Determination of Ion Structures in Structurally Related Compounds Using Precursor Ion Fingerprinting

Michelle T. Sheldon,[a] Robert Mistrik,[b] and Timothy R. Croley[a]

[a] Commonwealth of Virginia, Department of General Services, Division of Consolidated Laboratory Services, Richmond, Virginia, USA
[b] HighChem, Ltd., Bratislava, Slovakia

Structurally-related alkaloids were analyzed by electrospray ionization/multiple stage mass spectrometry (ESI/MS^n) at varying collision energies to demonstrate a conceptual algorithm, precursor ion fingerprinting (PIF). PIF is a new approach for interpreting and library-searching ESI mass spectra predicated on the precursor ions of structurally-related compounds and their matching product ion spectra. Multiple-stage mass spectra were compiled and constructed into "spectral trees" that illustrated the compounds' product ion spectra in their respective mass spectral stages. The precursor ions of these alkaloids were characterized and their spectral trees incorporated into an MS^n library. These data will be used to construct a universal, searchable, and transferable library of MS^n spectra. In addition, PIF will generate a proposed structural arrangement utilizing previously characterized ion structures, which will assist in the identification of unknown compounds. (J Am Soc Mass Spectrom 2009, 20, 370–376) © 2009 American Society for Mass Spectrometry

Mass spectrometry (MS) is a well-established and powerful analytical technique. Analytical laboratories throughout the world rely on MS to identify unknown compounds, quantitate ions, and determine pertinent structural information. Currently gas chromatography/mass spectrometry (GC/MS) is the most commonly used method for identification of unknown compounds [1]. GC/MS employs hard ionization techniques [e.g., electron ionization (EI)], which are independent of model or manufacturer, largely reproducible from instrument to instrument, and consequently incorporates standard mass spectral libraries to reliably confirm and/or identify compounds [2]. Accordingly, there are a large number of GC/MS instruments available on the market and a comparable number of commercial spectral libraries to go along with them; yet the technique is limited [3] in that many compounds require derivatization to be analyzed by GC/MS [4]. In instances where there are a large number of samples, this can be significantly time-consuming and labor intensive. More importantly, this means that while EI spectral libraries are undoubtedly reliable and effective, they are only applicable to a restricted number of compounds [3].

Recently, liquid chromatography/mass spectrometry (LC/MS) employing electrospray ionization (ESI) has gained more use as an effective analytical tool. By comparison, ESI/MS can be used to analyze more classes of compounds (to include polar compounds), requires less sample preparation (no derivatization step before analysis), is more sensitive, demonstrates increased specificity, and is quantitatively more accurate [3, 5, 6]. In addition, collision induced dissociation (CID) or tandem mass spectrometry (MS/MS) techniques can be incorporated to provoke fragmentation to elucidate considerable structural information pertaining to a compound [5]. Despite all these advantages, to-date there is no universal, searchable fragmentation spectral library that is not confined to a certain type of instrument and/or a specific class of compounds [7].

Achieving reproducible ESI/MS^n data to archive in a spectral library from different instrumental platforms is no easy task. Many have made attempts to utilize already commercially available resources, while even more have tried to develop their own ESI/MS^n spectral libraries. EI databases, such as the National Institute and Standards and Technology (NIST) and Wiley libraries, are fully evolved and retain an abundant number of compounds. In terms of relations, this would seem like the most reasonable place to start. EI spectra, however in most cases, are very different than those of ESI and fall short of a usable reference database [8]. Fredenhagen et al. attempted to utilize only the NIST algorithm to search through their own in-house ESI-MS/MS spectra [9]. The software used for data acquisition in this study could not export the precursor ion along with the product ion(s) into the NIST database [9]. As a result, only the product ion(s) was/were searchable. Many

Address reprint requests to Dr. T. R. Croley, Commonwealth of Virginia, Department of General Services, Division of Consolidated Laboratory Services, 600 N 5th Street, Richmond VA 23219, USA. E-mail: tim.croley@dgs.virginia.gov

compounds yield an ESI-MS/MS spectrum consisting of only one product ion. Without the presence of a discriminate precursor, there is no way to distinguish those compounds with identical product ions from one another. Any number of compounds can generate a single-ion product spectrum with the same m/z, thereby increasing the probability of false positives [9].

Others have tried employing their own in-house ESI libraries, which is a challenge because there are no sets of "universal" instrumental parameters (instrument design, collision energies, isolation width, etc); therefore, spectra not only vary between instruments but amongst the same models of instruments [3, 10]. Even when instrumental parameters are standardized, there are still incongruities in terms of reproducibility due performance variance from instrument to instrument. Rather than standardize instrumental settings, Hough et al. performed a study predicated on performance-based criteria [3]. The criterion was assigned according to ion ratios and MS/MS spectra acquired at three energy levels: high, medium, and low. This demonstrated reproducibility on the same instrument and different instruments of the same type [3]. While this shows promise, it is still a significant shortcoming. Due to variations in instrumental design, this approach lacks the instrumental versatility that an ESI-MS/MS library necessitates.

To date ESI/MS^n libraries have fallen short of the applicability that GC/MS libraries possess. Recently, a new concept has been introduced [10, 11]. Precursor ion fingerprinting (PIF) is a mode of identifying the structure of ions from their tandem product spectra to elucidate the chemical structure of an unknown. This concept relies on the quasi-equilibrium theory: the probabilities of various possible decomposition products of an ion depend upon its structure, internal energy (E_I), and the energy deposited during ion activation (E_A) but is independent of the ionization method, structure of the precursor ion, and the formation mechanism of the ion undergoing decomposition [12]. Matching spectra of identical ions can be generated by harnessing E_I and E_A, regardless of the original structure and MS^n stage, enabling unequivocal ion identification through PIF. Much like a family shares genes, families of structurally similar compounds share substructures, precursor ions. It is only the manner in which these substructures come together and consequently arrange themselves that make a compound unique. It is as distinctive to its identity as a fingerprint. PIF employs these substructures and fingerprints to make a searchable, applicable database of compounds.

Decomposition product-ions depend only on its original structure, E_A, and E_I. This implies analogous fragmentation patterns. However the degree of spectral analogy (i.e., ion intensities) and reproducibility are dependent on a number of factors. Analogous spectral fingerprints from identical precursor ions are modulated by sample preparation and experimental parameters. By standardizing these two things, highly comparable and searchable MS^n spectra can be produced [13, 14]. To compensate for further spectral variation within these set conditions, there are several other sample preparation processes (pH, solvent, buffer, salt content, etc.) and experimental conditions (instrument design, ionization energies, collision energies, collision gas, collision gas pressure, isolation width, accumulation cut-off, fragmentation amplitude, chromatographic system, etc.) that can be modified to ensure spectral reproducibility. Furthermore, there are certain mathematical methods of assigning a "weight" to known factors that influence spectral variations and that provide an adjustment within certain limitations to account for these factors, which is expressed by a match factor used for compound identification [15]. Thus, reproducible, searchable spectra are accomplished through experimental and instrumental conditions and through the efficiency of the search algorithm itself.

PIF is based on the premise that structurally similar compounds possess matching substructures and subsequent product ion spectra. Previous internal studies were performed on hundreds of linear molecules and molecules with limited structural complexity that confirmed the PIF concept. The PIF algorithm was applied towards the determination of ion structures of topologically complex alkaloids in the opiate family. Purposely, compounds with fused rings were chosen to demonstrate the presented concept for nonlinear molecules that exhibit complex fragmentation patterns. To do this, E_I and E_A were selectively modulated by changing the experimental conditions (normalized collision energy in this study), to generate MS^n spectra of the chosen alkaloids. Performing MS^n (where $n = 1-\infty$) is most important because smaller fragments do not possess all of the substituent groups from the original molecule; that is, it is more likely to identify identical substructures from the higher level MS^n stages (i.e., $n = 3$ or 4) because there are less substituent groups present to potentially effect fragmentation pathways that would result in very different product ion spectra. A hierarchy of spectral fragmentation was constructed to form spectral trees. These spectral trees illustrate the proof of concept that structurally identical precursor ions have matching product ion spectra independent of their original structure. These product ion spectra, or substructures, can be used construct a library of structurally-characterized product ion spectra. This library can be used to identify unknown compounds by generating a proposed structure or structures from its pre-determined ion fragments [10].

Experimental

Codeine and oxycodone hydrochloride (>99% purity) were obtained from Sigma-Aldrich, Inc. (Milwaukee, WI). Heroin and morphine were obtained from Cerilliant (Round Rock, TX) at a concentration of 1.0 mg/mL in acetonitrile. The model compounds were diluted
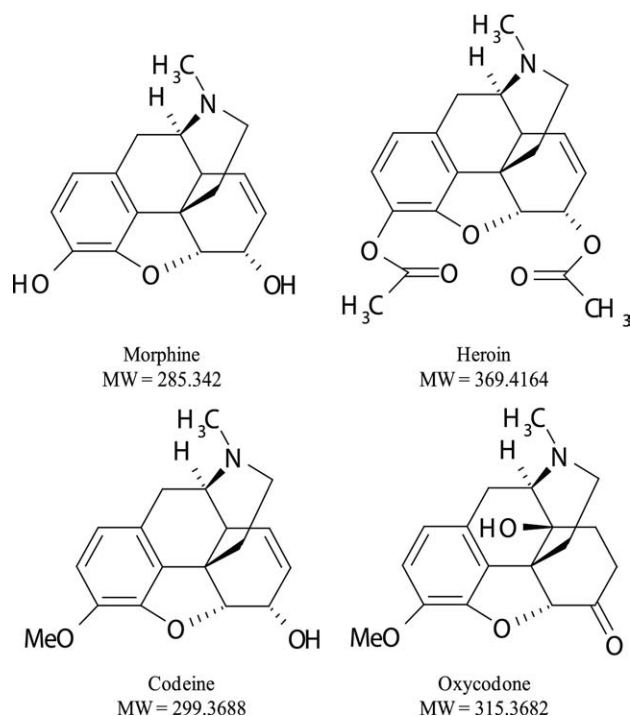
Figure 1. Structures of analytes chosen for this study. These analytes were chosen based their nonlinear, analogous structures.

with high purity grade acetonitrile (Burdick and Jackson, Muskegon, MI) to a concentration of 5 μg/mL.

Electrospray ionization mass spectrometry was performed using a ThermoFinnigan LTQ Linear Ion Trap mass spectrometer (Waltham, MA). The spray needle voltage was +4.5 kV, the heated capillary temperature was 275 °C, auxiliary gas setting 5 units, sheath gas setting 4 units, and sweep gas setting 2 units. Product ion spectra were acquired with three microscans and a maximum injection time of 500 ms.

MS^n experiments were performed on each compound, examining every ion in the mass spectrum. CID mass spectra were acquired at both low (25%) and high (40%) normalized collision energies. Typically 2 to 10 energy-resolved product ion scans for a precursor ion were acquired for up to MS^5 stages.

Database and search capabilities were performed with commercially available Mass Frontier™ 5.0 software (ThermoFisher Scientific, Waltham, MA). Microsoft SQL database is implemented in Mass Frontier. This software also contains search algorithm specially developed for comparison of MS/MS spectra. Average PC computer power is necessary. The complete PIF system will be implemented into this software.

Results and Discussion

In previous work PIF capabilities have been demonstrated for ions emerging from simple cleavage of linear moieties [10]. The objective of this work was to demonstrate this concept for complex structures, heroin, codeine, morphine, and oxycodone, which undergo a

cascade of rearrangement reactions and prove PIF's general applicability. These particular compounds were chosen as model compounds because they are structurally similar members of the opiate family (Figure 1). With PIF, precursor ions are considered "identical" if their product ion spectra exhibit analogous decomposition products. To generate comparable and reproducible spectral fingerprints of ions, factors influencing the appearance of product ion spectra and their effects on match factors were studied. E_I differences arising from ionization or ion activation processes were compensated for by applying the normalized collision energy in the range of 25% to 40%. During each collision E_A is converted to vibrational energy that is rapidly distributed throughout all covalent bonds increasing E_I. Fragment ions are formed when E_I exceeds the activation barrier required for bond cleavage. Therefore in energy-resolved product ion scans of all studied precursor ions, a single matching spectrum that served as a fingerprint of that ion was identified. Subsequently with increasing levels of MS^n, mass spectral fragmentation "trees" evolved, which had common features even though these data came from original structures that were different. The components of a spectral tree are shown in Figure 2.

The spectral tree encapsulates product-ion spectra generations of a single compound into a multiple-stage, hierarchically consistent tree structure. The tree consists of three demonstrative segments: the levels, the "branches", and the nodes. Each level is indicative of an MS^n stage; for example, four levels illustrate stages MS^1–MS^4. Each "branch" is representative of a quantitatively expressed m/z value of a precursor ion that connects a level to a node. A node illustrates the precursor ion's corresponding product ion spectra. Each node has multiple layers of alternative ion fingerprints, which are additional product ion spectra that were acquired under different
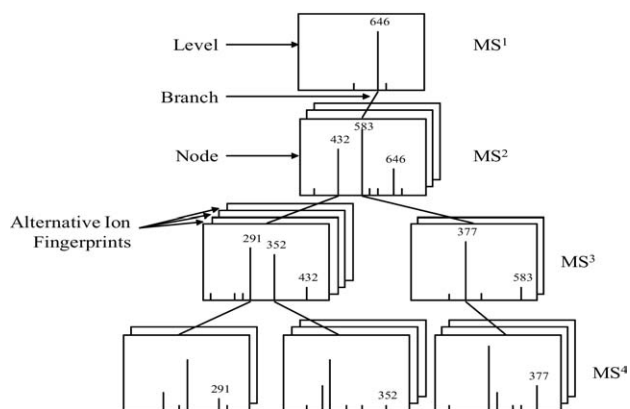

Figure 2. A model of a fragmentation tree used to illustrate all of the MS^n stages, precursor ions, and spectral fingerprints for the target compounds. A level pertains to a mass spectral stage; a node refers to a product ion spectrum; a branch connects a precursor ion to its resulted product ion spectrum (node); alternative ion fingerprints are the additional product ion spectra from precursor ion analyzed under varied instrumental parameters (i.e., collision energies).

experimental conditions to cover a wide range of possible factors affecting spectral dissimilarities. This feature averts the problem of spectral reproducibility. Therefore, an ion can be characterized by several replicate spectra (2 to 10 energy resolved product ion scans in this case), so that the search algorithm can browse through each of them to find the specific match for a query spectrum.

Spectral trees were constructed for the alkaloids chosen for this study. The spectral tree for oxycodone is given in Figure 3. The five "levels" indicated that MS^1–MS^5 was performed. MS^2 yielded two product ions at m/z 229 and m/z 298. Subsequent CID of m/z 298 (MS^3) pointed to secondary inductive cleavage resulting in the loss of an acetate group from oxycodone producing the tertiary cation at m/z 256, which subsequently yielded a product ion at m/z 183 (MS^5). In addition, three other product ions were observed at m/z 161, 175, and 187. Comparison of the spectral trees generated for codeine, morphine, and heroin (not shown) revealed MS^3 of codeine and morphine both yielded an ion at m/z 183. Heroin also contained the ion at m/z 183 observed from MS^4. The CID mass spectra of m/z 183 for each analog at a normalized collision energy of 25% is seen in

Figure 4b. Not surprisingly, the product-ion mass spectra are nearly identical. All four produced product ions at m/z 165 and 155 independent of MS^n stage or the original structure of the compound. Thus, subsequent levels of MS^n (MS^n: n = 2–5) produced product ion spectra which were identical when the precursor ion was the same.

The spectral trees from the model compounds identified 10 precursor ions that were common among a minimum of two of the four analytes (Table 1). For example, five ions (m/z 193, 201, 211, 229, and 239) were products of three compounds. The spectral trees highlight the need to scrutinize every ion in a mass spectrum and to not ignore those ions that are low in intensity. For example, codeine propagated 12 observable precursor ions in the MS^3 stage. Eight of these coincided with at least one other analog. Morphine generated nine ions in the MS^3 stage, in which seven of these matched another compound. The corresponding mass spectra for these precursor ions and the MS^n stage in which they were observed are shown in Figure 4. In Figure 4a, c, d, and i, the product ion intensities of these spectra are reasonably equivalent. However the product ion intensities observed in Figure 4b, e–h, and j are not. While
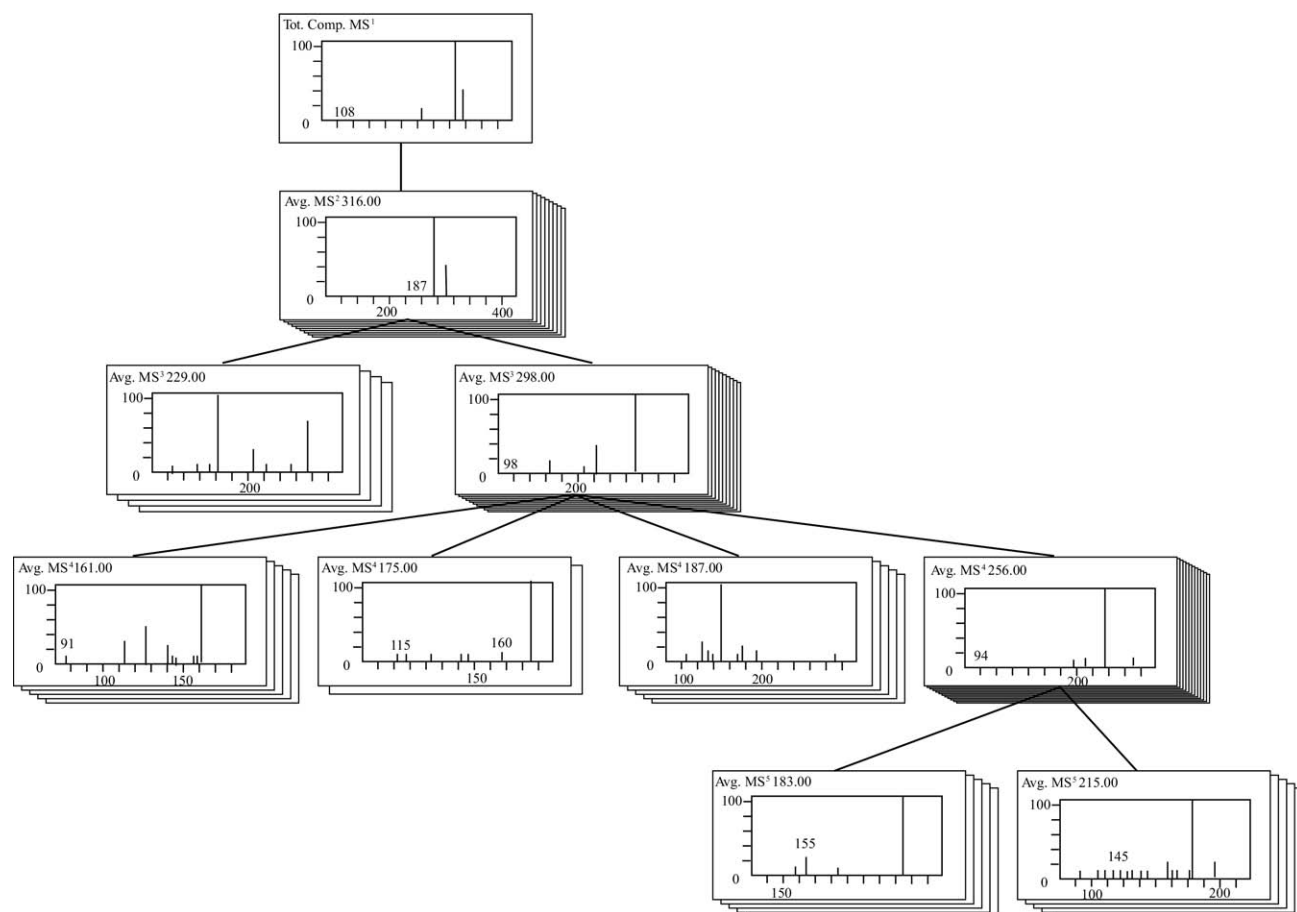
Figure 3. Spectral tree of oxycodone. Each level is indicative of the prospective MS^n stage, and each "branch" is representative of a m/z of a precursor ion that extends out to a node comprised of its corresponding product ion spectra.
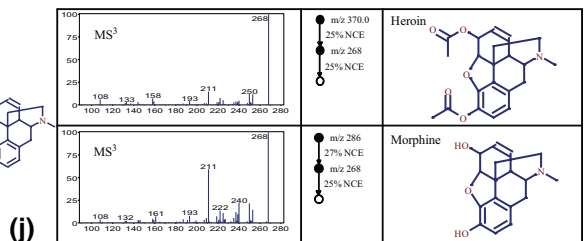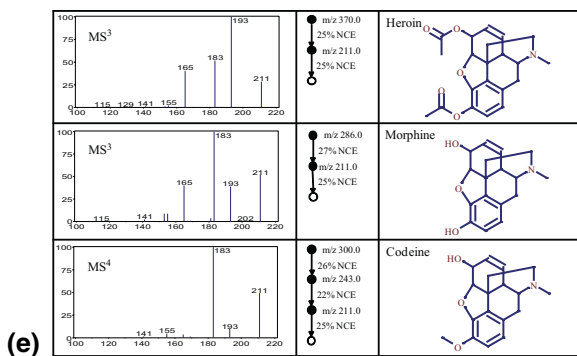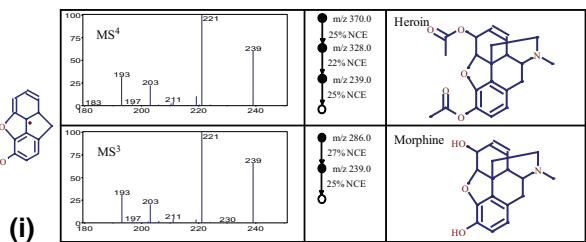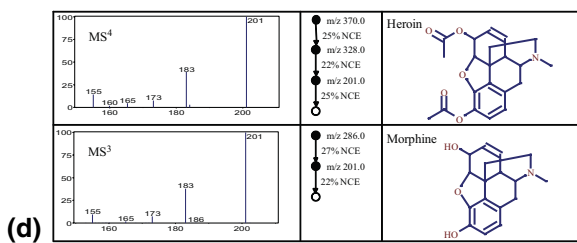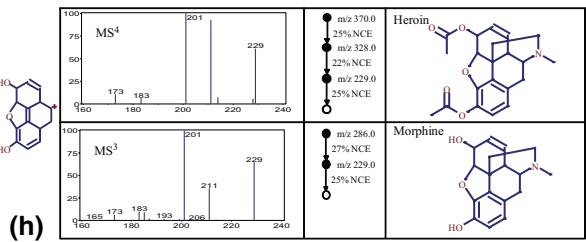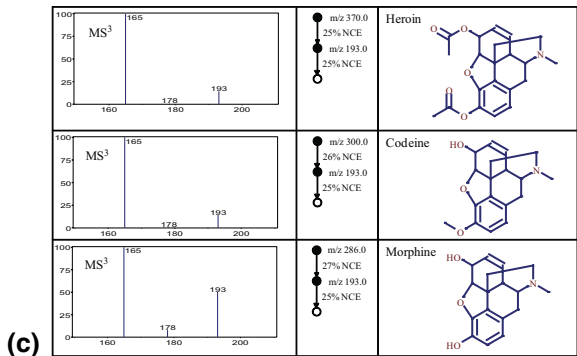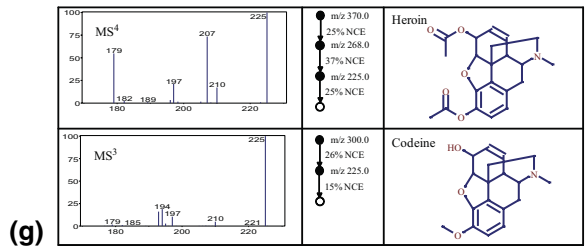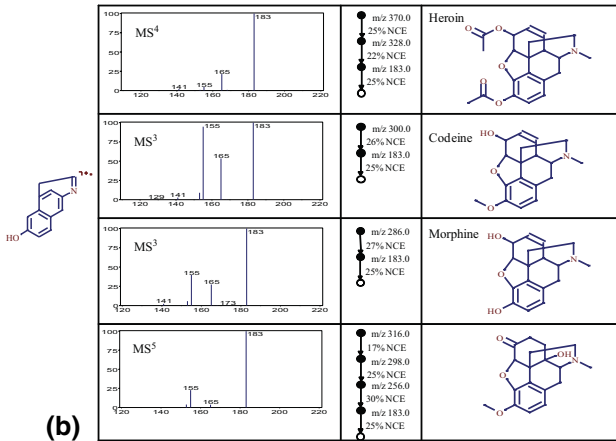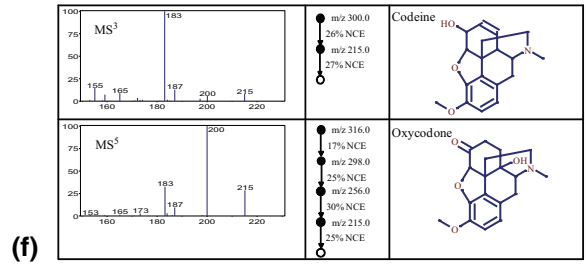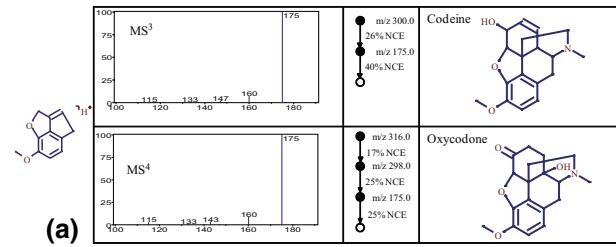
Table 1. Precursor ion fingerprinting matches for the four model compounds at different MS^n stages

m/z	Heroin	Codeine	Morphine	Oxycodone
175		MS^3		$MS^3/298$
183	$MS^3/328$ MS^4	MS^3	MS^3	$MS^4/256$ MS^5
193	MS^3	MS^3	MS^3	
201	$MS^3/328$ MS^4	$MS^3/282$	MS^3	
211	MS^3	$MS^3/243$ MS^4	MS^3	
215		MS^3		MS^3 298 $MS^4/256$ MS^5
225	$MS^3/268$ MS^4	MS^3		
229	$MS^3/328$ MS^4		MS^3	MS^3
239	$MS^3/328$ MS^4	$MS^3/282$	MS^3	
268	MS^2		MS^3	

the same ions are present in all the spectra, the ion intensities are disproportionate. These observations further emphasize the importance of acquiring data at multiple collision energies in each MS stage from multiple ions of a spectrum rather than focusing only on the major ions in a mass spectrum. This approach allowed the degree of similarity among these alkaloids to be better ascertained, and the variations among their mass spectra were further compensated. Overall, this ensures a more definitive ion fingerprint, maximizes comparability, and accounts for poor specificity and reproducibility.

Determining the identical precursor ions generated by the four model compounds was essential to corroborating the concept of PIF. Inherently, this must be taken one step further. To build a library of searchable ion spectra, accurate ion structures must be determined. The fact that at least two different structures were used for mechanistic ion structure assignment proved to be essential for compounds like heroin and similar alkaloids. In addition, an in-house fragmentation library of more than 100,000 decomposition mechanisms was collected from peer-reviewed literature supporting ion structure assignments [16]. The identical product ion spectra and corresponding precursor ion structure assignments are illustrated in Figure 4 a–j. Some precursor ions demonstrating identical product-ion spectra were not given a structure assignment (Figure 4 c–g). These substructures either could not be found in the literature or could not be appointed a conclusive structure. However, this

information is still very useful. Figure 4e shows the spectral fingerprint of the precursor ion m/z 211, characteristic of heroin, morphine, and codeine. While there is no structural assignment for this ion, there is one for the precursor ion m/z 183 (Figure 4b), which can be observed in all three decomposition spectra of the m/z 211 ion. The precursor ion m/z 183 was noted in the decomposition spectra of precursor ions m/z 201 (Figure 4d) and 215 (Figure 4f) as well. While the definitive structure of these ions may still be unknown, important structural information can be extrapolated, and each precursor ion remains an integral part of the structure of the three alkaloids, their spectral fingerprints, and the library.

Consistent CID behavior provides a reliable means for employing the concept of PIF for deriving structures of unknown compounds whose higher level CID spectra are the same as known structures in a database of ion structures. This, like traditional libraries, is largely dependent on the number of compounds in the database. The more characterized precursor ions in the library, the more likely it is to solicit a match. If an unknown spectrum can, in fact, be matched with a reference spectrum, and it is determined that they are "identical", a substructure of that compound has been identified. If nothing else, this implies that the unknown compound may be structurally related to other compounds in the library. Following the initial experiment with the four model compounds, MS^n (n = 1–4) was performed on apomorphine (Figure 5), another

←

Figure 4. (a–j) The common moieties corresponding to identical precursor ions of heroin, codeine, morphine, and oxycodone and their matching spectral fingerprints. Despite different structures and MS^n stage disparity, identical precursor ions yield nearly identical product-ion spectra. (a) Matching product ion of the m/z 175 precursor ion dissociated from codeine and oxycodone, and its precursor ion structure assignment. (b) Matching product ion of the m/z 183 precursor ion dissociated from heroin, codeine, morphine, and oxycodone, and its precursor ion structure assignment. (c) Matching product ion of the m/z 193 precursor ion dissociated from heroin, codeine, and morphine. (d) Matching product ion of the m/z 201 precursor ion dissociated from heroin and morphine. (e) Matching product ion of the m/z 211 precursor ion dissociated from heroin, morphine and codeine. (f) Matching product ion of the m/z 215 precursor ion dissociated from codeine and oxycodone. (g) Matching product ion of the m/z 225 precursor ion dissociated from heroin and codeine. (h) Matching product ion of the m/z 229 precursor ion dissociated from heroin and morphine and its precursor ion structure assignment. (i) Matching product ion of the m/z 239 precursor ion dissociated from heroin and morphine and its precursor ion structure assignment. (j) Matching product ion of the m/z 268 precursor ion dissociated from heroin and morphine and its precursor ion structure assignment.
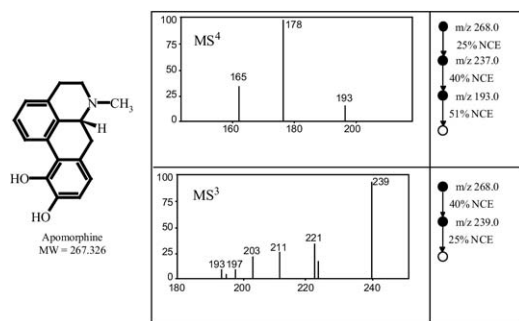
Figure 5.

member of the opiate family and somewhat structurally similar to the fingerprinted alkaloids. While apomorphine was not found in the database, compounds previously fingerprinted (codeine, morphine, etc.) were. Like the other model compounds, MS^n spectra for apomorphine were collected for many precursor ions rather than for only those over a certain intensity threshold. The data indicated that there were two precursor ions, m/z 239 and 193, which were common to heroin, codeine, and morphine. Observe the matching product ion spectra of fingerprinted alkaloids (Figure 4c and i) for the 239 and 193 ions compared with that of the apomorphine in Figure 5. Identifying these substructures, at the very least, suggested assimilation with the other structurally characterized compounds despite apomorphine itself not being present in the library. This information could be decisive in identifying unknowns not available to the database.

These results point to the potential for PIF to process identified ion substructures through a structure generator as an additional tool for identifying unknowns. This is performed by determining the potential structures of each product ion, including its respective precursor ion through the spectral library, and building the potential structure of the unknown compound from the ground up. This approach is unique in that all spectra are searched in spite of their mass spectral stage. In other words, an MS^2 ion from an unknown could be compared with and identified by an MS^3 ion from the library. If there is enough fragmentation information available (i.e., enough fragment structures identified), these are processed by a structure generator. This process allows a number of molecular structural candidates to be proposed that can be ranked according chemical relevance. If one conclusive structure can be determined, that structure can be subsequently identified, characterized, and added to the database. Over time, identification of more unknowns strengthens the size, effectiveness, and overall capability of the PIF library. This characteristic creates more versatility than traditional approaches.

Conclusions

Traditional library search methods are designed to identify compounds represented in a reference library;

when the unknown compound is not in the library, the compound cannot be identified. To overcome this shortcoming, precursor ion fingerprinting (PIF) has been developed to obtain the structural arrangements of unknown. Spectral fingerprints of fragment ions formed in the mass spectrometer allow identification of ion structures and subsequent reconstruction of the molecular structure of small organic compounds. This approach is particularly promising for applications operating in a structurally finite space, such as the characterization of structurally similar compounds, and enables identification of designer drugs or novel chemical analogs that are usually not present in any spectral library at the time of analysis.

Acknowledgments

References

1. Stein, S. E. Estimating Probabilities of Correct Identification from Results of Mass Spectral Library Searches. *J. Am. Soc. Mass Spectrom.* **1994,** *5,* 316–323.
2. Jiang, H.; Somogyi, A.; Timmermann, B. N. Instrument Dependence of Electrospray Ionization and Tandem Mass Spectrometric Fragmentation of the Gingerols. *Rapid Commun. Mass Spectrom.* **2006,** *20,* 3089–3100.
3. Hough, J. M.; Haney, C. A.; Voksner, R. D.; Bereman, R. D. Evaluation of Electrospray Transport CID for the Generation of Searchable Libraries. *Anal. Chem.* **2000,** *72,* 2265–2270.
4. Knapp, D. R. *Handbook of Analytical Derivatization Methods*; John Wiley: New York, 1979.
5. Josephs, J.L; Sanders, M. Creation and Comparison of MS/MS Spectral Libraries using Quadrupole Ion Trap and Triple-Quadrupole Mass Spectrometers. *Rapid Commun. Mass Spectrom.* **2004,** *18,* 743–759.
6. Lee, M. S.; Kerns, E. H. LC/MS Application in Drug Development. *Mass Spectrom. Rev.* **1999,** *18,* 190.
7. Bristow, A. W.; Nichols, W. F.; Webb, K. S.; Conway, B. Evaluation of Protocols for Reproducible Collisionally Induced Dissociation on Various Liquid Chromatography/Mass Spectrometry Instruments and the Development of Spectral Libraries. *Rapid Commun. Mass Spectrom.* **2002,** *16,* 2374–2386.
8. Milman, B. L. Identification of Chemical Compounds. *Trends Analyt. Chem.* **2005,** *24,* 493–508.
9. Fredenhagen, A.; Derrien, C.; Gassmann, E. An MS/MS Library on an Ion-Trap Instrument for Efficient Dereplication of Natural Products. Different Fragmentation Patterns for [M + H] and [M + Na] Ions. *J. Nat. Prod.* **2005,** *68,* 385–391.
10. Mistrik, R. Determination of Molecular Structures Using Tandem Mass Spectrometry. Patent no. U.S. 7197402 B2, 2007.
11. Sheldon, M. T.; Croley, T. R.; Mistrick, R. Determination of Ion Structures in Structurally Related Compounds Using Precursor Ion Fingerprinting. Poster presentation. *Proceedings of the 55th ASMS Conference on Mass Spectrometry and Allied Topics*; Indianapolis, IN June 2007.
12. McLafferty, F. W.; Turecek, F. *Interpretation of Mass Spectra*, 4th ed.; University Science Books: Mill Valley, CA, 1993; p. 118–120.
13. Bristow, A. W.; Webb, K. S.; Lubben, A. T.; Halket, J. Reproducible Product-Ion Tandem Mass Spectra on Various Liquid Chromatography/Mass Spectrometry Instruments for the Development of Spectral Libraries. *Rapid Commun. Mass Sperctrom.* **2004,** *18,* 1447–1454.
14. Pittenauer, E.; Zehl, M.; Mistrik, R.; Allmaier, G. Evelution of ESI-MS^n Ion Trap and MALDI-MS^n Ion Trap/Reflectron TOF Mass Spectra (n = 1–3) for Incorporation into a Searchable Database. *Proceedings of the 52nd ASMS Conference on Mass Spectrometry and Allied Topics*; Nashville, TN May 2004.
15. Sander, P. Substance Identification of Ion Trap MS/MS Spectra in a MS/MS Library. *Proceedings of 47th ASMS Conference on Mass Spectrometry and Allied Topics*; Dallas, TX June 1999.
16. Mass Frontier 6.0 software, HighChem, Ltd., 2006.