Similarity of High-Resolution Tandem Mass Spectrometry Spectra of Structurally-Related Micropollutants and Transformation Products

Jennifer E. Schollée,*[a,b] Emma L. Schymanski,[a] Michael A. Stravs,[a,b] Rebekka Gulde,[a] Nikolaos S. Thomaidis,[c] and Juliane Hollender[a,b]

[a]Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland

[b]Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, 8092 Zürich, Switzerland

[c]Laboratory of Analytical Chemistry, Department of Chemistry, National and Kapodistrian University of Athens, 157 71 Athens, Greece

*Corresponding Author

Author e-mail addresses:

Jennifer.schollee@eawag.ch

Emma.schymanski@eawag.ch

Michael.stravs@eawag.ch

Rebekka.gulde@eawag.ch

Ntho@chem.uoa.gr

Juliane.hollender@eawag.ch

Phone numbers of corresponding authors:

Jennifer Schollée +41-58-765-5512

Address reprint requests to Jennifer Schollée, Ueberlandstrasse 133, 8600 Duebendorf, Switzerland, +41-58-765-5512, +41-58-765-5028, Jennifer.schollee@eawag.ch

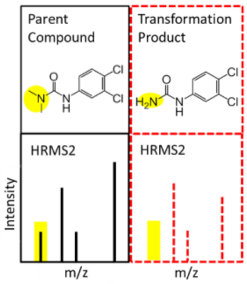
ABSTRACT

High-resolution tandem mass spectrometry (HRMS2) with electrospray ionization is frequently applied to study polar organic molecules such as micropollutants. Fragmentation provides structural information to confirm structures of known compounds or propose structures of unknown compounds. Similarity of HRMS2 spectra between structurally-related compounds has been suggested to facilitate identification of unknown compounds. To test this hypothesis, the similarity of reference standard HRMS2 spectra was calculated for 243 pairs of micropollutants and their structurally-related transformation products (TPs); for comparison, spectral similarity was also calculated for 219 pairs of unrelated compounds. Spectra were measured on Orbitrap and QTOF mass spectrometers and similarity was calculated with the dot product. The influence of different factors on spectral similarity (e.g., normalized collision energy (NCE), merging fragments from all NCEs, and shifting fragments by the mass difference of the pair) was considered. Spectral similarity increased at higher NCEs and highest similarity scores for related pairs were obtained with merged spectra including measured fragments and shifted fragments. Removal of the monoisotopic peak was critical to reduce false positives. Using a spectral similarity score threshold of 0.52, 40% of related pairs and 0% of unrelated pairs were above this value. Structural similarity was estimated with the Tanimoto coefficient and pairs with higher structural similarity generally had higher spectral similarity. Pairs where one or both compounds contained heteroatoms such as sulfur often resulted in dissimilar spectra. This work demonstrates that HRMS2 spectral similarity may indicate structural similarity and that spectral similarity can be used in the future to screen complex samples for related compounds such as micropollutants and TPs, assisting in the prioritization of non-target compounds.

Keywords: high-resolution tandem mass spectrometry; micropollutants; transformation products; non-target screening; spectral similarity

Abbreviations: AUC, area under the curve; EI, electron impact; ESI, electrospray ionization; FCS, flexible common substructure; FPR, false positive rate; GNPS, Global Natural Products Social Molecular Networking; HCD, higher-energy collision dissociation; HRMS2, high-resolution tandem mass spectrometry; LC, liquid chromatography; NCE, normalized collision energy; PR, precision-recall; QTOF, quadrupole time-of-flight; ROC, receiver operating characteristic; TP, transformation product; TPR, true positive rate.

Graphical Abstract:



1.0 Introduction

High-resolution tandem mass spectrometry (HRMS2) with electrospray ionization (ESI) has become vital in the identification of known and unknown compounds in fields as diverse as pharmacokinetics, human health studies, metabolomics, natural product research, food, and environmental analysis. HRMS2 has become more common for target screening of known compounds since detection limits have been decreasing in recent years. But the unique advantage of HRMS2 is best observed in non-target or untargeted screening methods which aim to identify compounds in the sample not previously known to the investigator. In this case, accurate mass measurements and resolution of isotope peaks make it possible to assign molecular formulas to unknown peaks, while fragmentation of the precursor ion provides information about the presence or absence of chemical functional groups or substructures, making structure elucidation possible.

When investigating the spectra of an unknown in non-target screening, a reasonable first step is to compare the experimental spectra to those of reference standards that are present in databases and spectral libraries. This search, often referred to as “dereplication” or identifying “known unknowns”, determines if the unknown spectrum belongs to a known compound. Confirmation of matches between the experimental spectrum and library spectra is regularly evaluated with a similarity or match score [1-3], which is based on matching of aligned peaks, and several algorithms are currently available to calculate similarity scores (*e.g.*, the dot product [4], Jaccard index [5], and X rank [6]). But whereas large libraries, such as NIST, exist for low-resolution, electron impact (EI) MS spectra, library resources are more limited for ESI-HRMS2 spectra, for a variety of reasons. The technique is newer and measurements are less standardized leading to varying fragmentation. Therefore library searches with HRMS2 data are less successful in identifying known compounds. Additionally, reference standards are rarely available for some compounds, *e.g.*, transformation products (TPs), which are formed from parent compounds through a multitude of reaction pathways, including metabolism, photolysis or hydrolysis in the environment, or biotransformation or ozonation during wastewater or drinking water treatment. Therefore, HRMS2 spectra for these compounds are also seldom present in spectral libraries.

Since spectra for many compounds may not be in libraries, other methods have been proposed to use HRMS2 spectra to identify unknown compounds, preferably in an automated fashion. One of these strategies is screening for characteristic fragments, thereby at least assigning the unknown compound to a particular class of structurally related compounds. Different resources ~~software~~ (*e.g.*, mzCloud (mzcloud.org), FT-BLAST [7], METLIN [8], MS2Analyzer [9], and CSI:FingerID [10]) have demonstrated the overall success of using fragments to assign chemical substructures. This approach has also been applied to identify TPs, where fragments characteristic of a parent compound have been used to screen for possible TPs [11, 12].

While the relationship between structural and spectral similarity has been previously explored for EI-MS data [13], it is not clear to what extent these results would be the same for ESI-HRMS2 data, and what similarity score corresponds to “similar” spectra, since it cannot be assumed that criteria previously established for EI-MS data also apply to ESI-MS2. Preliminary work, reported in [14], showed spectral similarity between parent compounds and TPs might not be as high as hypothesized. To address this open question, we investigated more than 10,000 HRMS2 spectra from reference standards of polar organic micropollutants, such as pharmaceuticals and pesticides, and associated TPs with various functional groups. The spectral similarity was calculated with the dot product between 243 pairs of parent micropollutants and known TPs. For comparison, similarity scores between 219 unrelated pairs were also calculated. Multiple scenarios were considered when comparing spectra, such as measuring at different collision energies and merging of different spectra, to determine the conditions resulting in the maximum spectral similarity score for each pair. Once similarity scores were maximized, a similarity score threshold was determined that could distinguish related from unrelated pairs. Finally, spectral similarity of each pair was compared to the corresponding structural similarity. The resulting best strategy and thresholds can be applied for future screening of related unknown compounds such as TPs.

2.0 Methods

2.1 Measurement and Data Analysis

Reference standards of 777 compounds were measured in-house with liquid chromatography (LC) – HRMS2 for entry into spectral libraries. The reference standards included a highly diverse group of micropollutants, such as pharmaceuticals, pesticides, artificial sweeteners, industrial chemicals, with various functional groups and heteroatoms, and TPs resulting from a variety of transformation processes, including human metabolism and microbial degradation, as well as from drinking water treatment processes such as ozonation. Seventy compounds were previously reported in Stravs *et al.* [15] along with the details of the measurement conditions, although here three Orbitrap instruments (Thermo Fisher Scientific, San Jose, USA) were used (*i.e.*, Orbitrap XL, Q-Exactive, and Q-Exactive Plus), depending on availability. For 370 compounds, HRMS2 measurement was done on an Orbitrap XL. For 196 compounds, a Q-Exactive was used and a Q-Exactive Plus was used for 224 compounds (13 compounds were measured on multiple instruments). No large differences were observed in fragmentation between the different instruments (Supplementary Material; Figure S1a-m). Ionization was done with either positive or negative ESI (or both). All fragmentation was performed with HCD at set energies (*i.e.*, 15, 30, 45, 60, 75, 90), reported as normalized collision energies (NCEs), using the minimum resolution for the MS2 (7,500 for Orbitrap and 17,500 for Q-Exactive/Q-Exactive Plus) and an isolation window of 1 *m/z*, such that no isotope peaks were present in the spectra.

The initial dataset was comprised of reference standard spectra processed using the R package *RMassBank* [15] and made available online at MassBank (www.massbank.eu) [16]. RMassBank retrieves spectra from raw files (mzML or mzXML) based on SMILES and retention time. The RMassBank workflow then starts with a recalibration of the fragment masses, where first, a mass recalibration is performed using mass errors of subformulas assigned to fragment masses for a set of known compounds, and second, using the recalibrated spectra, subformula assignment is performed again to remove noise peaks that do not match a chemical formula consistent with the parent formula. Further processing steps include 1) the removal of probable Fourier transform satellite peaks and (if activated) of known electronic noise peaks from the instrument, 2) reassignment of potential collision gas adducts, 3) filtering by multiplicity (occurrence in multiple spectra or repeated measurements), and finally 4) an export of intense peaks marked as noise for manual review (further details of the settings are in [13] and in the vignette in BioConductor (http://bioconductor.org/packages/release/bioc/vignettes/RMassBank/inst/doc/RMassBank.pdf)). The processed spectra are then annotated with metadata and exported into MassBank record format, or alternatively (*e.g.*, for this work) exported in tabular format for further processing. The basic settings (tailored in this case to the Orbitrap spectra and the chromatography) were as follows: RT margin = 0.4 min; include reanalyzed peaks (accounting for N_2 and O adducts, see [13]); add annotation; multiplicity filter = 2; recalibrate by ppm; MS1 and MS2 recalibration using the *loess* function; initial recalibration window 15, 10, and 15 ppm for MS1, MS2 $m/z > 120$, and MS2 $m/z < 120$, respectively; final recalibration window 5 ppm; intensity limit 10,000 (spectra are not extracted if the maximum MS2 intensity is below this level). As the reference standard spectra available in-house (cleaned records) were the starting point for this study, “uncleaned” spectra, which included all peaks, were subsequently extracted from the RMassBank archives to assess this approach on spectra more similar to routine data analysis. In total, 9413 spectra were processed, encompassing 289,615 fragments. A subset of compounds were measured on a QTOFMS instrument, details of which are in the Supplementary Material (Section S2) and in Gago-Ferrero *et al* [17].

All data processing was done in R [18] (v.3.2.1) using various packages as indicated below. Of the 777 reference samples measured, 243 related pairs of parent and TP were selected based on previous knowledge of possible transformations; additionally, 219 unrelated pairs were randomly generated. The transformations between the pairs consisted mainly of minor modifications resulting from environmentally relevant reactions. A small number of larger transformations (such as conjugation reactions) were included although these reactions are expected to be of less significance in the environment and only a few reference standards for these TPs were available. Sixty-seven parent compounds were associated with multiple TPs, while 53 TPs were paired to multiple parents. Full list of the pairs is available in the Supplementary Material, Table S1.

2.2 Spectral Similarity Calculations

Spectrum similarity was based on the distance between the aligned HRMS2 spectra as calculated by the cosine of the angle between them. It is referred to as the modified cosine or dot product, is often employed in database spectral search algorithms [16, 19, 20], and was used for a similar evaluation with low-resolution EI-MS data [13]. Calculations were done with an internal R script (https://github.com/dutchjes/MSMSsim) and were based on functions in the R package *OrgMassSpecR* [21]. Only the forward match score was considered in this analysis. In order to calculate similarity, *m/*z fragments are aligned and the intensities are compared. An *m/z* tolerance factor is applied to align fragments; 0.005 Da was used for Orbitrap data and 0.015 Da for QTOF data due to a higher mass error. A relative intensity cutoff of 0.5 was used to eliminate peaks of low intensity and fragments with no match were paired with an intensity of zero. The similarity score ranges from 0 to 1, with 1 being a perfect match and is calculated as

$$r = \frac{x_A \cdot x_B}{\sqrt{(x_A \cdot x_A)}\sqrt{(x_B \cdot x_B)}} \quad (1)$$

with x_A and x_B the aligned intensity vectors of compound A and compound B, respectively.

Rather than using only intensities, comparison of spectra can also be done using weighted vectors, where both mass and intensity are considered, using the formula

$$x_i = m^c I^d \quad (2)$$

where *m* is the mass and *I* is the intensity and *c* and *d* are weighting factors to optimize the dot product algorithm. For example, the NIST search algorithm uses c=3, d=0.6; MassBank uses c=2, d=0.5; and Demuth *et al*. found that c=0, d=0.33 produced the best results for correlating structural similarity to spectral similarity. For this work, these three weighting factors plus c=0 and d=1 were tested. Two examples of HRMS2 spectra comparison with very different similarity scores are shown in Fig. **1**.

2.2.1 Scenario 1: Single collision energy spectra

Measurements at six different NCEs were used to study changing fragmentation profiles and determine if there was an optimum NCE for comparison. To the extent possible, the measurements that were compared were collected at the same resolution and on the same instrument. Only measurements collected in the same ionization mode were compared. R package *lattice* [22] (v.0.20-

33) was used for box-whisker plots. Density distributions were generated with the R package *sm* [23] (v.2.2-5.4).

2.2.2 Scenario 2: Merged spectra

‘Merged’ spectra were produced by merging fragments from all collision energies measured using an internal R script (https://github.com/dutchjes/MSMSsim). The *m/z* tolerance for merging fragments was 0.001 Da and the fragment intensity in the merged spectra corresponded to the maximum intensity of the fragment across the collision energies, using either absolute intensities or relative intensities (both possibilities were considered).

2.2.3 Scenario 3: Shifted spectra

In addition to the measured (‘unshifted’) spectra, ‘shifted’ spectra were generated for each TP to understand if including the mass difference of the transformation resulted in higher spectral similarity; shifted spectra have previously been described for comparing spectra of different compounds [7, 24]. Unshifted spectra were simply the measured fragments of the TP. Shifted spectra were produced by shifting all fragments of the TP by the mass difference between the parent and TP. For example, for a pair where a demethylation occurred, all fragment masses of the TP were increased by 14.0157 Da, the mass of a methyl group minus one hydrogen. This shift was done to capture those cases where a TP fragmented at the same location in the molecule as the parent compound, but where the fragment masses do not match because the transformation occurred on this fragment. Spectral similarity to the parent compounds were then calculated for both the unshifted and shifted spectra. During this analysis the precursor ions of both the parent and TP were removed from the spectra, to remove the trivial match resulting from the TP shift and subsequent match of the parent precursor to the TP precursor, which lead to artificially high similarity scores (data in Supplementary Material, Section S8). Shifted spectra are denoted with the annotation ‘wMD’ (with mass difference). Additionally, ‘combined’ spectra, which included both shifted and unshifted fragments, were also analyzed.

2.3 Similarity Score Threshold Determination

After calculating the similarity scores of all the scenarios detailed above, stacked bar plots were used to visualize how the rates of false positives, false negatives, true positives, and true negatives changed at different similarity score thresholds. True positives were the number of related pairs with a spectral similarity score above the threshold, and false negatives the number of related pairs below the threshold; true negatives were the number of unrelated pairs with similarity scores below the threshold, while false positives the number of unrelated pairs above the threshold. Furthermore, the different scenarios were visually compared with the following two methods: (1) receiver operating characteristic (ROCs) curves, that visualize the rate of false positives (FPR) on the x-axis vs. the rate of true positives (TPR) on the y-axis and (2) precision-recall (PR) curves, where recall is plotted vs. precision (defined below). The FPR and TPR reflect the percent of unrelated pairs and related pairs that are above a given similarity score threshold, respectively, and are calculated as follows:

$$false\ positive\ rate\ (FPR) = \frac{\#\ of\ false\ positives}{\#\ of\ false\ positives + \#\ of\ true\ negatives} \qquad (3)$$

$$true\ positive\ rate\ (TPR) = \frac{\#\ of\ true\ positives}{\#\ of\ true\ positives + \#\ of\ false\ negatives} \qquad (4)$$

where the denominator in eq. 3 is equal to the total number of unrelated pairs, and the denominator in eq. 4 is equal to the total number of related pairs. Calculating precision and recall was done as follows:

$$precision = \frac{\#\ of\ true\ positives}{\#\ of\ true\ positives + \#\ of\ false\ positives} \qquad (5)$$

$$recall = \frac{\#\ of\ true\ positives}{\#\ of\ true\ positives + \#\ of\ false\ negatives} \qquad (6)$$

(note that recall is the same at TPR). In the ROC graphs, an ideal situation would be plotted in the top-left, with FPR equal to 0 and TPR equal to 1, while in PR curves the ideal case could be plotted in the top-right, with recall equal to 1 and precision equal to 1. Quantitatively the curves were compared by calculating the area under the curve (AUC) statistic. ROC curves, PR curves, ROC-AUCs, and PR-AUCs were calculated with the R package *PRROC* (v1.3).~~It~~ Additionally, it has been shown that the ROC-AUC statistic may include some bias and that the H-measure is a more reliable way to compare ROCs [25], therefore ROC-AUCs and H-measures were also calculated with the R package *hmeasure* [26] (v.1.0); however, for this data the results were found to be similar and available only in the Supplementary Material (Table S6). The scenario with the highest ROC-AUC and PR-AUC values was selected to be the best, as it was most successful in distinguishing related from unrelated pairs. Finally, the similarity score corresponding to a FPR of 0 was designated as an optimum threshold value. Bootstrapping (R=1000) was done with the R package *boot* [27] to determine the mean, standard deviation, and 95% confidence interval of the optimum similarity score threshold.

2.4 Spectral Similarity vs. Structural Similarity

Finally, to measure the structural similarity of each pair, JChem for Office [28] (15.7.2700.2799) was used to first retrieve SMILES codes from CAS numbers [29]. For a handful of compounds (namely TPs) without a CAS number, the structure of the compound was manually drawn in MarvinSketch [28] (v.15.8.3) and output as a SMILES code. MOL files were generated from the SMILES codes with the R package *RMassBank* [15] (v.1.10.0), SDF files were generated with the R package *ChemmineR* [30] (v.2.20.3), and structures were visualized with the flexible common substructure (FCS) algorithm available in the R package *fmcsR* [31] (v.1.10.3) to compare differences in functional groups between parent and TP. Three algorithms were considered for estimating structural similarity. First, TanimotoDissimilarity was calculated by JChem with the function JCDissimilarityCFTanimoto, and similarity was reported as 1 – TanimotoDissimilarity with values reported from 0 to 1, 1 being a perfect match. This algorithm uses substructure-based fingerprints to compare structures and the dissimilarity between these fingerprints is calculated with the Tanimoto distance. Second, *cmp.similarity* function from ChemmineR [30] was used, which is defined as the proportion of atom pairs shared between two compounds. Third, the *fmsc* function from fmcsR [31] was used, which is a graph-based similarity function based on the largest overlapping substructure.

3.0 RESULTS AND DISCUSSION

3.1 Fragment analysis

Fragments measured from 777 compounds across six NCEs were characterized and, as expected, smaller fragments were formed at higher NCEs (Supplementary Material, Figure S2). The *m/z* range of all detected fragments at NCE15 was 50–1040, while at NCE90 the *m/z* range was 50–692. Correspondingly, the number of fragments detected per compound increased (median 12 fragments per compound at NCE15 to 52 fragments per compound at NCE90) and the detection frequency increased for many fragments at higher NCEs. At NCE15 the most common fragment (*m/z* 91.0542) was detected 121 times (in 16% of spectra), while at NCE90, the most common fragment (*m/z*

65.0386) was detected in 74% of spectra. Fragments were annotated with formulas and the most common fragments are shown in Table S2. While the number of detections increased with NCE, the most frequently detected fragment formulas generally (and surprisingly) remained the same. It is postulated that these frequently detected fragments correspond to common substructures, especially since many micropollutants contain similar functional groups. For example, *m/z* 91.0542 ($C_7H_7^+$) and *m/z* 65.0386 ($C_5H_5^+$) both are formed during the fragmentation of aromatic compounds.

The fragment $C_6H_5N_2^+$ became increasingly common at the higher collision energies, while the fragment $C_3H_6N^+$ had decreasing rank at higher collision energies, even though the overall number of detections still increased. A recent publication by Böcker and Dührkop examined frequency of detection of fragment formulas in Agilent QTOF data and also regularly detected the fragments $C_7H_7^+$ and $C_3H_6N^+$, although $C_6H_5N_2^+$ was not reported [32]. The $C_6H_5N_2^+$ fragment is a nitrogen adduct associated mostly with NCE75 and above [15]. As Böcker and Dührkop considered only fragments that were a subformula of the parent, their method could not annotate this fragment but they did find occurrences of this peak in their unprocessed spectra (Böcker and Dührkop, pers. comm.), primarily in the 40 eV spectra.

3.2 Pairs characterization

From the 777 compounds with reference spectra, 243 related pairs were established; 198 measured in positive ESI mode and 45 in negative ESI mode. Within these pairs, 47 different transformation types were found and some parents or TPs were associated with multiple pairs. In general, TPs were more polar and smaller than their parent compounds. LogKow values corrected for pH (logDow at pH7) of the TPs were between -4.2 and 5.6 (median 0.7), while for the parents logDow ranged from -3.7 to 9.6 (median 1.9). Masses ranged from 86.03 to 764.50 Da (median 234.66 Da) for the TPs and from 70.04 to 990.98 Da (median 270.13 Da) for the parent compounds. The median absolute mass difference between the pairs was 28.03 Da and ranged from 0.04 Da (loss of CH_4, addition of O) to 446.0 Da (loss of a long fluorinated alkyl chain). For the QTOFMS analysis a smaller set of 73 pairs were analyzed.

3.3 Similarity Score Calculations

Different scenarios were considered to calculate similarity scores between parent compound and TP. The results of each scenario are presented in the following subsections, followed by an overall comparison of the different scenarios and the selection of the best scenario based on the ROC-AUCs and PR-AUCs. Although different weighing factors were considered for the similarity score calculations, the scenario resulting in the highest ROC-AUC and highest PR-AUC was the same with each of the weighting factors; therefore only the similarity score results using c=0 and d=1 are presented. A summary of the results from the other weighing factors is provided in the Supplementary Material, Section S5.

3.3.1 Scenario 1: Single collision energy spectra

First, the influence of collision energy of the similarity scores of pairs was investigated. It was of concern that the same fragments could be generated even from two structurally unrelated molecules, since quite a few fragments (especially smaller fragments) were frequently detected. High similarity scores (*i.e.*, scores close to 1) in the unrelated pairs could therefore indicate that the fragments were not very structure specific.

As shown in Figure S5a, spectral similarity of the unrelated pairs was very low at all NCEs. Even at NCE90, where the highest number of small fragments are expected to be formed, spectral similarity was very low (median 0), demonstrating that the spectra containing smaller fragments did not lead to high similarity scores. In the related pairs (Figure S5b), highest spectral similarity between parent and TP was observed at NCE90 (median similarity score 0.4) and pairs were less similar at lower NCEs. This result may simply be a result of having more fragments to match. For example, at NCE15 an average of 2.5 fragments matched per related pair, while at NCE90 an average of 18.5 fragments matched (Table S5).

3.3.2 Scenario 2: Merged spectra

The second scenario concerned merged spectra from all collision energies measured. Note that the fragments with the highest absolute intensities are generally larger fragments measured at lower NCEs (Figure S6), which would result in these fragments having a high influence on the similarity scores when spectra are merged using absolute intensities (Figure S7). Therefore, merged spectra using either the absolute intensity or relative intensity were evaluated separately.

The similarity scores of the related pairs using the relative intensities were overall substantially higher as compared to scores calculated using the absolute intensities (median 0.25 and 0.04, respectively; Figure S8), suggesting again that the smaller fragments formed at higher NCEs were critical in obtaining higher similarity scores. These small fragments still appeared to be structure specific, since in the similarity scores of the unrelated pairs were overall close to zero (median 0) for both the relative and absolute intensities.

3.3.3 Scenario 3: Shifted spectra

It was hypothesized that if TP fragment masses were adjusted for the transformation which had occurred, fragments would be aligned that were altered during the transformation. A similar idea has been used in molecular networking of metabolites [24] and has been implemented in GNPS [33]. During the course of this analysis, it became apparent that the monoisotopic precursor peak had a large influence on the spectral similarity, since this peak was, in many cases, the most intense peak in the spectrum. By shifting all fragments, the monoisotopic peaks artificially matched purely as a result of the mass difference shift (which was calculated as the difference of the monoisotopic masses), resulting in an increase in similarity scores of unrelated pairs. This increase was especially evident at low NCEs, where the monoisotopic peak dominated the HRMS2. When the precursor peak was removed, similarity scores of unrelated pairs decreased (further information in the Supplementary Material, Section S8). Therefore, the precursor peak was removed from the shifted spectra.

The similarity of the shifted spectra from the different collision energies was evaluated. Interestingly, the results had the opposite trend as the unshifted spectra (Section 3.3.1). The similarity of the shifted spectra decreased with increasing NCEs (Fig. **2**), indicating that shifting fragments was most beneficial when larger fragments were present (*i.e.*, those produced at the lower NCEs). A likely explanation is that shifting fragments is not very useful at higher NCEs, since many small fragments are produced at higher NCEs and only a few of those fragments are on locations of the molecule affected by the transformation. Furthermore, when the similarity scores at the single NCEs were compared between shifted and unshifted, even the highest similarity scores that were obtained with the shifted spectra (at NCE15; median 0.07) were much lower than those calculated for the unshifted spectra (highest scores at NCE90; median 0.43) (Table 1 and Fig. **2**). Therefore, adjusting all fragment masses to account for the change that is likely present on only one or two fragments has a detrimental effect on the spectral similarity scores, since it meant that previously matching fragments which did not contain the modification no longer matched.

3.4 Scenario Comparison and Similarity Score Threshold Determination

As shown in Section 3.3.2, using the relative intensity for merging resulted in higher similarity scores, either because more weight is given to the smaller, less intense fragments formed at higher collision energies or simply because more fragments are present. From the single collision energy analysis (Section 3.3.1), it was determined that these smaller fragments are useful for calculating spectral similarity. These results nicely substantiate each other and are further confirmed with the ROC curves and PR curves (Fig. 3) and the AUC values obtained (Table 1). From all scenarios analyzed (*i.e.*, single collision energies, merged spectra, and shifted spectra), the two combined merged spectra scenarios, with both shifted and unshifted TP fragments, had the highest ROC-AUCs (0.92; Table 1), indicating these scenarios were most successful at distinguishing between related and unrelated pairs. From these two, the highest PR-AUC and the higher true positive rate (TPR) was achieved with the combined merged spectra using relative intensities (PR-AUC = 0.94; 40% TPR at a false positive rate (FPR) of 0%; Fig. 4). But other scenarios, namely the unshifted NCE90 and unshifted relative merged spectra, actually had higher percentage of true positives captured (48% and 46%, respectively, at FPR of 0%). Therefore, related and unrelated pairs could also be separated simply by measuring at high collision energies or merging fragments from multiple collision energies, without needing to remove the monoisotopic peak and/or shift fragments.

Using the scenario with the highest ROC-AUC, PR-AUC, and lowest TPR (*i.e.*, the relative combined spectra), a similarity score threshold was selected that distinguished between the related pairs and the unrelated pairs. There are many different ways to select such a threshold value [34], but in the context of applying the similarity score threshold to screen for unknown TPs, it was decided that minimizing the false positives was most important, and therefore a FPR of 0% was desirable. In this way, in the future when screening unknown spectra, there would be more confidence that a pair with a similarity score above the given similarity score threshold is truly related; simultaneously there is a higher likelihood that related pairs may be missed. The similarity score threshold above which all unrelated pairs were discarded was determined to be 0.52 (95% confidence interval 0.41 – 0.78; Table 1).

3.5 Comparison to QTOF spectra

Overall QTOF data corroborated the Orbitrap results. Higher spectral similarity between related pairs was observed at higher collision energies (Figure S16a) and the best results were obtained with the relative merged data (Figure S17). Using the mass difference of the transformation to shift the fragment masses was not beneficial (Figure S16b; also here the monoisotopic peaks were removed prior to comparison of shifted spectra). These results indicate that the conclusions shown here for the Orbitrap data should be relevant also for HRMS2 spectra collected on QTOF instruments.

3.6 Spectral Similarity vs. Structural Similarity

Finally, it was tested if structural similarity of a pair was related to the spectral similarity of the HRMS2. The scenario with the highest AUCs (as described in Section 3.4), the relative combined merged spectra that included unshifted and shifted TP fragments, was used to calculate spectral similarity. The structural similarity between a pair was estimated using the Tanimoto coefficient and ranged from 0.06 to 1.0 for related pairs (Figure S18). To visualize how transformation type may influence fragmentation, two example pairs are shown in Fig. **1**. Atrazine is the parent molecule in both cases, with one TP the result of a substitution of a chlorine with a hydroxy group and the second a dealkylation reaction. In both pairs the Tanimoto coefficients were relatively high (0.55 for the hydroxyl TP, 0.97 for the desethyl TP), but the spectral similarity scores were very different for these two pairs (0.0 and 0.54, respectively). The substitution of the chlorine with a hydroxyl meant that most fragments no longer matched. In comparison, the ethyl group of the parent compound was one of the first functional groups cleaved, therefore the remaining fragments matched in many cases to the fragments of the desethyl-TP. More generally, it is clear from Fig. **5**a that pairs with low structural similarity were unlikely to produce similar spectra. However, the inverse statement, that two structurally similar compounds will produce similar spectra, is much more difficult to conclude. In general, increasing spectral similarity was observed with increasing structural similarity (Fig. 5a). Two other algorithms for estimating structural similarity were also considered, but the strongest relationship between structural similarity and spectral similarity was observed with the Tanimoto coefficient (Supplementary Material, Section S9 and Figure S19).

Some special cases were observed; 28 pairs were found to have high structural similarity (Tanimoto score >0.8) and low spectral similarity (dot product <0.4). For 53% of these pairs, either the parent or the TP (or both) was a sulfur-containing compound and in most cases the sulfur moiety was directly affected by the transformation (Table S10). Heteroatoms such as sulfur can have a large influence on the fragmentation behavior of molecule [35], resulting in dissimilar spectra. These results show that in some cases chemical characteristics which have a large influence on the fragmentation of a molecule are not always adequately captured by the structural similarity measure used here. Nevertheless, a thorough evaluation of structural similarity coefficients by Salim *et al*. found that the Tanimoto coefficient was an adequate single measurement of the chemical similarity as more complicated algorithms did not improve upon this greatly [36]. The similarity scores for different transformation types were analyzed to determine if certain parent/TP pairs had overall higher (or lower) spectral similarity but no firm conclusions could be drawn (Figure S20).

3.7 Uncleaned spectra

Uncleaned spectra were also analyzed to simulate real-world data. The same pairs were used but noise and unannotated peaks (removed by RMassBank during processing of the spectra used above) were retained. The similarity scores were calculated with the relative combined merged spectra with both unshifted and shifted fragments which had produced the best results in the cleaned spectra. It was observed that a lower similarity score threshold (0.29) could be used to achieve a FPR of 0%, likely because the overall distribution of similarity scores was lower. Interestingly, at this threshold the uncleaned spectra had a higher TPR compared to the cleaned spectra (69%). This result is surprising but very positive, since it indicates that the presence of noise peaks in the spectra did not lead to any reduction in the ability of the similarity score to discriminate between the related pairs and unrelated pairs. Additionally, when considering the relationship between the structural similarity and spectral similarity of the uncleaned spectra, the results were the same as with the cleaned spectra (Fig. **5**). It is clear that dissimilar pairs will not produce similar spectra and that increasing structural similarity did overall indicate increasing spectral similarity.

4.0 CONCLUSIONS

A detailed analysis of HRMS2 reference spectra of parent/TP pairs provided insight into how different measurement and data analysis parameters can influence spectral similarity and demonstrated that structural similarity is related to spectral similarity. Using optimized settings, 40% of the related pairs (and none of the unrelated pairs) were above the spectral similarity score threshold of 0.52. In uncleaned spectra, the similarity score threshold was lower (0.29) due to the presence of noise peaks; however, the percentage of related pairs above this threshold was substantially higher (69%). Although the 95% confidence interval for the similarity score threshold was quite large (0.41-0.78), it provides a starting point to determine if spectra are from structurally similar compounds. It should be noted that in the real world situation, many more unrelated pairs exist than related pairs; therefore higher rates of false positives can be expected, and the correct similarity score threshold applicable under these conditions would need to be further evaluated in future work. Nevertheless, these results demonstrate that pairs of related parent micropollutants and the corresponding TPs could be selected over unrelated pairs of compounds using the similarity of HRMS2 spectra, representing a step forward in the prioritization of potentially relevant non-target peaks amongst the tens of thousands of unknown peaks that remain unidentified in typical environmental investigations [37, 38]. Furthermore, as the link to the parent can be established, identification efforts can be performed on the substance most likely to be known, *i.e.*, the parent compound.

The similarity score threshold needed here to distinguish between related and unrelated pairs is lower than values recommended in other situations (*e.g.*, matching measured spectral with a database entry or matching predicted spectra with measured spectra). For example, in molecular networking, which builds nodes of similar MS2 spectra for the purposes of clustering structurally similar compounds, a similarity score threshold of 0.7 is recommended to build the nodes'[24, 39, 40]. This difference may partially be explained by the fact that natural products are in general larger than micropollutants, therefore more fragments are generated per compound. As was demonstrated here, the best results were obtained with those spectra containing the most fragments. Furthermore, it should be noted that results from positive and negative ionization modes were presented together due to a lack of negative ionization pairs for separate analysis. The similarity score thresholds needed to discriminate between related and unrelated pairs in the two ionization modes may be quite different and could be further explored. Particularly in the case of TPs, the dataset used here is one of the largest publically available for these types of compounds, but the conclusions of this work can be refined as new reference spectra become available for comparison. Additionally, in the single NCE comparison, spectral similarity scores were calculated only between spectra collected at the same NCEs. It would be interesting in the future to expand the comparison, such that the spectra collected at all energies are compared for each pair, to find the best matching spectra. Other algorithms for calculating merged spectra, *e.g.*, using the sum of raw intensities rather than the maximum intensity of each fragment, could also be considered. It should be stressed that the spectral similarity scores presented here are not intended for comparing unknown spectra to library spectra but rather for comparing two unknown spectra. The goal is that, after previous prioritization steps such as linkages through metabolic logic as conducted in our recent study [14], these similarity scores thresholds will be useful in selecting compounds that might be structurally related and therefore assisting in further structure elucidation.

The observed relationship between structural similarity and spectral similarity was in good agreement with a similar comparison conducted with low-resolution EI-MS data. It is perhaps surprising that the correlation observed is so similar, since one might expect that the accurate mass information provided by HRMS2 would be more specific. As detailed in the introduction, many groups have used spectral similarity to find structurally related compounds such as metabolites or TPs of known parent compounds. The work presented here indicates that some of the strategies proposed for metabolite discovery (*e.g.*, using a single diagnostic fragments from parent compounds to search for TPs) may still be overlooking TPs which do not produce these characteristic fragments. This work provides a way forward for incorporating information from the entire HRMS2 spectra when searching for structurally related compounds such as unknown TPs.

Appendix A. Supplementary Material

Supplementary material is available as indicated in the text.

Acknowledgements

Birgit Beck, Heinz Singer, and many members of the department of Environmental Chemistry at Eawag are gratefully acknowledged for the measurement of the standards for MassBank. We thank additionally Nikiforos Alygizakis from the University of Athens for the measurement of the QTOFMS spectra. Uwe Schmitt (ETH Zurich) and Leon Bichmann (Eawag), Sebastian Böcker and Kai Dührkop (University of Jena), and Oscar Yanes (Center for Omic Sciences, Spain) are thanked for helpful discussions. Funding for JES was provided by the EDA-Emerge project through the EU Seventh Framework Programme (FP7-PEOPLE-2011-ITN) under grant agreement number 290100 and from the Swiss Federal Office for the Environment. ELS was supported by the SOLUTIONS project (EU FP7, grant number 603437). Funding for MS and RG was provided by the Swiss National Science Foundation.

References

1. Wishart, D., Tzur, D., Knox, C., Eisner, R., Guo, A., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., Fung, C., Nikolai, L., Lewis, M., Coutouly, M., Forsythe, I., Tang, P., Shrivastava, S., Jeroncic, K., Stothard, P., Amegbey, G., Block, D., Hau, D., Wagner, J., Miniaci, J., Clements, M., Gebremedhin, M., Guo, N., Zhang, Y., Duggan, G., MacInnis, G.: HMDB: The human metabolome database. Nucleic Acids Res. **35**, D521--526 (2007)
2. Neumann, S., Böcker, S.: Computational mass spectrometry for metabolomics: Identification of metabolites and small molecules. Anal. Bioanal. Chem. **398**, 2779-2788 (2010)
3. Stein, S.: Mass spectral reference libraries: An ever-expanding resource for chemical identification. Anal. Chem. **84**, 7274 - 7282 (2012)
4. Stein, S.E., Scott, D.R.: Optimization and testing of mass spectral library search algorithms for compound identification. J. Am. Soc. Mass. Spectrom. **5**, 859-866 (1994)
5. Allen, F., Greiner, R., Wishart, D.: Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. Metabolomics. **11**, 98-110 (2015)
6. Mylonas, R., Mauron, Y., Masselot, A., Binz, P., Budin, N., Fathi, M., Viette, V., Hochstrasser, D., Lisacek, F.: X-rank: A robust algorithm for small molecule identification using tandem mass spectrometry. Anal. Chem. **81**, 7604 - 7610 (2009)
7. Rasche, F., Scheubert, K., Hufsky, F., Zichner, T., Kai, M., Svatos, A., Bocker, S.: Identifying the unknowns by aligning fragmentation trees. Anal. Chem. **84**, 3417 - 3426 (2012)
8. Smith, C., O'Maille, G., Want, E., Qin, C., Trauger, S., Brandon, T., Custodio, D., Abagyan, R., Siuzdak, G.: METLIN: A metabolite mass spectral database. Ther Drug Monit. **27**, 747 - 751 (2005)
9. Ma, Y., Kind, T., Yang, D., Leon, C., Fiehn, O.: MS2Analyzer: A Software for Small Molecule Substructure Annotations from Accurate Tandem Mass Spectra. Anal. Chem. **86**, 10724-10731 (2014)
10. Dührkop, K., Shen, H., Meusel, M., Rousu, J., Böcker, S.: Searching molecular structure databases with tandem mass spectra using CSI:FingerID. Proceedings of the National Academy of Sciences. **112**, 12580-12585 (2015)
11. Kern, S., Fenner, K., Singer, H.P., Schwarzenbach, R.P., Hollender, J.: Identification of Transformation Products of Organic Contaminants in Natural Waters by Computer-Aided Prediction and High-Resolution Mass Spectrometry. Environ. Sci. Technol. **43**, 7039-7046 (2009)
12. Majewsky, M., Glauner, T., Horn, H.: Systematic suspect screening and identification of sulfonamide antibiotic transformation products in the aquatic environment. Anal. Bioanal. Chem., 1-11 (2015)
13. Demuth, W., Karlovits, M., Varmuza, K.: Spectral similarity versus structural similarity: mass spectrometry. Anal. Chim. Acta. **516**, 75-85 (2004)
14. Schollée, J.E., Schymanski, E.L., Avak, S.E., Loos, M., Hollender, J.: Prioritizing Unknown Transformation Products from Biologically-Treated Wastewater Using High-Resolution Mass Spectrometry, Multivariate Statistics, and Metabolic Logic. Anal. Chem. **87**, 12121-12129 (2015)
15. Stravs, M.A., Schymanski, E.L., Singer, H.P., Hollender, J.: Automatic recalibration and processing of tandem mass spectra using formula annotation. J. Mass Spectrom. **48**, 89-99 (2013)
16. Horai, H., Arita, M., Kanaya, S., Nihei, Y., Ikeda, T., Suwa, K., Ojima, Y., Tanaka, K., Tanaka, S., Aoshima, K., Oda, Y., Kakazu, Y., Kusano, M., Tohge, T., Matsuda, F., Sawada, Y., Hirai, M.Y., Nakanishi, H., Ikeda, K., Akimoto, N., Maoka, T., Takahashi, H., Ara, T., Sakurai, N., Suzuki, H., Shibata, D., Neumann, S., Iida, T., Tanaka, K., Funatsu, K., Matsuura, F., Soga, T., Taguchi, R., Saito, K., Nishioka, T.: MassBank: a public repository for sharing mass spectral data for life sciences. J. Mass Spectrom. **45**, 703-714 (2010)
17. Gago Ferrero, P., Schymanski, E.L., Bletsou, A.A., Aalizadeh, R., Hollender, J., Thomaidis, N.S.: EXTENDED SUSPECT AND NON-TARGET STRATEGIES TO CHARACTERIZE EMERGING POLAR ORGANIC CONTAMINANTS IN RAW WASTEWATER WITH LC-HRMS/MS. Environ. Sci. Technol. **49**, 12333-12341 (2015)

18. A language and environment for statistical computing. R Foundation for Statistical Computing, 2014. http://www.R-project.org/

19. Stein, S.E.: Chemical substructure identification by mass spectral library searching. J. Am. Soc. Mass. Spectrom. **6**, 644-655 (1995)

20. Huan, T., Tang, C., Li, R., Shi, Y., Lin, G., Li, L.: MyCompoundID MS/MS Search: Metabolite Identification Using a Library of Predicted Fragment-Ion-Spectra of 383,830 Possible Human Metabolites. Anal. Chem. **87**, 10619-10626 (2015)

21. OrgMassSpecR: Organic Mass Spectrometry. R package version 0.4-4, 2014. http://CRAN.R-project.org/package=OrgMassSpecR

22. Sarkar, D. Springer, New York (2008)

23. R package 'sm': nonparametric smoothing methods, 2014. http://www.stats.gla.ac.uk/~adrian/sm

24. Watrous, J., Roach, P., Alexandrov, T., Heath, B., Yang, J., Kersten, R., van der Voort, M., Pogliano, K., Gross, H., Raaijmakers, J., Moore, B., Laskin, J., Bandeira, N., Dorrestein, P.: Mass spectral molecular networking of living microbial colonies. Proc Natl Acad Sci USA. **109**, E1743--E1752 (2012)

25. Hand, D.J.: Measuring classifier performance: a coherent alternative to the area under the ROC curve. Machine Learning. **77**, 103-123 (2009)

26. hmeasure: The H-measure and other scalar classification performance metrics., 2012. http://CRAN.R-project.org/package=hmeasure

27. boot: Bootstrap R (S-Plus) Functions, 2015.

28. JChem for Office, 2015. www.chemaxon.com

29. Daylight Chemical Information Systems, Inc., http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html

30. Cao, Y., Charisi, A., Cheng, L.-C., Jiang, T., Girke, T.: ChemmineR: a compound mining framework for R. Bioinformatics. **24**, 1733-1734 (2008)

31. Wang, Y., Backman, T.W.H., Horan, K., Girke, T.: fmcsR: mismatch tolerant maximum common substructure searching in R. Bioinformatics. **29**, 2792-2794 (2013)

32. Böcker, S., Dührkop, K.: Fragmentation trees reloaded. Journal of Cheminformatics. **8**, 1-26 (2016)

33. GNPS: Global Natural Products Social Molecular Networking, 2015. https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp

34. López-Ratón, M., Rodríguez-Álvarez, M.X., Cadarso-Suárez, C., Gude-Sampedro, F.: OptimalCutpoints: An R Package for Selecting Optimal Cutpoints in Diagnostic Tests. 2014. **61**, 36 (2014)

35. Holčapek, M., Jirásko, R., Lísa, M.: Basic rules for the interpretation of atmospheric pressure ionization mass spectra of small molecules. J. Chromatogr. A. **1217**, 3908-3921 (2010)

36. Salim, N., Holliday, J., Willett, P.: Combination of Fingerprint-Based Similarity Coefficients Using Data Fusion. Journal of Chemical Information and Computer Sciences. **43**, 435-442 (2003)

37. Schymanski, E.L., Singer, H.P., Longrée, P., Loos, M., Ruff, M., Stravs, M.A., Ripollés Vidal, C., Hollender, J.: Strategies to Characterize Polar Organic Contamination in Wastewater: Exploring the Capability of High Resolution Mass Spectrometry. Environ. Sci. Technol. **48**, 1811-1818 (2014)

38. Schymanski, E.L., Singer, H.P., Slobodnik, J., Ipolyi, I., Oswald, P., Krauss, M., Schulze, T., Haglund, P., Letzel, T., Grosse, S., Thomaidis, N.S., Bletsou, A., Zwiener, C., Ibáñez, M., Portolés, T., de Boer, R., Reid, M., Onghena, M., Kunkel, U., Schulz, W., Guillon, A., Noyon, N., Leroy, G., Bados, P., Bogialli, S., Stipaničev, D., Rostkowski, P., Hollender, J.: Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. Anal. Bioanal. Chem. **407**, 6237-6255 (2015)

39. Barupal, D.K., Haldiya, P.K., Wohlgemuth, G., Kind, T., Kothari, S.L., Pinkerton, K.E., Fiehn, O.: MetaMapp: mapping and visualizing metabolomic data by integrating information from biochemical pathways and chemical and mass spectral similarity. BMC Bioinformatics. **13**, 1-15 (2012)

40. Allard, P.-M., Péresse, T., Bisson, J., Gindro, K., Marcourt, L., Pham, V.C., Roussi, F., Litaudon, M., Wolfender, J.-L.: Integration of Molecular Networking and In-Silico MS/MS Fragmentation for Natural Products Dereplication. Anal. Chem. **88**, 3317-3323 (2016)

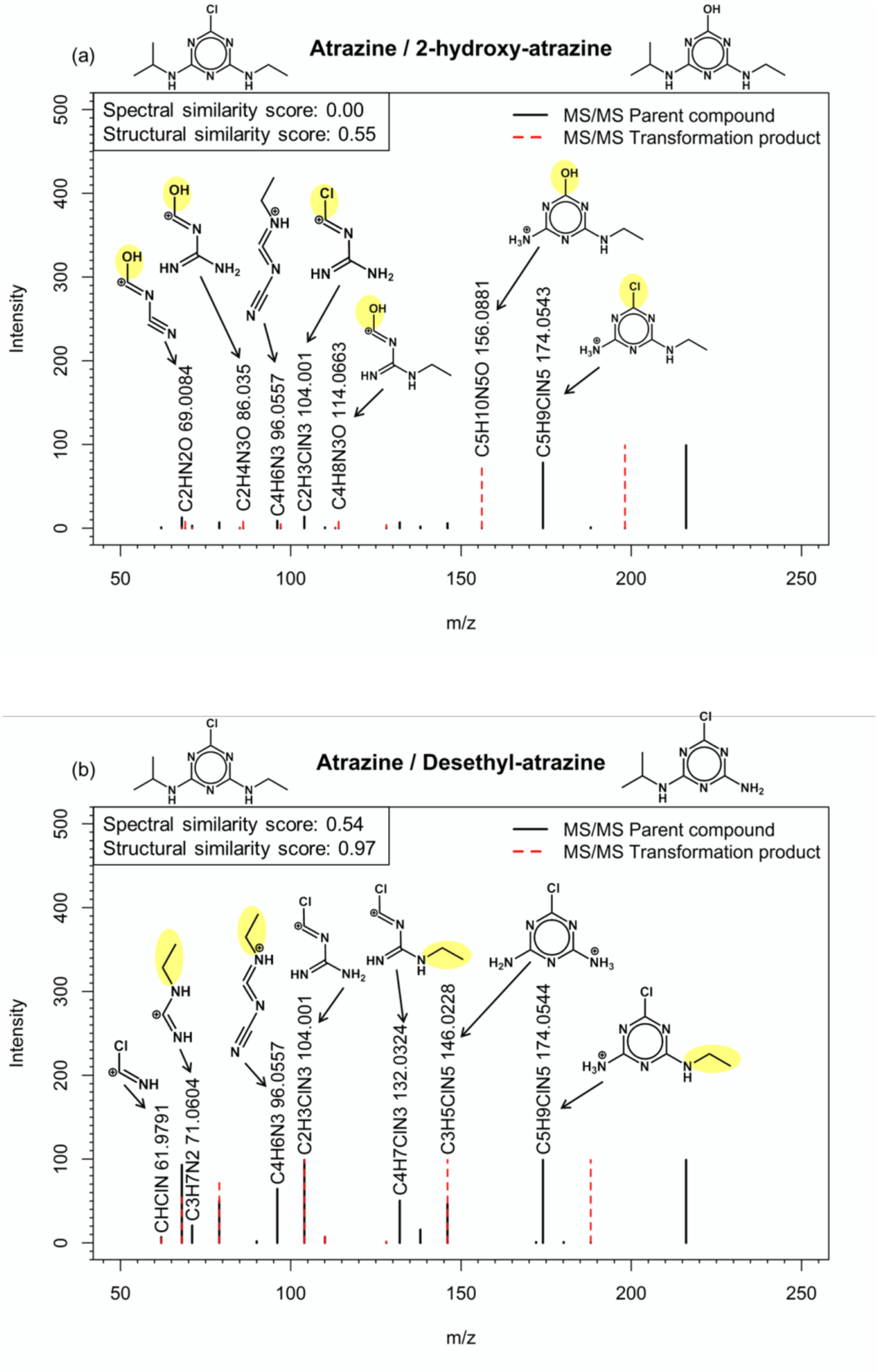


Fig. 1. Comparison of two HRMS2 spectra for two pairs (a) Atrazine and 2-hydroxy atrazine and (b) Atrazine and desethyl-atrazine with different similarity scores but high structural similarity. More fragments overlap in (b), demonstrating that the location of the transformation as well as the transformation itself may have a large influence on fragmentation.

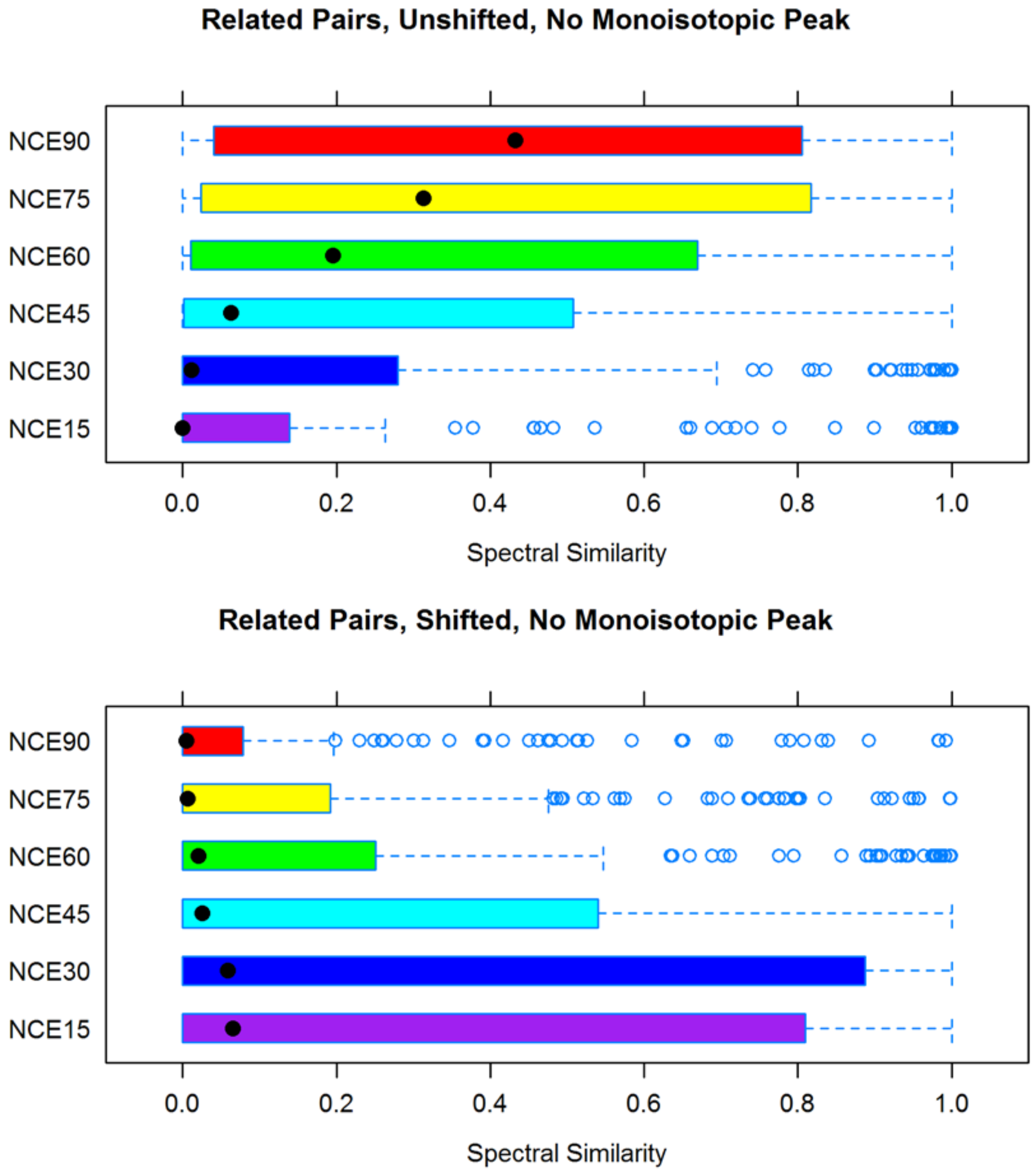


Fig. 2. Box-whisker plots comparing the similarity scores calculated at each NCE. Both (a) unshifted and (b) shifted spectra were used. It is possible to see that there are opposing trends for the shifted and unshifted spectra – highest similarity scores were achieved at highest NCEs for the unshifted spectra, while the opposite was true for the shifted spectra. But the average scores even at the best conditions for the shifted spectra (*i.e.*, NCE30) were much lower than that achieved with the best unshifted spectra conditions (NCE90).

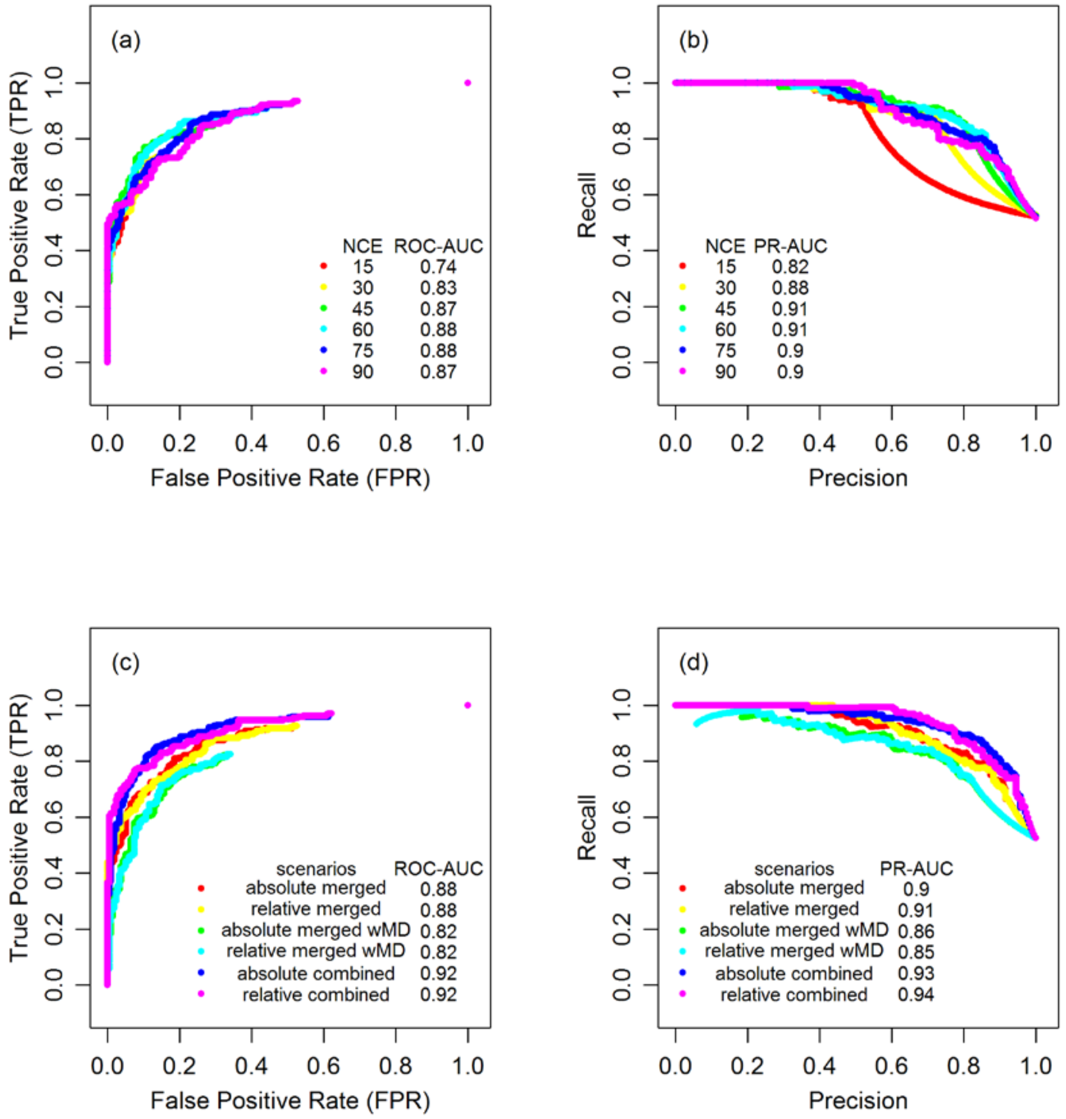


Fig. 3. (a) Shown Receiver Operating Characteristic (ROC) curve for the single collision energy analysis. (b) Precision-Recall (PR) curve for the single collision energy analysis. (c) ROC curve for the merged spectra analysis. (d) PR curve for the merged spectra analysis. Figures (a) and (b) show the results of the single collision energy analysis, while figures (c) and (d) show the results from the merged spectra analysis. In each plot, the different colors designate the different scenarios and the Area under the Curve (AUC) statistic is reported for each scenario.

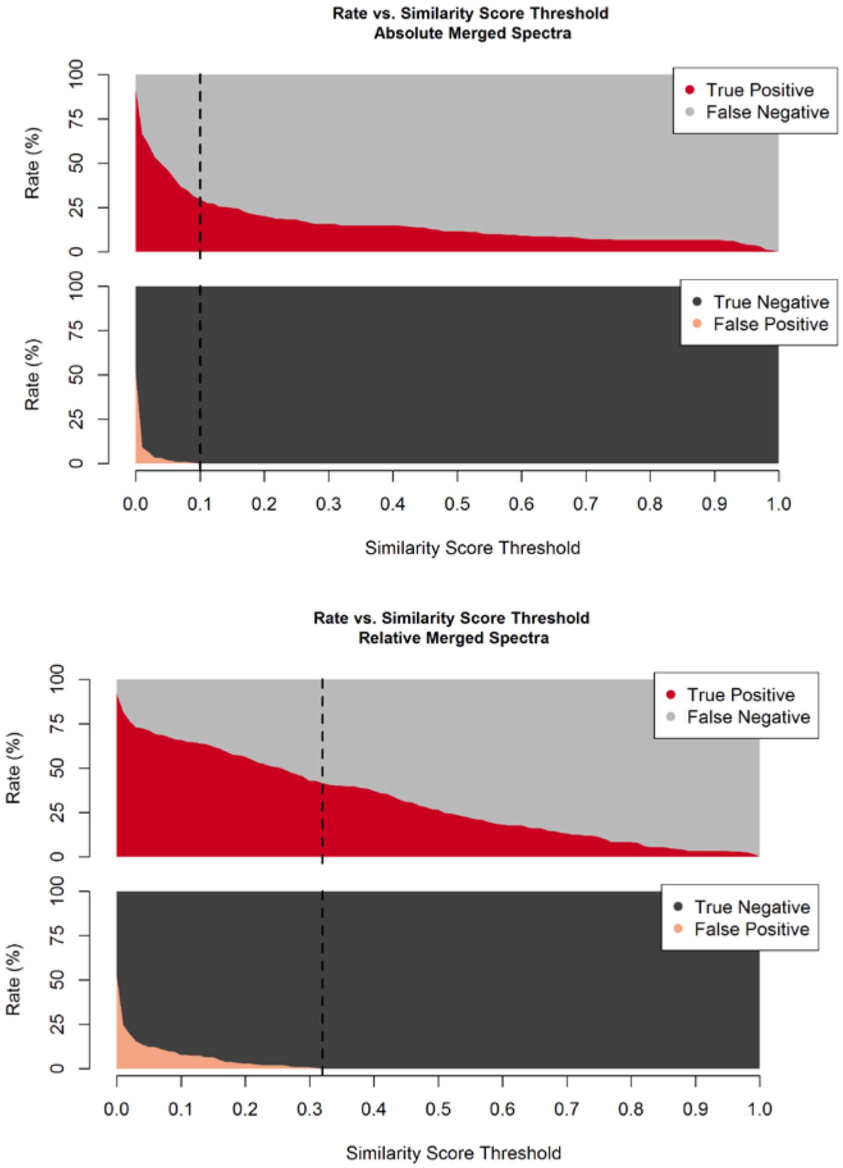
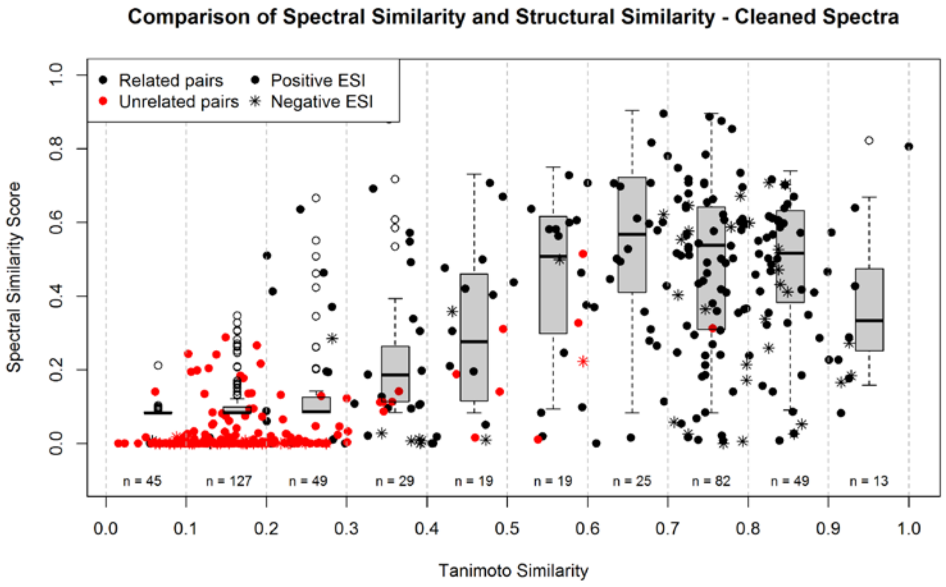


Fig. 4. Similarity score threshold vs. rate of true positives and false negatives in the upper plot and vs. true negatives and false positives in the lower plot. Results shown for (a) absolute merged spectra and (b) relative merged spectra. It is clear that when increasing the similarity score threshold that there is a decrease in the rate of true positives and false positives, while true negatives and false negatives are increasing. For the purposes of this study it was chosen that the optimum similarity score threshold was when false positives was equal to zero (indicated here with a dashed black line).



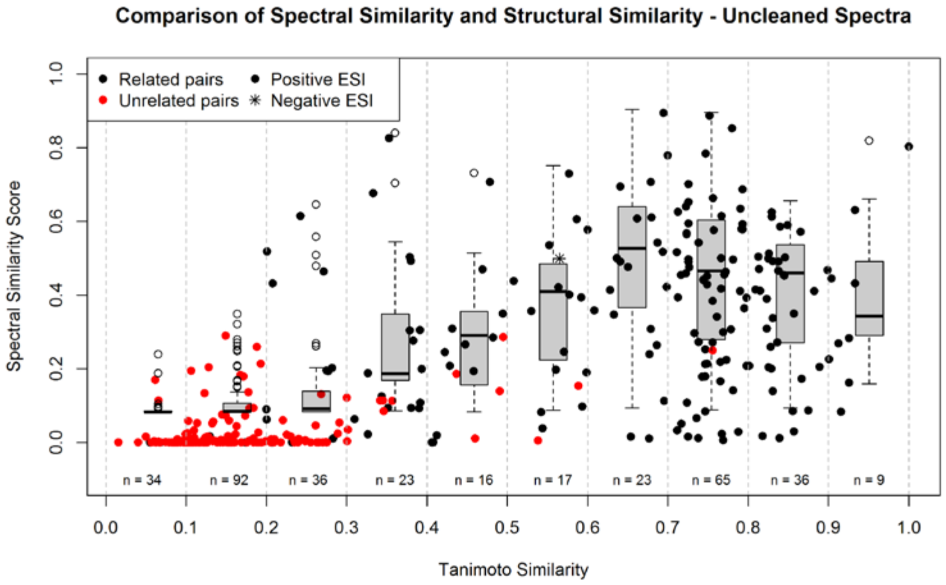


Fig. 5. Structural similarity of the pairs as estimated by Tanimoto Index compared to the spectral similarity as calculated by the cosine dot product (a) for the cleaned spectra (b) for the uncleaned spectra. The spectral similarity used the relative combined merged spectra with both shifted and unshifted fragments while excluding the monoisotopic peak. Positive and negative electrospray ionization pairs are marked separately. In both figures a small correlation can be seen between the two indices and there is a strong resemblance between the cleaned and uncleaned figures.

Table 1. Summary statistics of the scenarios. Areas under the curve (AUCs) of Receiver Operating Characteristic (ROC) and Precision-Recall (PR) curves are reported, as well as spectral similarity score threshold and true positive rate[a] (TPR) at false positive rate[b] (FPR) of 0%. Scenario 1-8 used all fragments, while scenarios 9-20 excluded the monoisotopic precursor peaks (further details in the text and Supplementary Material).

Section		Scenario	Spectral Similarity Scores		ROC-AUC	PR-AUC	TPR[a]	Similarity Score Threshold
			Related Pairs mean +/- std (median)	Unrelated Pairs *mean +/- std (median)*				
Single collision energies	1	NCE15 (unshifted fragments)	0.08 +/- 0.24 (0.0)	0.0 +/- 0.0 (0.0)	0.73	0.80	24%	0.01
	2	NCE30 (unshifted fragments)	0.17 +/- 0.31 (0.01)	0.0 +/-0.0 (0.0)	0.82	0.88	32%	0.09
	3	NCE45 (unshifted fragments)	0.29 +/- 0.37 (0.06)	0.0 +/- 0.0 (0.0)	0.87	0.91	46%	0.16
	4	NCE60 (unshifted fragments)	0.35 +/- 0.38 (0.20)	0.01 +/- 0.04 (0.0)	0.88	0.91	42%	0.33
	5	NCE75 (unshifted fragments)	0.41 +/- 0.38 (0.33)	0.02 +/- 0.07 (0.0)	0.88	0.90	43%	0.46
	6	NCE90 (unshifted fragments)	0.46 +/- 0.39 (0.43)	0.04 +/- 0.09 (0.0)	0.87	0.90	48%	0.44
Merged spectra	7	Absolute merged (unshifted fragments)	0.15 +/-0.26 (0.04)	0.0 +/- 0.01 (0.0)	0.87	0.90	33%	0.10
	8	Relative merged (unshifted fragments)	0.31 +/- 0.29 (0.25)	0.02 +/- 0.06 (0.0)	0.87	0.90	46%	0.32
Shifted spectra	9	NCE15 (shifted fragments)	0.34 +/- 0.41 (0.07)	0.0 +/- 0.0 (0.0)	0.80	0.85	20%	0.9
	10	NCE30 (shifted fragments)	0.34 +/- 0.42 (0.06)	0.0 +/- 0.01 (0.0)	0.81	0.83	0%	1.0
	11	NCE45 (shifted fragments)	0.29 +/- 0.38 (0.03)	0.01 +/- 0.03 (0.0)	0.81	0.84	0%	1.0
	12	NCE60 (shifted fragments)	0.22 +/-0.35 (0.02)	0.01 +/- 0.05 (0.0)	0.80	0.84	6%	0.98
	13	NCE75 (shifted fragments)	0.16 +/- 0.29 (0.01)	0.02 +/- 0.07 (0.0)	0.79	0.83	5%	0.82
	14	NCE90 (shifted fragments)	0.11 +/- 0.23 (0.0)	0.04 +/- 0.09 (0.0)	0.77	0.81	4%	0.71
	15	Absolute merged (unshifted fragments)	0.27 +/- 0.32 (0.10)	0.01 +/- 0.03 (0.0)	0.88	0.90	36%	0.31
	16	Relative merged (unshifted fragments)	0.38 +/- 0.33 (0.31)	0.03 +/- 0.07 (0.0)	0.88	0.91	47%	0.41
	17	Absolute merged (shifted fragments)	0.29 +/- 0.35 (0.08)	0.03 +/- 0.11 (0.0)	0.82	0.86	18%	0.8
	18	Relative merged (shifted fragments)	0.21 +/- 0.26 (0.07)	0.02 +/- 0.08 (0.0)	0.82	0.85	6%	0.72
	19	Absolute combined merged (shifted + unshifted fragments)	0.37 +/- 0.27 (0.40)	0.03 +/- 0.08 (0.0)	0.92	0.93	36%	0.57
	20	Relative combined merged (shifted + unshifted fragments)	0.39 +/- 0.25 (0.42)	0.04 +/- 0.07 (0.0)	0.92	0.94	40%	0.52 **(0.41-0.75)**[c]

[a] Number of related pairs in Scenarios 1-8 was 243 and in Scenarios 9-20 was 240

[b] Number of unrelated pairs in Scenarios 1-8 was 219 and in Scenarios 9-20 was 217

[c] 95% confidence intervals calculated with bootstrapping (n-1000)