B159 A Quantitative Reliability Metric (QRM) for Querying Large Databases

Peter de B. Harrington, PhD, Ohio University, Athens, OH 45701-2979; Zewei Chen, MS, Ohio University, Athens, OH 45701-2979; Vivekananda Shetty, PhD, Houston Forensic Science Center, Houston, TX 77002; Preshious Rearden, PhD, Houston Forensic Science Center, Houston, TX 77002; Angelica Torres Noyola, BS, Houston Forensic Science Center, Houston, TX 77002*

Learning Overview: The goal of this presentation is to showcase a novel measure of database search reliability. This is demonstrated for a Gas Chromatography/Mass Spectrometry (GC/MS) study of novel synthetic opioids in participation with the Houston Forensic Science Center.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by showing that this metric gives a probability that an unknown object matches the closest reference objects returned from the database search. This has broad application because it can be used with any database and any comparison measure (e.g., distance or similarity). It generates a statistical probability of the quality of the match compared to a more similar matching object excluded from the database.

A modern tool for identifying data objects is by comparison to libraries of reference objects; in this case, it is mass spectra. Comparison metrics, such as correlation, are used to search for a list of the closest matching objects. In some instances, the correct identification is not the closest match because the object is not in the reference library or is distorted by noise; therefore, having a quantitative measure of the search reliability can help the analyst determine whether the queried data object is contained in the reference collection or finding the best match that may not be the closest match.

The QRM can be used with any database and any comparison measure. It was recently developed to assist in a National Institute of Justice (NIJ) - funded project for the identification of Novel Synthetic Opioids (NSOs) by GC/MS. A customized library of more than 223,000 mass spectra was queried with several thousand NSO samples. The QRM provided a statistical measure of the search quality but also proved invaluable for chromatographic peak detection, background removal, and library search optimization. The statistical probability compares the ordering of the closest matching reference spectra with the ordering of each matching reference spectrum when it is searched against the library. A variance can be calculated by the sum of the squared differences of the comparison metrics with the intra-library representing the ideal order. These variances are then compared to the variance that would be obtained if the correct matching spectrum was not in the library using the F-distribution.

The QRM is an independent measure of the quality of each library search result that can be used along with the comparison metric. The key advantage is that it can be used with any kind of library and any kind of similarity metric, and it provides a statistical result of the match reliability. For example, if the similarity measure is high and the QRM is low, then there is not much confidence in the result. However, when the similarity measure is low and the QRM is high, than the analyst can be confident in the identification. Last, the QRM provides greater contrast between the list of the closest matching reference objects than the comparison metrics.

Table 1. 4-ANPP (4-aminophenyl-1-phenethylpiperidine, CAS 21409-26-7) ***0.02 ng injected***

Hit Number	Similarity	QRM(%)	Name
1	0.717	92.7	4-ANPP
2	0.713	84.3	4-ANPP
3	0.705	84.1	4-ANPP
4	0.704	83.2	4-ANPP
5	0.698	82.8	4-ANPP
6	0.663	15.0	*N*-methyl-1-phenylcyclohexylamine
7	0.655	25.8	1-ethyl-6,7-dimethyl-2-indolinone
8	0.655	12.4	1,1-dideuterio-2-hydroxyethylcyclohexylcarbamate
9	0.654	10.0	3-butyl-5-methyl-1,8a-dihydroindolizine
10	0.652	3.3	3-methyl-5-*N*-butyl-1,2-benzisoxazole

QRM, Quantitative Reliability Metric, Statistical Probability of Match

*Presenting Author