS/MS fragment spectra (Figure 3B) and triggered a CNL-MS<sup>3</sup> spectrum for the resulting product peak of  $m/z$  207.1134 (Figure 3C). The high-resolution accurate mass measurement of 342.1680  $[M+H]^+$  yielded a molecular formula of C<sub>16</sub>H<sub>19</sub>N<sub>7</sub>O<sub>2</sub> (calculated, 342.1673) with 11 degrees of unsaturation. Tandem MS/MS fragmentation exhibited one major fragment ion of  $m/z$  207.1134  $[M+H\text{-Adenine}]^+$ . Product ion CNL-MS<sup>3</sup> fragmentation displayed three additional fragment ions of  $m/z$  179.0825  $[M+H\text{-C}_2\text{H}_4]^+$ , 152.0715  $[M+H\text{-C}_3\text{H}_5\text{N}]^+$ , and 124.0766  $[M+H\text{-C}_4\text{H}_5\text{NO}]^+$ . A mass tolerance of 10 ppm was used for the assignment of the fragment ions.

Based on these mass spectra, we propose the  $m/z$  342.1680 parent ion structure (adduct 342; Figure 4) to be a 4-hydroxy-pyrrolidin-2-one, which closely resembles the structures of the  $m/z$  540.1765 adducts and other precolibactin metabolites.<sup>23,29</sup> In Figure 4, we provide a fragmentation mechanism that not only rationalizes the appearance of all detected MS<sup>2</sup> and MS<sup>3</sup> ions but also predicts their structures, which all support our proposal for adduct 342. The most striking feature of adduct 342 is the enol hydroxyl group located where a carbon side chain normally resides. We posit that the hydroxyl group is introduced through formal hydrolysis either via a retro-aldol or retro-Mannich type mechanism (Figure S-3). Importantly, this mechanism suggests adduct 342 is a novel decomposition product of colibactin and its DNA cross-links and offers another explanation for colibactin's notorious instability. However, in the absence of internal standards and additional experimentation, we are currently unable to unambiguously confirm the structures of adduct 342 and its fragment ions. Additionally, the lack of standards limits our ability to quantify adduct 342 and understand its relative contribution to colibactin DNA adduct decomposition and its potential as a biomarker.

### Expanding the Detection of Putative DNA Adducts.

The strength of diagnostic fragmentation filtering is the ability to alter analysis parameters post-acquisition and probe any diagnostic fragmentation patterns characteristic of the compound class of interest.<sup>15</sup> The DDA-CNL-MS<sup>3</sup> screening approach used to date



biases adduct discovery to low molecular weight, singly-charged ions and is limited to previously described neutral loss fragmentation pathways.<sup>12</sup> It has been recently observed that the ionization of DNA cross-links (two or more nucleosides connected through a covalent modification) and other bulky covalent modifications can be dominated by double-charge ionization and can exhibit more complex fragmentation patterns not predictably characterized by single neutral loss transitions (e.g. the 2'-deoxyribose moiety).<sup>30</sup> In addition, the fragmentation of depurination DNA adducts lacking basic sites will result in the production of the charged nucleic acid (e.g. *N*<sup>7</sup>-ethylguanine). To demonstrate the potential of our DFBuilder workflow to evaluate new fragmentation patterns, we re-analyzed the previous HeLa cell colibactin-*pks* experiment with updated monitoring criteria. We supplemented our list of fragmentation patterns to include product ions of each nucleobase: cytosine (112.0506 *m/z*), thymine (127.0502 *m/z*), adenine (136.0618 *m/z*), and guanine (152.0567 *m/z*).

Using the modified pattern list, a total of 1,069 MS/MS spectra exhibited diagnostic fragmentation. After removing non-reproducible peaks, duplicate hits, isotopes, and ion complexes, our workflow detected 343 putative DNA adducts in the *pks*<sup>+</sup> and *pks*<sup>-</sup> infected HeLa cell DNA. The inclusion of nucleobase product ions resulted in a six times increase in the number of detected spectra relative to a CNL search. Over one-third of these signals represented doubly [M+2H]<sup>2+</sup> or triply charged [M+3H]<sup>3+</sup> ion species. In contrast, only two features detected via CNL monitoring were multiply charged. A total of seven putative DNA adduct features were found to be unique to *pks*<sup>+</sup> infected cells. In addition to the three CNL-detected features, four novel putative adducts were detected via the presence of an adenine product ion upon fragmentation, exhibiting an *m/z* of 333.5243 [M+3H]<sup>3+</sup> at 14.91 min, 499.7857 [M+2H]<sup>2+</sup> at 15.09 min, 535.2576 [M+3H]<sup>3+</sup> at 15.67 min, and 423.6687 [M+2H]<sup>2+</sup> at 24.33 min (Figure 5A-D). In contrast to the neutral loss-derived putative adducts, all of the newly detected putative DNA adducts were multiply charged. Two of these features, *m/z* of 499.7857 [M+2H]<sup>2+</sup> and 333.5254 [M+3H]<sup>3+</sup>, likely represent multiply charged variations of the same compound due to their identical neutral mass of 997.5568 amu and overlapping elution patterns (Figure 5A,C). With only tandem MS/MS fragmentation, elucidating chemical structures for these signals is challenging. Hundreds of molecular formulae are possible for each precursor, even with high-resolution accurate mass measurement. These bulky putative adducts could represent DNA interstrand cross-links missed through CNL monitoring alone. However, these features do not correspond with any proposed colibactin-associated DNA cross-link structures.<sup>30,31</sup> Unlike the near-universal loss of deoxyribose for singly charged DNA adducts, the propensity of multiply charged cross-linked DNA adducts to fragment to protonated nucleobases has not been established. Therefore, it is possible that these signals are false positive signals from background contamination. In the absence of a comprehensive evaluation of the sensitivity and specificity of the nucleobase product ions, we are unable to assess the false discovery rate of this method. However, it should be noted that the ion signals attributed to the putative multiply charged cross-linked DNA adducts were not present in the negative control sample. With a quadrupole isolation window of 1.5 *m/z*, co-fragmentation of independent precursors may inadvertently occur during an analysis. A thorough evaluation will be required to fully assess the diagnostic potential of product ions for the discovery of DNA adducts.

The strength of post-acquisition monitoring of characteristic neutral loss and product ions is the expedient evaluation of any fragmentation pattern for class-specific small molecule discovery. Although the results presented here focus on the detection of DNA adducts using a hybrid Orbitrap mass analyzer, this approach may be applied to any other compound class with diagnostic fragmentation using any mass analyzer capable of generating MS/MS fragmentation in either positive or negative ion polarity. We foresee this approach being especially useful for the detection of other conjugated and covalently modified compounds indicative of chemical exposure, such as the cleavage of the thioether bond that is characteristic of mercapturic acids conjugates in negative ion mode LC-MS/MS (129.0426 amu).<sup>32</sup> Additionally, the DFBuilder approach may be used to discriminate specific moieties of diverse compound classes, such as saturated or unsaturated fatty acids. Using complex diagnostic pattern combinations, users may simultaneously discriminate chain length neutral losses and head group fragment ions for selective detection and characterization of glycerophospholipids and other complexes.

## CONCLUSION

We developed the DFBuilder MZmine module and automated workflow for the detection of DNA adducts using diagnostic fragmentation filtering. Our method screens any data-dependent MS/MS spectra for fragmentation patterns (neutral losses and product ions) and builds targeted precursor EICs in the corresponding chromatographic region. Using the suite of characterization, filtering, and alignment tools available in MZmine, the resulting peak list is automatically reduced to a list of high confidence, reproducible putative DNA adduct signals. In contrast to existing techniques, our method can be used with any tandem mass spectrometer and may be applied post-acquisition for repeated evaluation of experimental results. Using a mixture of authentic standards, we validated our workflow for the automated detection of DNA adducts with minimal false detection. We demonstrated the ability of our approach to reproduce results and discover novel putative adducts in complex mixtures. Using the DFBuilder workflow, data analysis of LC-MS/MS DNA adductomics experiments is simplified and enables further development of fragmentation filtering for the detection of DNA adducts and other compounds.

The source code repository for the DFBuilder module and our executable MZmine 2.54 forked repository is freely available on Github (<https://github.com/Kevin-Murray/mzmine2-DFBuilder>). The DFBuilder module will be incorporated into main branch of MZmine in the next available update. The DNA adduct standard data used in this project are available for public download from the Metabolomics Workbench, study ID: ST001661.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGEMENT

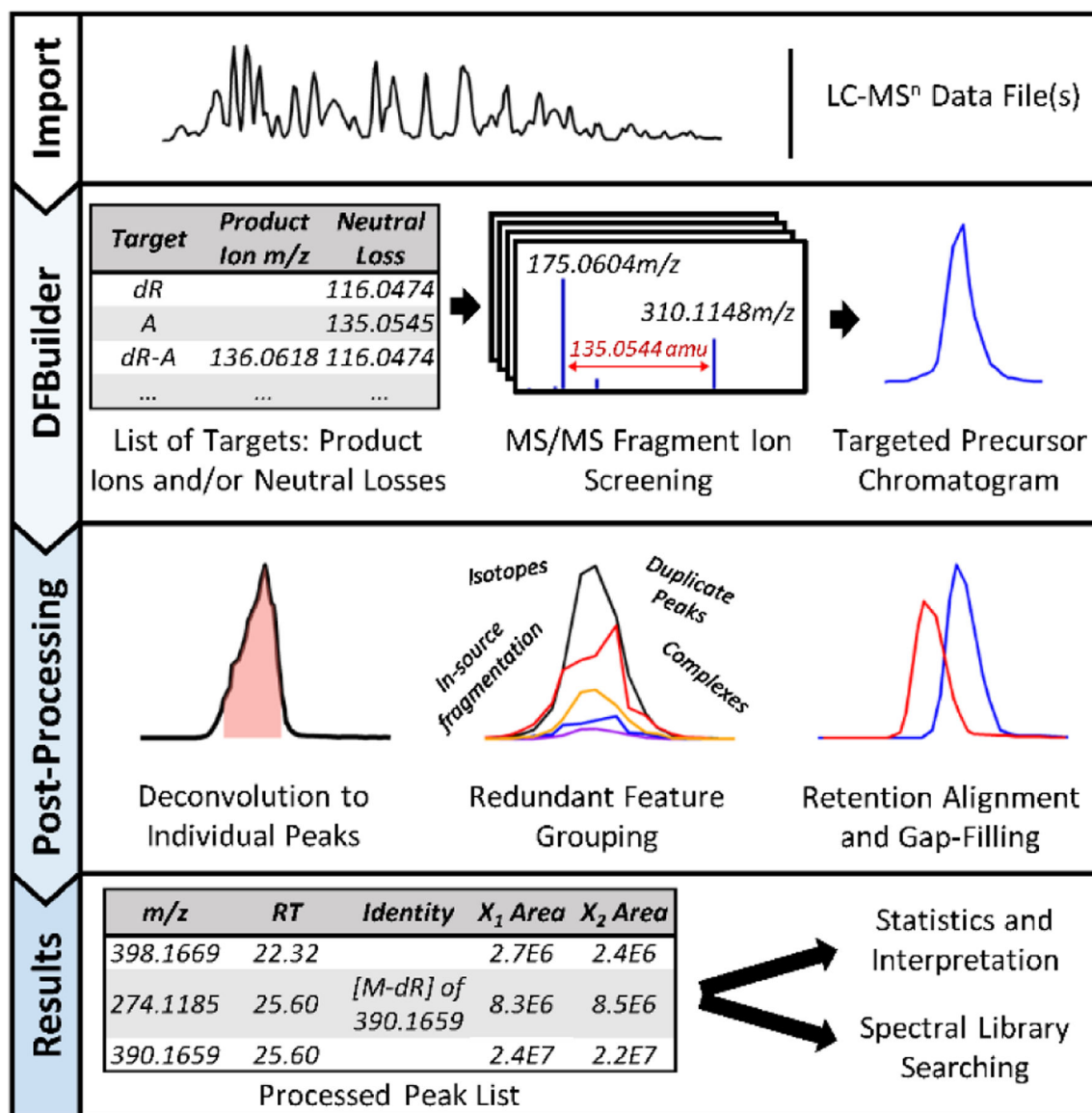
The authors would like to thank Dr. Romel Dator and Dr. Morwena Solivio of Masonic Cancer Center of the University of Minnesota for their comments and recommendations during workflow testing, and Dr. Ansgar Korf for advice developing an MZmine module. This work was funded in part by NIH grants R01 CA220376, R01 CA208834, and F32 CA254165, the Packard Fellowship for Science and Engineering, and the Damon

Runyon-Rachleff Innovation Award. Salary support for P.W.V. was provided by the National Cancer Institute (R50-CA211256) and mass spectrometry analysis was supported by the Masonic Cancer Center's National Cancer Institute support grant CA077598.

## REFERENCES

- (1). Dipple A DNA Adducts of Chemical Carcinogens; 1995; Vol. 16.
- (2). Wogan GN; Hecht SS; Felton JS; Conney AH; Loeb LA Environmental and Chemical Carcinogenesis. *Semin. Cancer Biol* 2004, 14 (6), 473–486. [PubMed: 15489140]
- (3). Wiencke JK DNA Adduct Burden and Tobacco Carcinogenesis. *Oncogene* 2002, 21 (48 REV. ISS. 6), 7376–7391. [PubMed: 12379880]
- (4). Xiao S; Guo J; Yun BH; Villalta PW; Krishna S; Tejpaul R; Murugan P; Weight CJ; Turesky RJ Biomonitoring DNA Adducts of Cooked Meat Carcinogens in Human Prostate by Nano Liquid Chromatography-High Resolution Tandem Mass Spectrometry: Identification of 2-Amino-1-Methyl-6-Phenylimidazo[4,5-b]Pyridine DNA Adduct. *Anal. Chem* 2016, 88 (24), 12508–12515. [PubMed: 28139123]
- (5). Singer B; Grunberger D Molecular Biology of Mutagens and Carcinogens; Springer US, 1983.
- (6). Balbo S; Hecht SS; Upadhyaya P; Villalta PW Application of a High-Resolution Mass-Spectrometry-Based DNA Adductomics Approach for Identification of DNA Adducts in Complex Mixtures. *Anal. Chem* 2014, 86 (3), 1744–1752. [PubMed: 24410521]
- (7). Guo J; Turesky RJ Emerging Technologies in Mass Spectrometry-Based DNA Adductomics. *High-throughput* 2019, 8 (2).
- (8). Balbo S; Turesky RJ; Villalta PW DNA Adductomics. *Chemical Research in Toxicology American Chemical Society* March 17, 2014, pp 356–366.
- (9). Farmer PB; Singh R Use of DNA Adducts to Identify Human Health Risk from Exposure to Hazardous Environmental Pollutants: The Increasing Role of Mass Spectrometry in Assessing Biologically Effective Doses of Genotoxic Carcinogens. *Mutation Research - Reviews in Mutation Research* 7 2008, pp 68–76.
- (10). Gangl ET; Turesky RJ; Vouros P Determination of in Vitro- and in Vivo-Formed Dna Adducts of 2-Amino-3- Methylimidazo[4,5-f]Quinoline by Capillary Liquid Chromatography/Microelectrospray Mass Spectrometry. *Chem. Res. Toxicol* 1999, 12 (10), 1019–1027. [PubMed: 10525280]
- (11). Wolf SM; Vouros P Application of Capillary Liquid Chromatography Coupled with Tandem Mass Spectrometric Methods to the Rapid Screening of Adducts Formed by the Reaction of N-Acetoxy-N-Acetyl-2-Aminofluorene with Calf Thymus DNA. *Chem. Res. Toxicol* 1994, 7 (1), 82–88. [PubMed: 8155830]
- (12). Guo J; Villalta PW; Turesky RJ Data-Independent Mass Spectrometry Approach for Screening and Identification of DNA Adducts. *Anal. Chem* 2017, 89 (21), 11728–11736. [PubMed: 28977750]
- (13). Tsugawa H; Cajka T; Kind T; Ma Y; Higgins B; Ikeda K; Kanazawa M; Vandergheynst J; Fiehn O; Arita M MS-DIAL: Data-Independent MS/MS Deconvolution for Comprehensive Metabolome Analysis. *Nat. Methods* 2015, 12 (6), 523–526. [PubMed: 25938372]
- (14). Stornetta A; Villalta PW; Hecht SS; Sturla SJ; Balbo S Screening for DNA Alkylation Mono and Cross-Linked Adducts with a Comprehensive LC-MS3 Adductomic Approach. *Anal. Chem* 2015, 87 (23), 11706–11713. [PubMed: 26509677]
- (15). Walsh JP; Renaud JB; Hoogstra S; McMullin DR; Ibrahim A; Visagie CM; Tanney JB; Yeung KK-C; Sumarah MW. Diagnostic Fragmentation Filtering for the Discovery of New Chaetoglobosins and Cytochalasins. *Rapid Commun. Mass Spectrom* 2019, 33 (1), 133–139. [PubMed: 30325552]
- (16). Roushan A; Wilson GM; Kletter D; Sen KI; Tang WH; Kil YJ; Carlson E; Bern M Peak Filtering, Peak Annotation, and Wildcard Search for Glycoproteomics. *Mol. Cell. Proteomics* 2020, 13, mcp.RA120.002260.
- (17). Medzihradsky KF; Kaasik K; Chalkley RJ Characterizing Sialic Acid Variants at the Glycopeptide Level. *Anal. Chem* 2015, 87 (5), 3064–3071. [PubMed: 25654559]

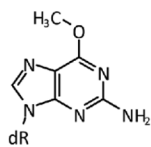
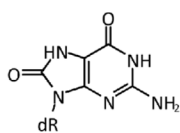
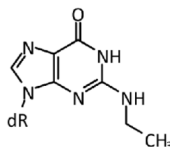
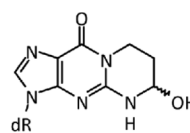
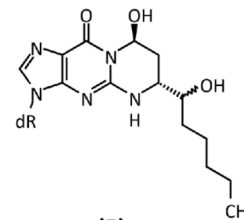
- (18). Pluskal T; Castillo S; Villar-Briones A; Orešič M MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data. *BMC Bioinformatics* 2010, 11 (1), 395. [PubMed: 20650010]
- (19). Lao Y; Villalta PW; Sturla SJ; Wang M; Hecht SS Quantitation of Pyridyloxobutyl DNA Adducts of Tobacco-Specific Nitrosamines in Rat Tissue DNA by High-Performance Liquid Chromatography- Electrospray Ionization-Tandem Mass Spectrometry. *Chem. Res. Toxicol* 2006, 19 (5), 674–682. [PubMed: 16696570]
- (20). Upadhyaya P; Kalscheuer S; Hochalter JB; Villalta PW; Hecht SS Quantitation of Pyridylhydroxybutyl-DNA Adducts in Liver and Lung of F-344 Rats Treated with 4-(Methylnitrosamino)-1-(3-Pyridyl)-1-Butanone and Enantiomers of Its Metabolite 4-(Methylnitrosamino)-1-(3-Pyridyl)-1-Butanol. *Chem. Res. Toxicol* 2008, 21 (7), 1468–1476. [PubMed: 18570389]
- (21). Zhang S; Villalta PW; Wang M; Hecht SS Detection and Quantitation of Acrolein-Derived 1,N2-Propanodeoxyguanosine Adducts in Human Lung by Liquid Chromatography- Electrospray Ionization-Tandem Mass Spectrometry. *Chem. Res. Toxicol* 2007, 20 (4), 565–571. [PubMed: 17385896]
- (22). Guidolin V; Carra' A; Carlson ES; Villalta PW; Balbo S Molecular Characterization of Alcohol-Induced DNA Damage for Cancer Prevention. Poster Presentation at the ACS Conference Boston, Massachusetts 2018.
- (23). Wilson MR; Jiang Y; Villalta PW; Stornetta A; Boudreau PD; Carrá A; Brennan CA; Chun E; Ngo L; Samson LD; Engelward BP; Garrett WS; Balbo S; Balskus EP The Human Gut Bacterial Genotoxin Colibactin Alkylates DNA. *Science* (80-. ) 2019, 363 (6428).
- (24). Chambers MC; MacLean B; Burke R; Amodei D; Ruderman DL; Neumann S; Gatto L; Fischer B; Pratt B; Egerton J; Hoff K; Kessner D; Tasman N; Shulman N; Frewen B; Baker TA; Brusniak M-YY; Paulse C; Creasy D; Flashner L; Kani K; Moulding C; Seymour SL; Nuwaysir LM; Lefebvre B; Kuhlmann F; Roark J; Rainer P; Detlev S; Hemenway T; Huhmer A; Langridge J; Connolly B; Chadick T; Holly K; Eckels J; Deutsch EW; Moritz RL; Katz JE; Agus DB; MacCoss M; Tabb DL; Mallick P A Cross-Platform Toolkit for Mass Spectrometry and Proteomics; Nature Publishing Group, 2012; Vol. 30, pp 918–920.
- (25). Bryant MS; Lay JO; Chiarelli MP Development of Fast Atom Bombardment Mass Spectral Methods for the Identification of Carcinogen-Nucleoside Adducts. *J. Am. Soc. Mass Spectrom* 1992, 3 (4), 360–371. [PubMed: 24243047]
- (26). Pluskal T; Uehara T; Yanagida M Highly Accurate Chemical Formula Prediction Tool Utilizing High-Resolution Mass Spectra, MS/MS Fragmentation, Heuristic Rules, and Isotope Pattern Matching. *Anal. Chem* 2012, 84 (10), 4396–4403. [PubMed: 22497521]
- (27). Cuevas-Ramos G; Petit CR; Marcq I; Boury M; Oswald E; Nougayrède JP *Escherichia Coli* Induces DNA Damage in Vivo and Triggers Genomic Instability in Mammalian Cells. *Proc. Natl. Acad. Sci. U. S. A* 2010, 107 (25), 11537–11542. [PubMed: 20534522]
- (28). Buc E; Dubois D; Sauvanet P; Raisch J; Delmas J; Darfeuille-Michaud A; Pezet D; Bonnet R High Prevalence of Mucosa-Associated *E. Coli* Producing Cyclomodulin and Genotoxin in Colon Cancer. *PLoS One* 2013, 8 (2), e56964. [PubMed: 23457644]
- (29). Faïs T; Delmas J; Barnich N; Bonnet R; Dalmasso G Colibactin: More than a New Bacterial Toxin. *Toxins (Basel)* 2018, 10 (4), 16–18.
- (30). Jiang Y; Stornetta A; Villalta PW; Wilson MR; Boudreau PD; Zha L; Balbo S; Balskus EP Reactivity of an Unusual Amidase May Explain Colibactin's DNA Cross-Linking Activity. *J. Am. Chem. Soc* 2019, 141 (29), 11489–11496. [PubMed: 31251062]
- (31). Xue M; Kim CS; Healy AR; Wernke KM; Wang Z; Frischling MC; Shine EE; Wang W; Herzon SB; Crawford JM Structure Elucidation of Colibactin and Its DNA Cross-Links. *Science* (80-. ) 2019, 365 (6457).
- (32). Jamin EL; Costantino R; Mervant L; Martin JF; Jouanin I; Blas-Y-Estrada F; Guéraud F; Debrauwer L Global Profiling of Toxicologically Relevant Metabolites in Urine: Case Study of Reactive Aldehydes. *Anal. Chem* 2020, 92 (2), 1746–1754. [PubMed: 31854978]



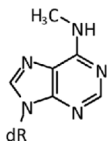
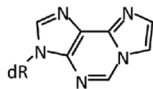
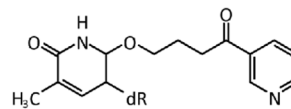
**Figure 1.**

Data analysis workflow using DFBuilder. Using a list of diagnostic fragmentation patterns, LC-MS<sup>n</sup> data files are screened for putative DNA adducts and targeted-EICs are built for precursors of interest. Chromatographic features are further processed to produce a peak list of high-quality, reproducible putative DNA adducts. The final result of the workflow is a feature list of quantitative peak metrics and associated fragment spectra for each input data file that may be utilized for additional statistical analyses and identification purposes.

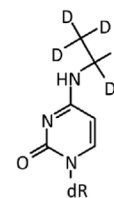
## Deoxyguanosine adducts

**(1)**  
*O*<sup>6</sup>-Me-dG**(2)**  
8-Oxo-dG**(3)**  
*N*<sup>2</sup>-ethyl-dG**(4)**  
OH-PdG**(5)**  
HNE-dG

## Deoxyadenosine adducts

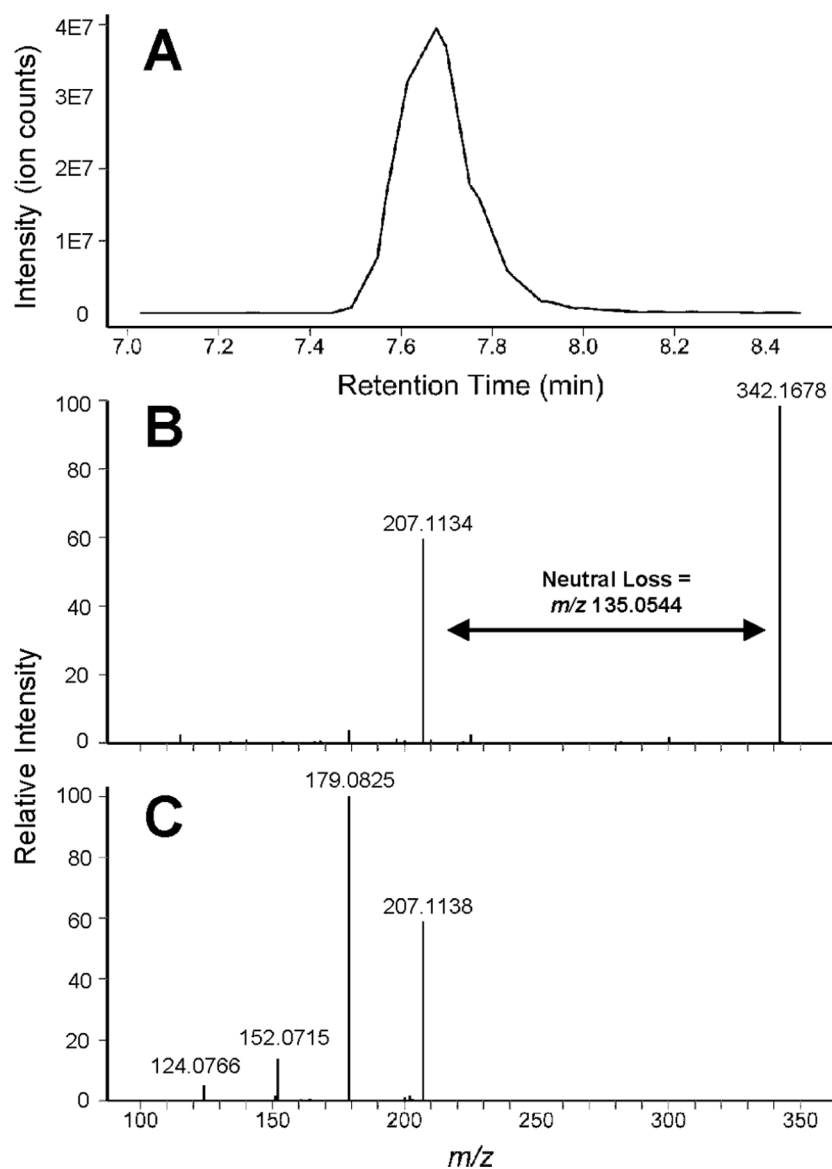
**(6)**  
*N*<sup>6</sup>-Me-dA**(7)**  
ε-dA**(8)**  
*O*<sup>2</sup>-POB-dT

## Deoxycytosine adduct

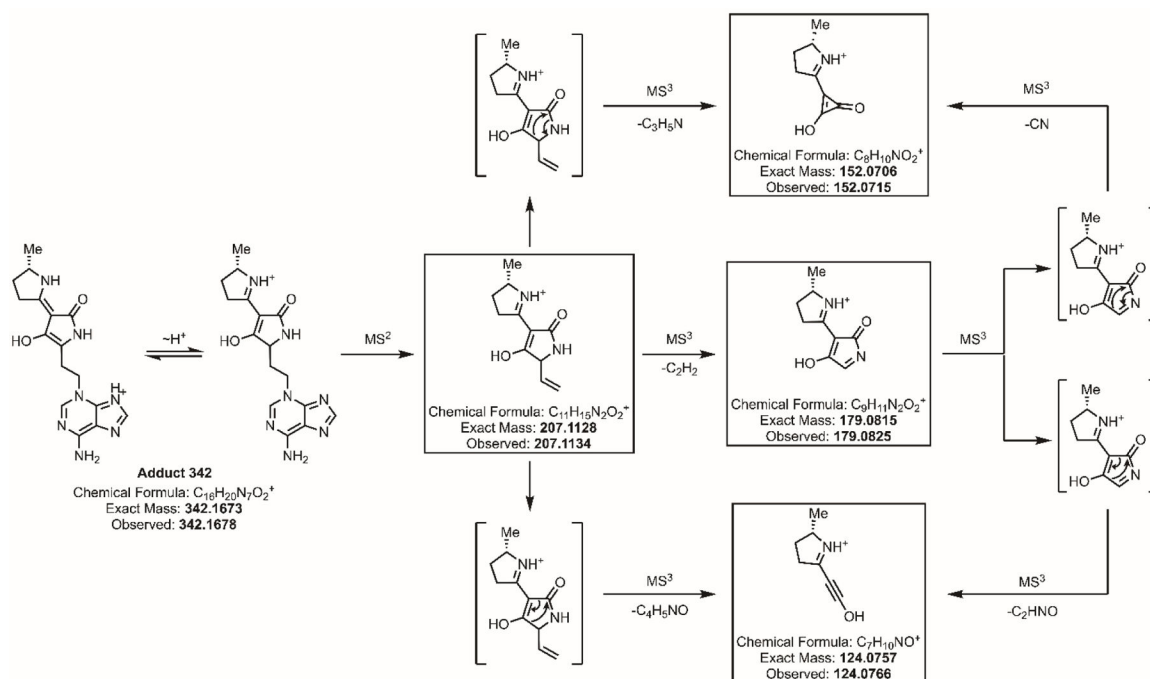
**(9)**  
D<sub>5</sub>-ethyl-dC**Figure 2.**

Structures of the DNA adducts in the authentic standard mixture, dR = 2'-deoxyribose.

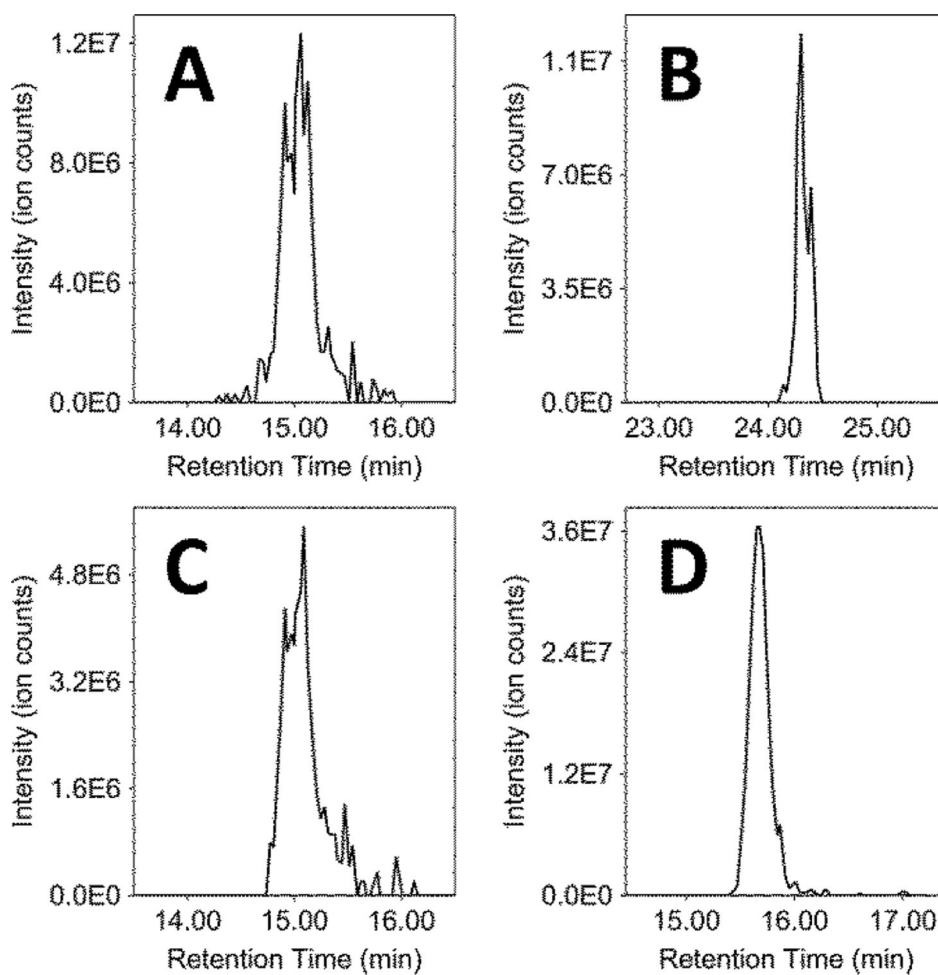




**Figure 3.** LC-MS characteristics of novel colibactin-induced DNA adduct; (A) EIC of precursor m/z 342.1673 [M+H]<sup>+</sup> at 7.71 min; (B) MS<sup>2</sup> precursor fragmentation exhibiting the neutral loss of adenine; (C) CNL-triggered MS<sup>3</sup> on product ion m/z 207.1134.



**Figure 4.** Proposed structure of adduct 342 and mechanism for  $MS^2$  and  $MS^3$  fragmentation. The proposed structures of all observed ions are enclosed and accompanied by their chemical formula, exact mass, and the  $m/z$  experimentally observed by LC-MS analysis.



**Figure 5.** EICs of putative colibactin-induced DNA adducts detected by adenine fragment monitoring exhibiting an  $m/z$  of; (A) 333.5243  $[M+3H]^{3+}$  at 14.91, (B) 423.6687  $[M+2H]^{2+}$  at 24.33 min, (C) 499.7857  $[M+2H]^{2+}$  at 15.09 min, and (D) 535.2576  $[M+3H]^{3+}$  at 15.67 min.

**Table 1.**  
Results of Automated DFBuilder Workflow on Mixture of Authentic DNA Adduct Standards

No.	m/z	Retention time (min)	Compound Identity	Neutral Loss	Neutral Mass	Neutral Formula [Predicted*]
1	261.1605	9.34	D <sub>5</sub> -ethyl-dC	dR	260.1532	C <sub>11</sub> H <sub>17</sub> D <sub>3</sub> N <sub>3</sub> O <sub>4</sub>
2	266.1247	16.62	N <sup>6</sup> -Me-dA	dR	265.1174	C <sub>11</sub> H <sub>15</sub> N <sub>5</sub> O <sub>3</sub>
3	276.1091	12.43	e-dA	dR	275.1018	C <sub>12</sub> H <sub>13</sub> N <sub>5</sub> O <sub>3</sub>
4	282.1197	21.50	O <sup>6</sup> -me-dG	dR	281.1124	C <sub>11</sub> H <sub>15</sub> N <sub>5</sub> O <sub>4</sub>
5	284.0990	15.76	8-oxo-dG	dR	283.0917	C <sub>10</sub> H <sub>13</sub> N <sub>5</sub> O <sub>5</sub>
6	296.1352	26.08	N <sup>2</sup> -ethyl-dG	dR	295.1279	C <sub>12</sub> H <sub>17</sub> N <sub>5</sub> O <sub>4</sub>
7	324.1303	16.96	OH-PdG	dR	323.1230	C <sub>13</sub> H <sub>17</sub> N <sub>5</sub> O <sub>5</sub>
8	390.1659	25.60	O <sup>2</sup> -POB-dT	dR	389.1586	C <sub>19</sub> H <sub>23</sub> N <sub>5</sub> O <sub>6</sub>
9	424.2189	37.65	HNE-dG	dR	423.2116	C <sub>19</sub> H <sub>29</sub> N <sub>5</sub> O <sub>6</sub>
10	211.1329	27.13	Unknown #1	dR	210.1256	C <sub>12</sub> H <sub>18</sub> O <sub>3</sub> *
11	296.1353	28.51	Unknown #2	dR	295.1280	C <sub>12</sub> H <sub>17</sub> N <sub>5</sub> O <sub>4</sub> *
12	398.1669	22.32	Unknown #3	dR	397.1596	C <sub>16</sub> H <sub>23</sub> N <sub>5</sub> O <sub>7</sub> *

\* Neutral formula predicted using natural elemental and heuristic constraints with a 5 ppm mass tolerance.<sup>26</sup>