Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography A, 1050 (2004) 159–171
JOURNAL OF CHROMATOGRAPHY A
www.elsevier.com/locate/chroma

Automated peak tracking for comprehensive impurity profiling in orthogonal liquid chromatographic separation using mass spectrometric detection

Gang Xue*, Anne D. Bendick, Raymond Chen, Sonja S. Sekulic

Pfizer Global R&D, Easter Point Road MS 8118 D-3087, Groton, CT 06340, USA

Received 27 February 2004; received in revised form 1 June 2004; accepted 12 August 2004

Abstract

The presence and quantity of impurities in pharmaceutical drugs can have a significant impact on their quality and safety. With the continuous pressure for increased industry productivity, there is urgent need for a systematic and comprehensive drug impurity profiling strategy. We report here our development of the fully automated Comprehensive Orthogonal Method Evaluation Technology (COMET) system. The system includes five columns, seven orthogonal HPLC methods, and hyphenated UV–MS detections, which provides automated generic impurities screening for any drug sample. An automated MS peak tracking approach by program-based mass spectral interpretation is devised to unambiguously track impurities among all orthogonal HPLC methods. The program passes electro-spray ionization mass spectra (ESI-MS) through four sequential decision-making mass ion tests and determines molecular weights for every peak. The system reduces the time required to obtain impurity profile from weeks to days, while the automated MS peak tracking takes only minutes to interpret all MS spectral data of interest. Up-to-date, impurity contents of 56 in-development drug candidate samples have all been successfully illustrated by COMET, which contained more than 500 chemical entities. The program is able to track more than 80% of the compounds automatically with majority of the failure due to insufficient ionization for some impurities by ESI. This system is well suited for efficient drug development and ensuring the quality and safety of drug products.

Keywords: Peak tracking; Impurity profiling; Orthogonal liquid chromatography

1. Introduction

During the production of active pharmaceutical ingredients (API), many opportunities for the generation of impurities may arise [1]. The safety of the drug depends not only on the toxicological properties of the API itself, but on the impurities it contains. For this reason, accurate assessment of impurity profiles of API is one of the most important fields of activity in pharmaceutical analysis. In the meanwhile, any changes in the synthetic process, such as synthetic routes, reaction conditions, purification processes, etc. which are common as a drug candidate goes through various development stages and process scale up, may lead to changes in the impurity profile and in questions about safety. Hence, the impurity profiling becomes a frequent task, which often time-consuming and labor-intensive [2]. It is, therefore, essential to have a strategy that enables the task to be performed efficiently and automatically.

As chromatographic techniques, such as HPLC been established as main workhorse for impurity analysis, it is natural to devise a generic HPLC method for this purpose. However, APIs in different therapeutic area may have dramatically different physiochemical properties, not to mention yet even more diversity of process related impurities. It is hardly true that a single generic HPLC method can provide sufficient resolution for impurity profiling of all drugs. Instead, orthogonal

* Corresponding author. Tel.: +1 860 715 6697; fax: +1 860 686 1244.
E-mail address: gang_xue@groton.pfizer.com (G. Xue).

0021-9673/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2004.08.030

HPLC methods, i.e. methods that differ significantly in chromatographic selectivity, would jointly provide peak capacity and resolving power magnitudes higher than any individual method, as demonstrated in many two-dimensional HPLC applications [3–7]. Hyphenated detections, such as ultraviolet (UV), mass spectrometer (MS), infrared (IR), and nuclear magnetic resonance (NMR), etc. offer additional means of orthogonality [8–14]. Automated orthogonal HPLC screening system with hyphenated detection would provide true comprehensive drug impurity profile in an "one-shoot" fashion for maximum efficiency.

The screening using orthogonal HPLC and hyphenated detection, on the other hand, inevitably lead to increased complexity in the acquired data and so forth information processing. The chromatograms that result from orthogonal HPLC methods exhibit changes in impurity retention orders. Therefore, recognition of the same impurities in multiple chromatograms, i.e. peak tracking, becomes crucial in successful impurity profiling. A wide range of peak tracking strategies has been proposed. One straightforward method is to inject individual pure standard in the same experimental conditions as the samples, which is very time consuming, not to mention that in many cases pure impurity standards are not available [15,16]. Some other early approaches include comparison of normalized peak areas [17], wavelength ratios [18,19] or retention values [20–22]. These methods are too sensitive to changes in experimental conditions and require all components have distinct peak areas. Therefore, the tracking capabilities of these methods are very limited, especially when peaks overlap. The introduction of photodiode array detector (DAD) for HPLC instrumentation, which made it possible to acquire full UV–vis spectra, considerably increases the information content of the chromatograms [23]. Computerized point-by-point numerical comparison of spectra data provides more precise peak tracking capabilities [24,25]. Mathematically enhanced spectral analysis and spectral contrast techniques are introduced to quantify the UV spectra shape difference, which is defined as match angle [26]. The match angle increases as the shape difference increases, resulting in values from 0° (identical) to 90° (maximum dissimilarity), which allows automated spectra library search. Round et al. reported a peak tracking software based normalized spectra overlay comparison (NSOC), which utilized weighted combinations of zero-order UV spectra and its first-order and second-order derivatives to enhance the resolution between spectral differences [27]. The use of multi-wavelength detection also offers the possibility to resolve overlapping peaks. A peak tracking method based on multivariate analysis of DAD spectra data is introduced by Strasters et al. [28–32]. In order to compare the spectral characteristics of overlapping components in peak clusters, each cluster is analyzed separately using principal component analysis (PCA) to determine the number of components, followed by iterative target testing factor analysis (ITTFA) to obtain pure spectra and chromatographic profiles. Then, peak tracking is performed based on comparison of spectra, combined with an examination of peak areas. Clusters with too little resolution for the ITTFA are resolved by means of target factor analysis, using the already obtained pure spectra as targets. A more recent work by van Zomeren et al. further extends the multivariate peak tracking approach with augmented ITTFA in order to remove the intensity ambiguity and rotational ambiguity, which are caused by factor analysis [33].

However, there are some limitations for the aforementioned UV spectra peak tracking methods. First of all, all components of the sample must have sufficiently different spectra. Unfortunately, UV spectra are rather broad and lack of distinct fingerprint spectra zones as those in infrared spectra. It is, therefore, difficult to distinguish various structurally related compounds. Even the enhanced cross-match normalization showed only limited success in the tracking of various peptide solutes from tryptic digests of porcine growth hormone [27]. More seriously, varying the pH and solvent in orthogonal HPLC methods may substantially change the spectral characteristics of the components [26]. For impurity profiling specifically, many of the impurity concentrations are within the 0.1% regime relative to the API. Their weak UV absorption signals are generally obscured by the dominating background spectra of mobile phases [24,25]. If the spectra difference induced by the eluent composition variation or required background subtraction is larger than that of different solutes, peak tracking using UV spectra match would be impractical.

In recent years, the hyphenation of mass spectrometer to HPLC analyses has become increasingly common with the proliferation of low-cost, smaller and more user-friendly instruments [34–36]. On the contrary to the ambiguity in UV spectra match, mass spectra offer unrivaled detection accuracy and specificity [37]. Mass-to-charge ratios of quasi-molecular ion ($[M + H]^+$) and its fragmentation ions provide high confidence in compound characterization and structure elucidation [38]. Matrix effects, as long as it does not have iso-mass ions with respect to the analyte, would be resolved by the mass selectivity. For these reasons, MS undoubtedly turns out to be a more dependable source for peak tracking.

In this article, we present an automated peak tracking approach by electrospray mass spectrometry (ESI-MS) for the comprehensive impurity profiling by Comprehensive Orthogonal Method Evaluation Technology (COMET). The system integrates five columns, seven orthogonal HPLC methods, and hyphenated UV–MS detections to achieve automated generic impurities screening for any drug sample. A series of decision-making mass ion tests are applied to all the mass spectra of interest to unambiguously identify the molecular weight of each impurity. The successful peak tracking based on the assigned molecular weights thus enables correlating results from multiple orthogonal HPLC methods and elucidates comprehensive impurity profile for any pharmaceutical samples, all in fully automated fashion.

2. Experimental

2.1. Materials and supplies

HPLC grade acetonitrile (ACN), methanol (MeOH), tetrahydrofuran (THF) and ACS Reagent grade ammonium hydroxide solution (28.0–30.0%) were obtained from J.T. Baker (Phillipsburg, NJ). HPLC grade glacial acetic acid, ammonium acetate and ACS reagent grade formic acid (+96%) were obtained from Fisher Scientific (Pittsburgh, PA). The 99.5% purity trifluoroacetic acid (TFA) was obtained from Fluka (Buchs, Switzerland). House de-ionized water was filtered through a Millipore (Bedford, MA) Milli-Q system using a 0.22 μm Millipak filter.

Trifluoroacetic acid, formic acid and glacial acetic acid solutions were prepared at 0.1% (v/v) by adding 1 mL of acid to 1 l of water and mixing well. The pH of the acetic acid solution was adjusted to 3.5 using ammonium hydroxide. Two solutions of ammonium acetate were prepared at 10 mM concentration by adding 0.77 g of ammonium acetate to 1 l of water. One solution was adjusted to pH 5 using glacial acetic acid. The other has a pH of about 7 without pH adjustment. Both ammonium acetate solutions were filtered using 0.2 μm Nylon 66 filters from Pierce (Rockford, IL).

YMC J'Sphere ODS H80 column (4.6 mm × 150 mm, 4 μm) and Spherisorb ODS1 (4.6 mm × 100 mm, 5 μm) were obtained from Waters (Milford, MA). Luna phenyl-hexyl column (4.6 mm × 100 mm, 5 μm) was obtained from Phenomenex (Torrance, CA). Kromasil C4 (4.6 mm × 250 mm, 5 μm) and Monitor C18 (4.6 mm × 150 mm, 5 μm) were purchased from Column Engineering (Ontario, CA).

2.2. COMET instrumentation

A Hewlett-Packard 1100 liquid chromatography system (Agilent Technologies, Palo Alto, CA) equipped with a vacuum degasser, quaternary pump, auto-sampler, UV diode array detector and HP-MSD model D mass spectrometer is used as the backbone of the automated system. The system is modified to include a thermostated Spiderling CS400010-position column selector from Chiralizer Services (Newtown, PA) and a six-position Cheminert valve (Valco Instruments Co. Inc., Houston, TX) is inserted on the "D" line of the quaternary pump to extend mobile phase selection (Fig. 1). Both the column switcher and the Cheminert valve are connected through contact closures on the back of the chromatography system and can be controlled electronically through the ChemStation software.

The mobile phase components are placed in the reservoirs for the Agilent 1100 HPLC as follows: line A contains tetrahydrofuran; line B contains acetonitrile; line C contains methanol and line D contains the aqueous solutions connected to the six-position Cheminert valve. The mobile phase valve positions 1–6 contain the aqueous solutions water, 0.1% formic acid (pH ∼2.1), 0.1% glacial acetic acid (pH ∼3.5), 10 mM ammonium acetate (pH 7.0), 0.1% trifluoroacetic acid (pH ∼1.9) and 10 mM ammonium acetate (pH 5.0), respectively.

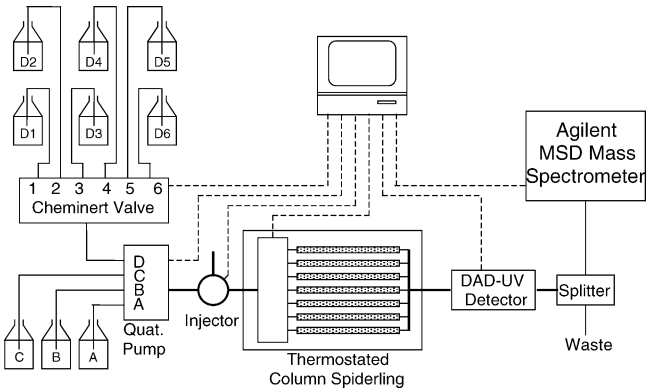

Fig. 1. Schematic diagram of automated COMET system.

2.3. LC/MS methods

Extensive evaluation of 32 HPLC columns and 18 mobile phase combinations were performed to provide a series of MS compatible methods, each with unique selectivity. The column evaluation mixture containing 18 Pfizer proprietary active pharmaceutical ingredients at various stages of development was run through each column and mobile phase combination (data not shown). The geometric factor analysis approach was applied to the chromatographic data to evaluate method orthogonality and practical peak capacity (N_p) [39]. A subset of the column and mobile phase combinations

Table 1
Orthogonal HPLC methods selected for COMET screen

Method 1 (M1)	*Column*: Kromasil C4, 4.6 × 250 mm, 5 μm *Mobile phase*: 0.1% trifluoroacetic acid pH ∼1.9/acetonitrile
Method 2 (M2)	*Column*: Luna phenyl-hexyl, 4.6 mm × 100 mm, 5 μm *Mobile phase*: 0.1% acetic acid pH 3.5/acetonitrile
Method 3 (M3)	*Column*: Luna phenyl-hexyl, 4.6 mm × 100 mm, 5 μm *Mobile phase*: ammonium acetate pH ∼7.0/acetonitrile
Method 4 (M4)	*Column*: Luna phenyl-hexyl, 4.6 mm × 100 mm, 5 μm *Mobile Phase*: ammonium acetate pH 5.0/tetrahydrofuran
Method 5 (M5)	*Column*: Spherisorb ODS1, 4.6 mm × 150 mm, 3 μm *Mobile phase*: ammonium acetate pH ∼7.0 /MeOH
Method 6 (M6)	*Column*: Monitor C18, 4.6 mm × 150 mm, 5 μm *Mobile phase*:0.1% acetic acid pH 3.5/MeOH
Method 7–9 (M7–9)	*Column*: YMC J'sphere ODS H80, 4.6 mm × 100 mm, 4 μm *Mobile phase*: 0.1% formic acid pH ∼2.1/acetonitrile

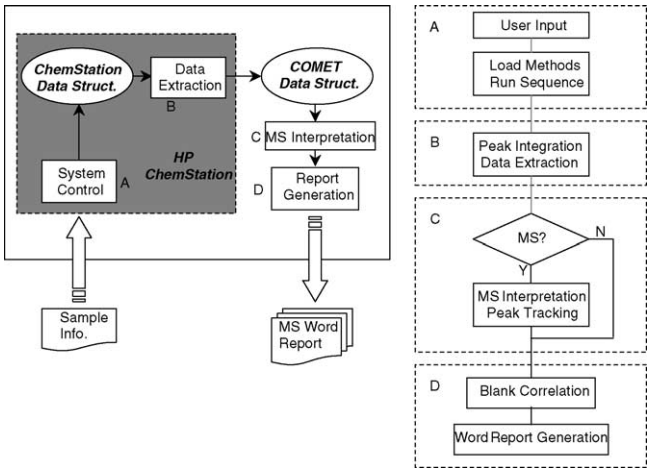

Fig. 2. COMET software layout and flow chart.

(Table 1), demonstrating both maximum orthogonality and highest N_p, were chosen as pre-determined methods for the routine COMET screening. M1–M6 and M8 are 60 min linear gradient methods; while M7 is 30 min short gradient and M9 being 90 min long gradient method, respectively, all ramping from 5 to 95% organic modifier. Column equilibrate time is set to 12 min for each method. M7–M9 use the same column-mobile phase combination, which show little orthogonality. The inclusion of two additional gradients for M8 is for the purpose of interfacing to DryLab for simulated method optimization. Thus, the COMET system essentially comprises seven orthogonal methods. As a general rule, fairly low additive concentrations (0.1% or 10 mM) were chosen to enhance absorption detection sensitivity at lower UV range. The column compartment is thermostated at 30 °C.

2.4. Control software

The complete system is controlled by a single IBM NetVista PC with house-developed software, COMET Version 5.0. The COMET software communicates with ChemStation (Version 9.1, Agilent Technologies, Palo Alto, CA) and automates the HPLC running sequence, simultaneous DAD/MS detection and subsequent data processing, including peak tracking.

As shown in Fig. 2, the COMET software is comprised of four standalone modules: (A) System Control, (B) Data Extraction, (C) MS Interpretation and (D) Report Generation. Module A provides the user interface to schedule and initiate the analysis. The second module, "Data Extraction", is a ChemStation post-run macro, which preprocesses the chromatographic data and exports the chromatograms, peak tables and spectral information of interests to COMET data structure.

The "MS Interpretation" module is a Visual Basic (VB) executable routine, which scrutinizes the extracted mass spectral data and deducts the molecular weights for each detected component. The processes by which the program determines the molecular weights are discussed in the following Section 3. Mass interpretation parameters shall be adjusted in system configuration as needed to maximize the interpretation accuracy. Once component molecular weight assignment is complete, the program normalizes the retention time of each chromatogram and generates peak tracking plots for individual components.

Finally, the "Report Generation" routine compiles all the results, identifies system related artifact peaks by correlating the sample chromatogram with that from the blank run, and generates Microsoft Word report. In order not to devastate the analysts with the overwhelming amount of information that the orthogonal LC/MS screening presents, the reports are prepared in hyperlinked three-layered Word documents: (a) nine-method summary with peak tracking; (b) LC/MS report for individual method; (c) mass spectra and extracted ion chromatograms (EIC). Depending on the level of detail needed, the hyperlinks guide the analyst to quickly jump to the specific layer of report.

The integration of the four modules casts the fully automated walk up COMET system capable of round-the-clock unattended impurity screening directly from submitted sample to printed reports.

3. Results and discussion

The orthogonality, i.e. the selectivity differences, of the nine COMET methods forms the basis for a comprehensive characterization of an impurity profile. Examples of COMET screening results for two Pfizer proprietary drug samples, PF#1 and PF#2 are shown in Figs. 3 and 4, respectively. Table 2 illustrates the method orthogonalities calculated using retention time correlation coefficients between any two chromatograms [39]. Many of the correlation coefficients are far less than 1, which indicates good orthogonalities. For M3 and M4, the r^2 is fairly high (~0.9), which is comprehensible since both methods are using the same column and similar eluent. However, the THF used in M4 does give very unique selectivity for some impurities such as impurity PF#1-B, although it is less significant in the PF#2 screening. The method orthogonality is also reflected in the chromatographic profiles. As shown in Fig. 3, four major peaks, including API (P), were detected for PF#1 in each method. The stacked chromatograms clearly show the significant change of elu-

Table 2
Orthogonality (r^2) evaluation of COMET methods

	M1	M2	M3	M4	M5	M6	M8
M1		0.031	0.033	0.037	0.020	0.536	0.533
M2			0.489	0.676	0.675	0.152	0.237
M3				0.896	0.028	0.152	0.054
M4					0.175	0.212	0.119
M5						0.301	0.599
M6							0.821
M8							

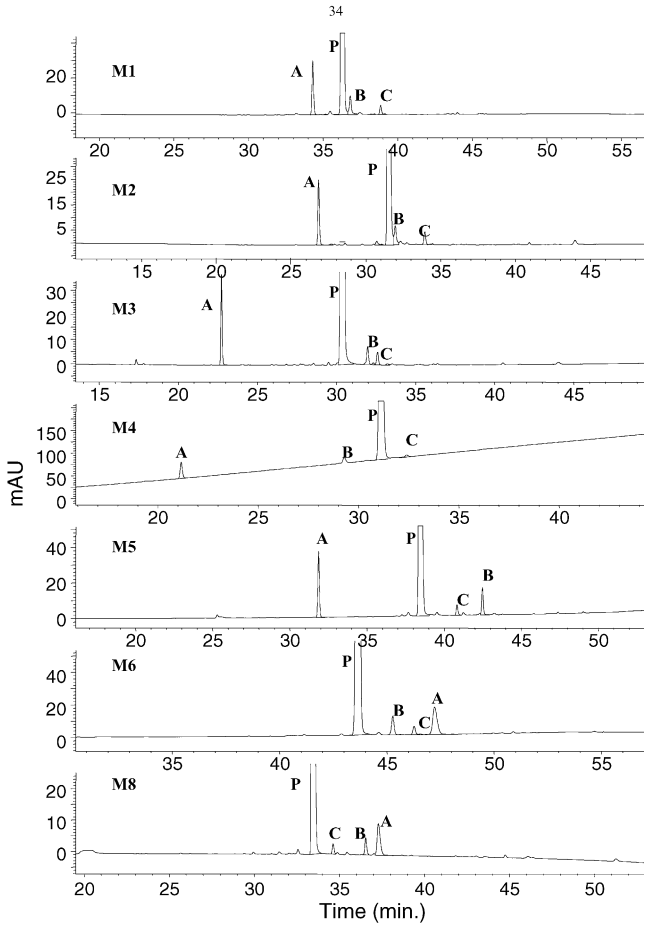

Fig. 3. COMET separations of impurity drug sample PF#1 (methods as listed in Table 1, UV detection at 254 nm). P denotes the API, while A, B, C stand for detected three major impurities, respectively.

tion order from method to method. The API peak eluted next to impurity PF#1-A and PF#1-B in M4, while it became the first in M6 and M8. Between M1 and M8, the elution orders of impurity PF#1-A, PF#1-B and PF#1-C were completely reversed. Accurate peak tracking is thus essential for the compilation of the complementary chromatographic information in order to illustrate the comprehensive impurity profile.

3.1. Peak tracking with DAD

For simple chromatograms as shown in Fig. 3, normalized UV peak areas probably would be sufficient to track each peak over different methods, since all peaks are well resolved and distinctive in peak area. Unfortunately, such nearly perfect separations are rare in generic screening. For most drug samples, such as PF#2 shown in Fig. 4, the complexity of the chromatograms makes peak tracking based solely on peak areas fairly unreliable. Coelution is obliviously the biggest hurdle, where fuzzy theory is required to "guess" the possible peak overlap [20,21]. At acidic pH, partially positive charged impurities tend to adsorb to the residual silanol groups on the stationary phase, causing severe tailing as shown in Fig. 4—M2, which further complicates accurate peak identification. Therefore, spectral information is truly vital for the unambiguous identification and tracking of chromatographic peaks.

We first create a customized UV spectra library from spectra inputs of the first HPLC method separation (M1). Each peak in the rest of the HPLC chromatograms is then searched against the local library using Agilent spectral match factor analysis software. The software performs a linear regression of the two spectra and multiplies the square of regression coefficient by 1000 to give the match factor. If the match factor of the best hit is below a cutoff limit (Agilent literature suggested 950), no hit will be reported and a new spectrum is added into the library. We applied this library search approach to the PF#2 COMET screening data. Unfortunately, the tracking result is discouraging. Out of the total of 40 peaks under evaluation, only 20 peaks are identified correctly. A closer examination of the data showed three major causes for the errors. First of all, there are two pairs of impurities in the mixture that have very similar spectra: PF#2-B and PF#2-E, as well as PF#2-D and PF#2-F (Fig. 5). The mislabeling of these two compound pairs accounts for eight incidents of peak tracking error. Secondly, the library search failed to identify the seven coelutions observed. Finally, five impurities (B–F) showed very different spectra in the pH 1.9 TFA solution (M1) compared with those acquired in the other mobile phases, mostly due to the low UV cutoff of the mobile phase and pH related chromophore shift. Also, worth mentioning is that the PF#2 sample used in this screening is spiked with about 0.2 mg/mL impurity standards. In routine drug samples, impurity levels are generally below 0.5% with respect to API. More significant interference from mobile phase background will further compromise the quality of spectra and thereby peak tracking. Fig. 5(e) shows the spectrum of impurity PF#2-F acquired in unspiked PF#2 drug sample at which its impurity level is about 0.47%. Obvious bias could be clearly observed compared with its more accurate acquired spectrum at high concentration (Fig. 5(d)).

3.2. Manual peak tracking with mass spectra

In this COMET system, an Agilent MSD mass spectrometer equipped with electro-spray ionizer (ESI) is available to acquire MS during the analysis. Fig. 6 lists one of the HPLC chromatograms from the COMET screening of drug sample PF#2 and the acquired mass spectra for each peak detected. The interpretation of the mass spectra can be addressed by the highly trained specialist, who would review the spectra and unambiguously assign molecular weights to each detected peak as shown in Fig. 6. Four of the five peaks are single component peaks, with the molecular weights identified as 465, 447,463, and 447, respectively. The second peak is proved to be a coelution of two impurities with molecular weights of 261 and 463, respectively. Peak tracking, therefore, gets no more complicated than assigning annotations for each molecular weight, for example, give each compound an alphabetic

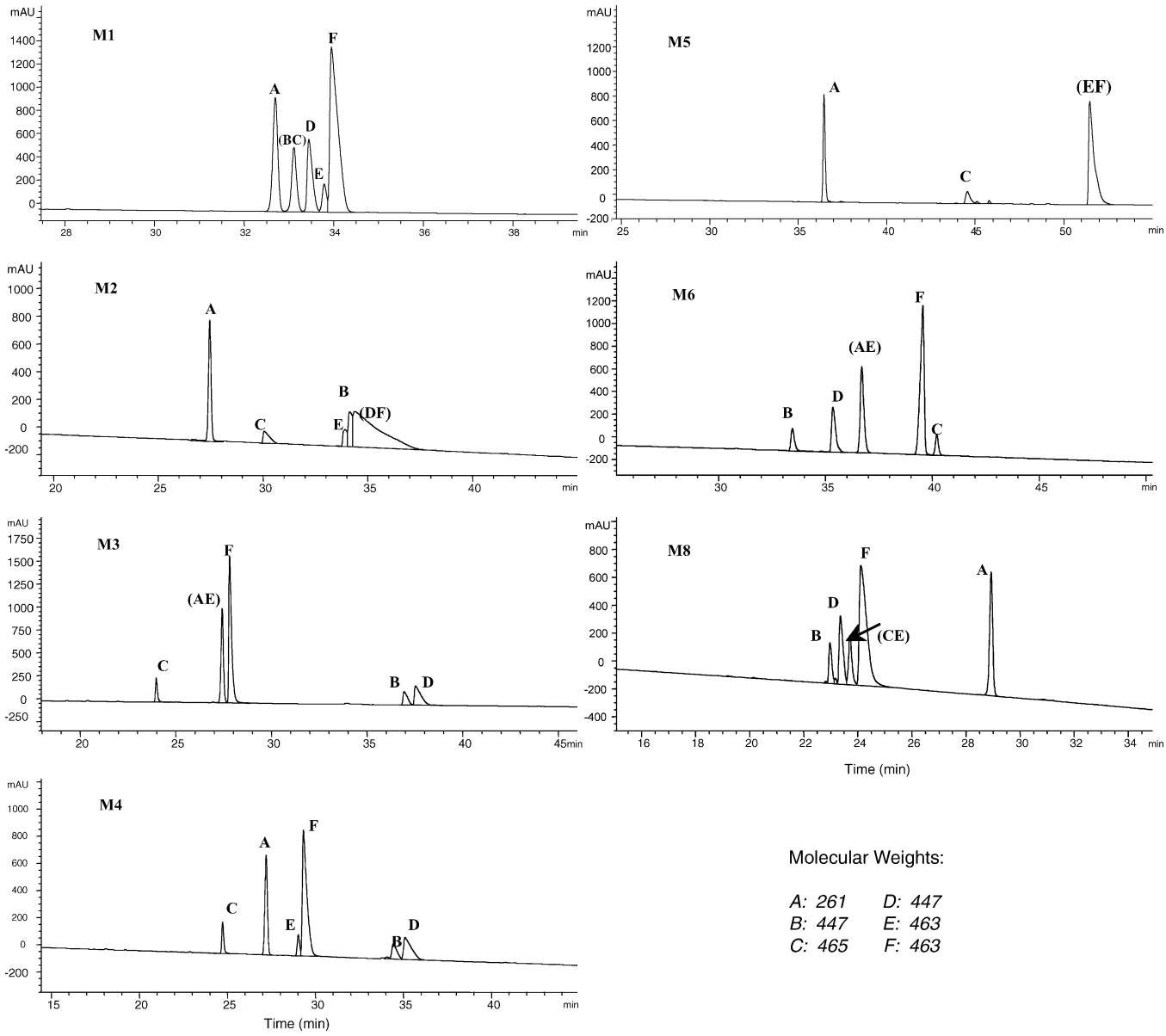

Fig. 4. COMET separations of spiked drug sample PF#2 (methods as listed in Table 1, UV detection at 220 nm). Co-eluting analytes are indicated by parentheses. The determined molecular weights for each analyte are: (A) 261 Da; (B) 447 Da; (C) 465 Da; (D) 447 Da; (E) 463 Da; (F) 463 Da.

designation starting with "A". One limitation, though, is that single-quadruple MS alone would not be able to differentiate isomers, such as the two 447 peaks in the PF#2 sample. For such cases, additional UV peak area ratio would be helpful for further identification within the isomer pairs. Fig. 4 illustrates the final peak tracking for PF#2 COMET screening results. Despite the combined multiple differences in elution order, peak shape, peak height, and even the number of peaks observed in varied LC methods, the assigned MS molecular weights provides accurate peak identification and unequivocal peak tracking for every compound.

Generally referred as a "soft" ionization technique due to its minimal internal energy deposition, ESI usually produces fairly simple mass spectra with limited fragmentation, which is dominated by the quasi-molecular ions. However, due to the generic screening nature of COMET system, it is impractical to optimize the ionization conditions for each compound. Consequently, many highly functionalized drug molecules often lead to various unconventional and unexpected behaviors when analyzed by ESI. Alkali metal adducts ($[M + Na]^+$ and $[M + K]^+$), solvation adducts ($[M + H + ACN]^+$), dimers ($[2M + H]^+$ and $[2M + Na]^+$) and doubly protonated ions ($[M + 2H]^{2+}$) are just some of the examples that complicate the mass spectra. Co-eluting compounds, which give overlapped spectra, further increase the complexity of the MS data. Even for experienced mass spectroscopists, the spectra interpretation can be time consuming. Thus, manual peak tracking via MS becomes the bottleneck

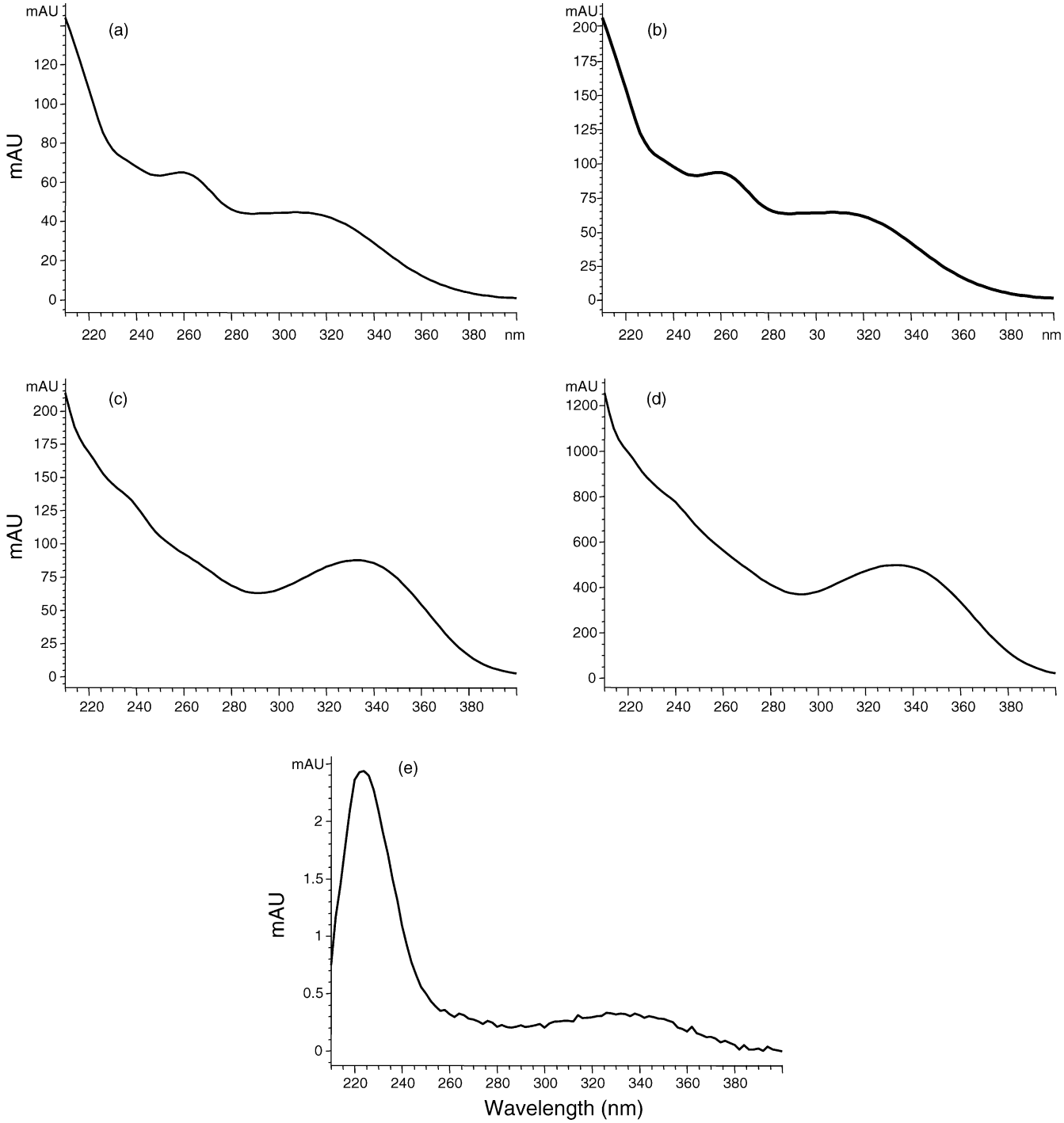

Fig. 5. UV spectra of: (a) PF#2-B; (b) PF#2-E; (c) PF#2-D; (d) PF#2-F; (e) PF#2-F. Spectra (a)–(d) were acquired from the spiked PF#2 sample shown in Fig. 4. Spectrum (e) was acquired from unspiked PF#2 drug sample solution, in which its impurity level is about 0.47%.

of the comprehensive impurity profiling. The running time for COMET screen of single drug sample is roughly 1.5 days, with no manual intervention needed. But it usually takes the analyst more than 5 days to examine all the MS data before concluding the comprehensive impurity content. Automation in the MS peak tracking process obviously becomes critical for the success of comprehensive and timely impurity profiling.

3.3. Automated peak tracking with mass spectra

3.3.1. Molecular weight determination

The key of automated MS peak tracking is to emulate the expert mass spectroscopists' knowledge-based activities of spectral interpretation. Some early efforts incorporated decision-making criteria into mass spectrometric analysis for protein structure determinations and peptide sequencing

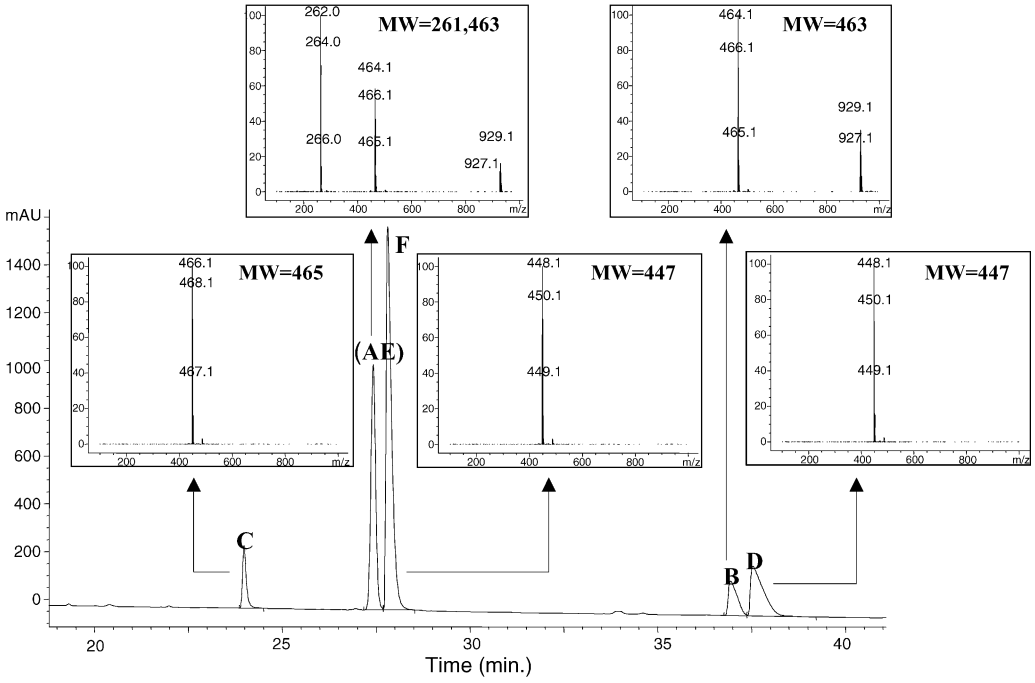

Fig. 6. Mass spectra and manual molecular weight assignments for M3 separation of PF#2 in COMET screening.

[40,41], characterization of polymeric molecular weight distributions [42], and classification of non-peptidic combinatorial compounds [43,44]. Huang et al. [45] and Tong et al. [46] even demonstrated complete automation package for data acquisition, processing, interpreting, and e-mailing mass spectrometry data. However, many of these approaches require a list of expected product molecular weights, which is not readily available during impurity profiling. More recently, Williams et al. reported an automated molecular weight assignment approach by evaluating all acquired mass ions and identifying the quasi-molecular ion [47]. A computer program was developed, which applied a set of simple quasi-molecular ion recognition rules to categorize and reject various interfering mass ions. In a separate work, Gorlach and Richmond demonstrated another interesting quasi-molecular ion discovery algorithm for adduct abundant combinatorial samples by adduct stencil search [48].

Both Williams et al. and Gorlach and Richmond's work provide ground for our automation of MS peak tracking. As aforementioned, in ESI mass spectra, the quasi-molecular ions are obscured mainly by five categories of interferences: (1) coelution (including noise spikes and eluent ions); (2) isotopic ions; (3) adduct ions; (4) dimers or doubly protonated ions; and (5) fragment ions. The fragment ions can be easily differentiated from the quasi-molecular ions, since their mass to charge ratios (m/z) are always less than that of their quasi-molecular ions. To identify the rest interference ions, four mass ion recognition tests, as illustrated in Table 3, get sequentially applied to all mass ions acquired in each mass spectrum, each eliminating one category of remaining interference ions.

3.3.1.1. Extracted Ion Chromatogram (EIC) test. The EIC test is to probe potential coeluting compounds as well as eliminate the noise spikes and background ions from the mobile phase. For the mass spectrum acquired for each chromatographic peak at retention time t_r, all major ions (relative abundance >10% that of the top ion in the spectrum) are pooled as quasi-molecular ion candidates. The EIC (also called single ion monitoring or SIM) of each ion is extracted and integrated. Noise spikes generally exhibit very sharp peak (typically 1 scan wide), which is less than

Table 3
Mass ion tests for automated spectra interpretation

Validation tests	Functionality	Rejected ions
1. EIC test	Integrate extracted ion chromatograms (EIC)	Noise spikes
	Compare peak retention time	Solvent ions
	Detect co-eluters	
2. Isotope test	Identify isotope ions	A + 1, A + 2 ions
3. Adduct test	Search for possible adducts	Common adduct ions: Na^+, K^+, NH_4^+, ACN
4. Dimer test	Check for "$(m/z)' = (m/z + 1)/2$" relationship	$[2M + H]^+$; $[M + 2H]^{2+}$
	Evaluate [A + 1]/[A] abundance ratio	

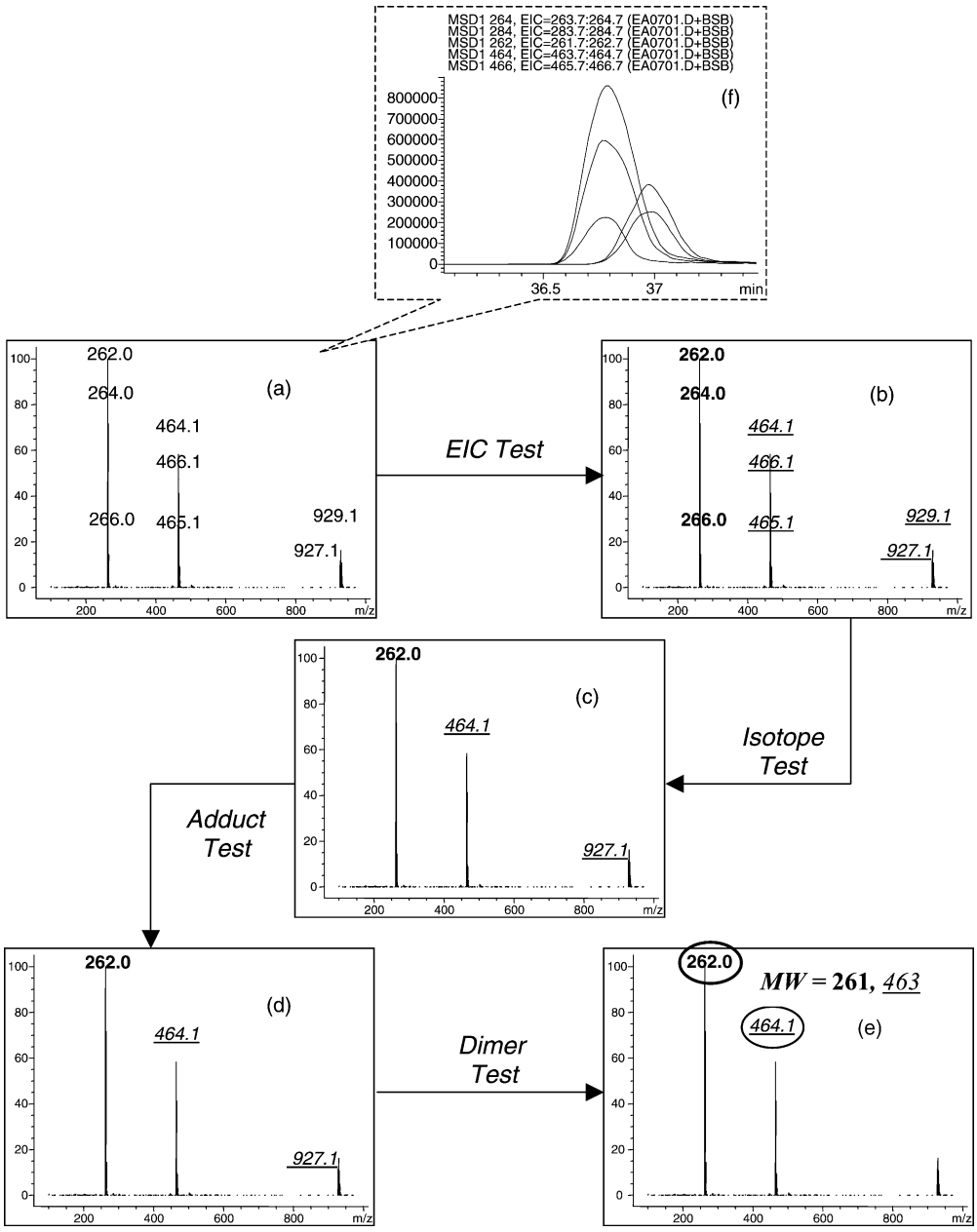

Fig. 7. Automated mass spectra interpretation. (a) Mass spectrum of Peak 2 in M3 separation of PF#2; (b) EIC test separates mass ions into two groups (bold and underlined italic) based on the slight difference in EIC peak retention time—each group indicates ions from one analyte. (c) Isotope test identifies and rejects A + 1, A + 2 isotope ions such as 264^+, 266^+, 465^+, 466^+, 929^+ ions. (d) Adduct test searches for and eliminates common adduct ions including Na^+, K^+, NH_4^+, ACN. No adduct ions are found in this example. (e) Dimer test differentiates proton-bound dimers vs. doubly protonated ions and rejects determined invalid ions. The survivor ions in each group after the Dimer test determine the molecular weights of the two coeluting analytes: A, 261 Da, E, 463 Da. (f) Overlaid EICs of major mass ions listed in (a).

the minimum peak width for integration, while the eluent ions typically show only slow signal drift over time. Therefore, no peak will be detected for their EIC at time t_r. For the rest of the ions, an integrated peak near time t_r is expected.

Co-eluting analytes, if any, usually exhibit slight difference in peak retention time such as the example shown in Fig. 7(f). For cases like this, the ions need to be clustered into groups based on the retention time, as shown in Fig. 7(b) the bold annotated ion group versus the italic underlined ion group. Each group, from then on, should be treated as independent mass spectrum and one quasi-molecular ion is expected from each group.

3.3.1.2. Isotope test. The search for isotope peaks is fairly straightforward since the natural isotope distribution has been so well studied. Basically, the *Isotope test* identifies all the A + 1 and A + 2 ions in each group, eliminates them from the ion pool, and passes the rest to the *Adduct test* as shown in Fig. 7(c).

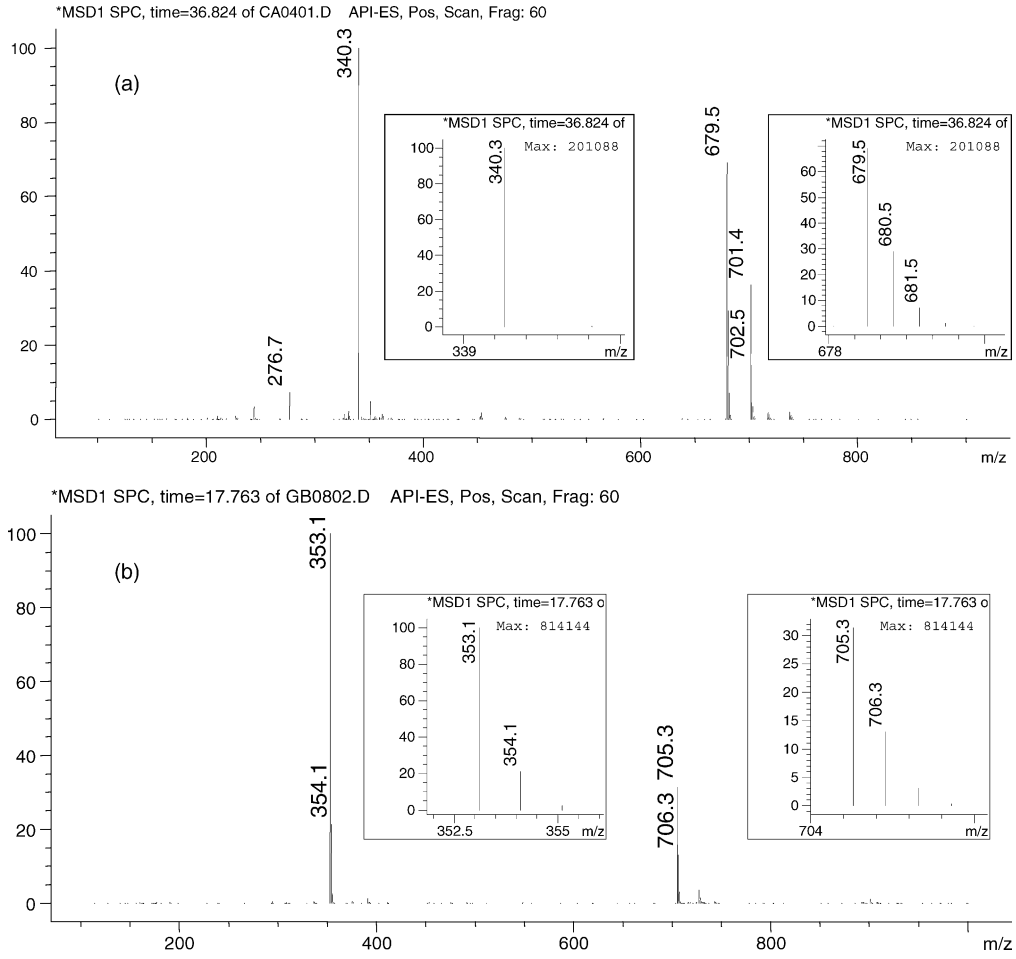

Fig. 8. Illustration of Na^+ adduct, doubly protonated ion and proton-bound dimer recognition. (a) The mass ion at *m*/*z* 701 is determined to be $[M + Na]^+$ adduct of *m*/*z* 679 ion. The mass ion at *m*/*z* 340 is determined to be $[M + 2H]^{2+}$ of the signal at *m*/*z* 679. (b) Signal at *m*/*z* 705 is determined to be proton bound dimer while *m*/*z* 353 being the $[M + H]^+$.

3.3.1.3. Adduct test. In COMET screening, the likely adduct ions are fairly easy to predict since all methods are predefined. In this development, we specifically look for four common adducts: $[M + Na]^+$, $[M + K]^+$, $[M + NH_4]^+$ and $[M + H + ACN]^+$, in which the moveable adduct stencil with *m*/*z* differences of 22, 38, 17, and 41 scans for potential adduct ions [49]. In Fig. 7(d), the *Adduct test* determines that no adducts exist in the spectrum, thus all ions are passed to the next test. Meanwhile, in the spectrum shown in Fig. 8(a), mass ions at *m*/*z* of 701^+ and 679^+ show mass difference of 22, indicating 701^+ being a Na^+ adduct ion and gets eliminated.

3.3.1.4. Dimer test. An interesting situation occurs while performing the *Dimer tests*: when a doubly charged ion ($[M + 2H]^{2+}$) is present, the $[M + H]^+$ and $[2M + H]^+$ ion pair and the $[M + 2H]^{2+}$ and $[M + H]^+$ pair share the same apparent *m*/*z* relationship:

$$(m/z)' = (m/z + 1)/2$$

However, since quasi-molecular ions $[M + H]^+$ have half the number of carbons as the dimer ions $[2M + H]^+$, their relative A + 1 abundance should also have the 1:2 relationship. For example, as shown in Fig. 8(b), the A + 1 (354^+) abundance with respect to 353^+ is about 0.21, while the abundance of 706^+ with respect to 705^+ ion is about 0.41.

$$\frac{[354^+]/[353^+]}{[706^+]/[705^+]} = \frac{0.21}{0.41} = 0.51$$

On the other hand, the relative A + 1 abundance ratios for the $[M + 2H]^{2+}$ and $[M + H]^+$ pair are typically much less than 0.5 or much greater than 0.5 if the molecule includes Cl or Br atoms. As shown in Fig. 8(a), the A + 1 abundance for $680^+/679^+$ is roughly 0.43, while no 341^+ ion is detected:

$$\frac{[341^+]/[340^+]}{[680^+]/[679^+]} = \frac{0.00}{0.43} = 0.00$$

Therefore, 705^+ and 340^+ are identified as proton-bound dimer and doubly charged ion respectively. For the compound in Fig. 7(d), the 927^+ is determined to as the proton-bound dimer of 464^+.

Once all of the four criteria have been applied to the test spectrum, the only survival ions should be the quasi-molecule

Table 4
Impurity profile and retention times of drug sample PF#2

Peak	MW	Area[a] %	Retention time (min)								
			M1[b]	M2	M3	M4	M5	M6	M7	M8	M9
A	261	15.52	32.82	27.58	27.55	27.32	36.58	36.83	19.51	29.04	40.35
B	447	6.47	33.24	34.27	37.09	34.56		34.95	15.79	23.10	32.59
C	465	4.77	33.24	30.22	24.11	24.84	44.70	40.32	15.94	24.25	34.01
D	447	15.55	33.57	35.29	37.67	35.29		35.35	15.79	23.50	33.34
E	463	3.86	33.79	34.56	27.95	29.44	51.62	39.66	15.94	24.25	33.70
F	463	52.27	34.10	34.27	27.55	29.14	51.62	36.83	16.21	23.84	34.58

[a] Peak area percentage calculated with UV absorbance at 220 nm.
[b] Methods M1–M9 as described in Table 1.

ions and their fragments (if any). For such cases, the signal with the highest mass in each ion group is assigned as the quasi-molecular ion (262^+ and 464^+ in Fig. 7(e)), from which the molecular weight is deduced (MW = 261 and 463).

3.3.2. Automated peak tracking results

The successful automated ESI-MS interpretation unambiguously identifies each chromatographic peak by its molecular weights (isomers need additional peak area information). As shown in Table 4, a total of six chemical entities are identified in PF#2 and their retention times in each orthogonal method are listed. Co-eluters, such as PF#2-B and C in M1, exhibit the same retention time. Compound PF#2-B and D are not detected in M5 because they are strongly retained on the column and do not elute within the 60 min gradient. However, since they are captured in many other methods, no critical impurity content information is missing, which represents one of the most appealing advantages of the orthogonal system.

Fig. 9 is the graphical representation of the peak tracking results included in the automatically generated reports. Each axis in the radar plot represents one COMET HPLC method. The scale of the axis stands for the normalized retention time for each method with the latest eluting peak being 1. The peak tracking is illustrated by connecting the retention times of each analyte across the radar, which forms a loop if the analyte elutes in every method. Any crossover represents a retention order change between the two neighboring methods. For example, compound PF#2-E elutes first in M5, while it elutes next to compound PF#2-C in M4.

The number of crossovers, to some extent, is indicative of the method orthogonality. The three methods on the left side (M7–M9) have very little orthogonality ($r^2 > 0.9$), which is reflected in the similar retention pattern. The cross-overs between M4 and M5 illustrate more orthogonality ($r^2 = 0.175$).

On the other hand, the shape of these retention loops signifies the trend of the retention with respect to the changes in stationary hydrophobicity, eluent pH and solvent strength. For PF#2, one of the most noticeable trends is that most analytes tend to elute closer in more acidic eluent (M1 and M7–M9). Besides, all analytes get more retained in higher pH eluent, which is consistent with the properties of basic drug compounds. At the extreme case, compound PF#2-B and PF#2-D did not elute in M5, in which the eluent is at pH 7.0 and a weak organic solvent MeOH is used. For PF#1, significant retention change is observed for impurity PF#1-A when changing from C18 columns to more hydrophilic columns such as phenyl-hexyl or C4, while the THF used in M4 seems to have dramatic impact on the retention of impurity PF#1-B specifically (Table 5). Therefore, besides unveiling the comprehensive impurity profile, the successful

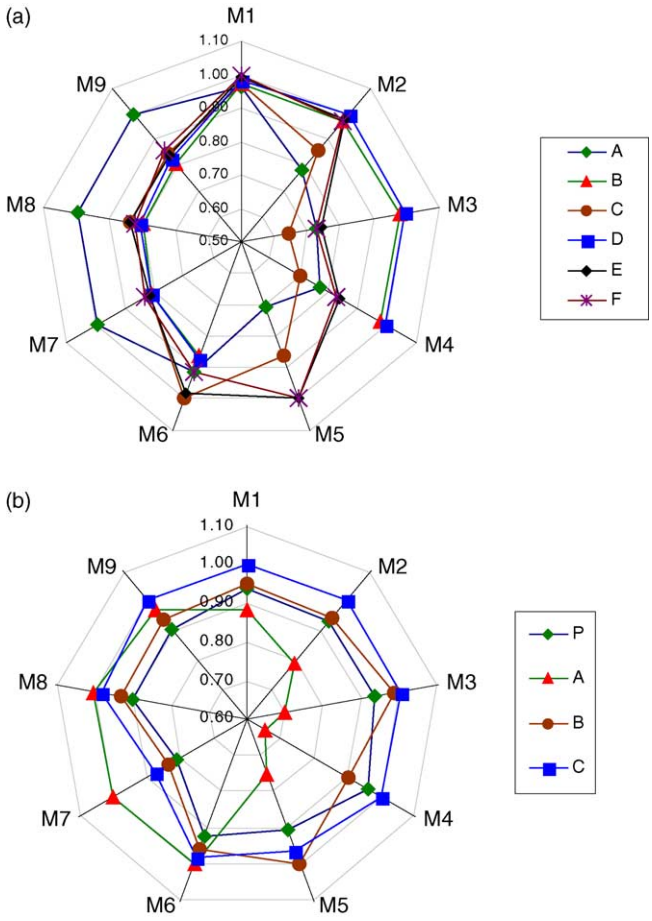

Fig. 9. Retention plots for COMET screening peak tracking. Each radar axis denotes one COMET HPLC method as described in Table 1. The scale on the axis is the normalized retention time with the latest eluting peak being 1. Drug samples: (a) PF#2 and (b) PF#1.

Table 5
Impurity profile and retention times of drug sample PF#1

Peak	MW	Area[a] %	Retention time (min)								
			M1[b]	M2	M3	M4	M5	M6	M7	M8	M9
P	504	95.76	36.30	31.53	30.41	31.11	38.44	43.64	21.03	33.44	49.08
A	522	1.71	34.31	26.84	22.74	21.17	31.86	47.22	25.90	37.30	52.41
B	478	0.70	36.82	31.96	31.96	29.30	42.50	45.27	21.68	34.64	50.97
C	582	0.33	38.86	33.93	32.59	32.40	40.84	46.27	22.57	36.55	54.34

[a] Peak area percentage calculated with UV absorbance at 254 nm.
[b] Methods M1–M9 as described in Table 1.

peak tracking also provides valuable hints on method optimization.

Since system deployment, a total of 56 drug samples submitted by chemists from various departments have been screened by the COMET system. The samples varied in complexity and composition, including both impurity and degradation samples. For the 56 samples, which contained more than 500 chemical entities, COMET provided successful peak tracking for 80% of the compounds. The most prominent failure, representing 18% of the compounds, was the insufficient MS signal for accurate molecular weight determination. A minimum signal to noise ratio of 5 in EIC was determined for reliable peak integration. For these MS inert compounds, more sophisticated chemometrics processing of spectral data (UV or MS) is needed for more successful peak tracking, e.g. component data analysis [50,51]. Another solution is to apply more efficient ionization technique or more sensitive mass detection scheme such as MS/MS. But the latter is more difficult for generic screening due to the compound versatility. The remainder of the failures was attributed to: (1) mislabeling solvent ion; and (2) misinterpreting quasi-molecular ion as dimer. Case (1) failure was due to less than perfect peak integration of EICs. Occasionally, ChemStation software assigned peaks to solvent ion EIC profile by mistake. Case (2) was very rare. It only happened when the MS signal was very weak and the noise at the *m*/*z* of A + 1 ion of doubly protonated ion happened to show the 2:1 relative A + 1 abundance ratio as the dimer/quasi-molecular ions relationship. Similarly, appropriate data preprocessing by chemometrics can be useful to reduce these failures.

4. Conclusions

A fully automated comprehensive impurity profiling system employing orthogonal HPLC separations and hyphenated UV–MS detections is developed. The system is capable of tracking each impurity peak over all chromatograms of a drug sample recorded under different chromatographic conditions by automatically assigned molecular weights. The process is based on automated ESI-MS interpretation, which utilizes a set of interpretive rules to systematically eliminate the common interfering mass ions. It is proved to be more accurate and reliable than the preceding UV-based peak tracking approaches [24–26]. Furthermore, it endows the project analyst with profound MS knowledge yet eliminates the tedious spectral interpretation. Initial testing of the automated peak tracking yields 80% success rate for over 500 drug impurities. The observed failures are mostly due to the low ionization efficiency of some impurities. Further work using chemometrics processing and deconvolution of the mass spectral data is on-going and will provide additional improvements. The availability of ESI/APCI dual mode ionization will also greatly widen the range of compounds could be detected with good sensitivity [52]. The peak tracking method described in this paper is not limited to tracking drug impurities. The method should be generally applicable to any LC/MS or capillary electrophoresis (CE)/MS separations of organic molecules. The peak tracking procedure should also be very useful in the development of new HPLC optimization protocols [22,33,53–55], providing a basis for improved strategies for the monitoring, purification, and analysis of biological macromolecules, such as peptide digestion [56,57].

Acknowledgements

The authors acknowledge Mark Niermeyer from Intekra LLC for his support of software programming for this work.

References

[1] T.W. Ryan, Anal. Lett. 31 (1998) 2447.
[2] S. Gorog, M. Babjak, G. Balogh, J. Brlik, A. Csehi, F. Dravecz, M. Gazdag, P. Horvath, A. Lauko, K. Varga, Talanta 44 (1997) 1517.
[3] G.J. Opiteck, S.M. Ramirez, J.W. Jorgenson, M.A. Moseley III, Anal. Biochem. 258 (1998) 349.
[4] G.J. Opiteck, J.W. Jorgenson, M.A. Moseley III, R.J. Anderegg, J. Microcolumn Sep. 10 (1998) 365.
[5] K.K. Unger, K. Racaityte, K. Wagner, T. Miliotis, L.E. Edholm, R. Bischoff, G. Marko-Varga, J. High Resolut. Chromatogr. 23 (2000) 259.
[6] K. Wagner, K. Racaityte, K.K. Unger, T. Miliotis, L.E. Edholm, R. Bischoff, G. Marko-Varga, J. Chromatogr. A 893 (2000) 293.
[7] K. Wagner, T. Miliotis, G. Marko-Varga, R. Bischoff, K.K. Unger, Anal. Chem. 74 (2002) 809.
[8] S. Gorog, G. Balogh, A. Csehi, E. Csizer, M. Gazdag, Z. Halmos, B. Hegedus, B. Herenyi, P. Horvath, A. Lauko, J. Pharm. Biomed. Anal. 11 (1993) 1219.
[9] D. Louden, A. Handley, R. Lafont, S. Taylor, I. Sinclair, E. Lenz, T. Orton, I.D. Wilson, Anal. Chem. 74 (2002) 288.
[10] E. Lenz, S. Taylor, C. Collins, I.D. Wilson, D. Louden, A. Handley, J. Pharm. Biomed. Anal. 27 (2002) 191.

[11] D. Louden, A. Handley, S. Taylor, E. Lenz, S. Miller, I.D. Wilson, A. Sage, R. Lafont, J. Chromatogr. A 910 (2001) 237.
[12] D. Louden, A. Handley, S. Taylor, E. Lenz, S. Miller, I.D. Wilson, A. Sage, Anal. Chem. 72 (2000) 3922.
[13] D. Louden, A. Handley, S. Taylor, E. Lenz, S. Miller, I.D. Wilson, A. Sage, Analyst 125 (2000) 927.
[14] M. Ludlow, D. Louden, A. Handley, S. Taylor, B. Wright, I.D. Wilson, J. Chromatogr. A 857 (1999) 89.
[15] P. Schoenmakers, Optimization of Chromatographic Selectivity: A Guide to Method Development, 1986.
[16] G. Szepesi (Ed.), HPLC in Pharmaceutical Analysis: General Considerations, vol. 1, 1990.
[17] I. Molnar, R. Boysen, P. Jekow, J. Chromatogr. 485 (1989) 569.
[18] A.C.J.H. Drouen, H.A.H. Billiet, L. De Galan, Anal. Chem. 56 (1984) 971.
[19] A.C.J.H. Drouen, H.A.H. Billiet, L. De Galan, Anal. Chem. 57 (1985) 962.
[20] M. Otto, W. Wegscheider, E.P. Lankmayr, Anal. Chem. 60 (1988) 517.
[21] E.P. Lankmayr, W. Wegscheider, J. Daniel-Ivad, I. Kolossvary, G. Csonka, M. Otto, J. Chromatogr. 485 (1989) 557.
[22] N.M. Djordjevic, F. Erni, B. Schreiber, E.P. Lankmayr, W. Wegscheider, L. Jaufmann, J. Chromatogr. 550 (1991) 27.
[23] A.C. Gilby, M.J. Leveille, in: Proceedings of SPIE—The International Society for Optical Engineering, vol. 1681, 1992, p. 54.
[24] D.W. Hill, T.R. Kelley, K.J. Langner, Anal. Chem. 59 (1987) 350.
[25] E.C. Nicolas, T.H. Scholz, J. Pharm. Biomed. Anal. 16 (1998) 813.
[26] M.E. Swartz, P.R. Brown, Chirality 8 (1996) 67.
[27] A.J. Round, M.I. Aguilar, M.T. Hearn, J. Chromatogr. A 661 (1994) 61.
[28] J.K. Strasters, H.A.H. Billiet, L. De Galan, B.G.M. Vandeginste, J. Chromatogr. 499 (1990) 499.
[29] J.K. Strasters, H.A.H. Billiet, L. De Galan, B.G.M. Vandeginste, G. Kateman, J. Chromatogr. 385 (1987) 181.
[30] J.K. Strasters, H.A.H. Billiet, L. De Galan, B.G.M. Vandeginste, G. Kateman, Anal. Chem. 60 (1988) 2745.
[31] J.K. Strasters, H.A.H. Billiet, L. De Galan, B.G.M. Vandeginste, G. Kateman, J. Liquid Chromatogr. 12 (1989) 3.
[32] J.K. Strasters, F. Coolsaet, A. Bartha, H.A.H. Billiet, L. De Galan, J. Chromatogr. 499 (1990) 523.
[33] P.V. van Zomeren, A. Hoogvorst, P.M.J. Coenegracht, G.J. de Jong, Analyst (Cambridge, UK) 129 (2004) 241.
[34] J. Ermer, J. Pharm. Biomed. Anal. 18 (1998) 707.
[35] A. Thomasberger, F. Engel, K. Feige, J. Chromatogr. A 854 (1999) 13.
[36] M. Babjak, G. Balogh, M. Gazdag, S. Gorog, J. Pharm. Biomed. Anal. 29 (2002) 1153.
[37] O.M. Denk, G.G. Skellern, D.G. Watson, J. Pharm. Pharmacol. 54 (2002) 87.
[38] J. Ermer, M. Vogel, Biomed. Chromatogr. 14 (2000) 373.
[39] Z. Liu, D.G. Patterson Jr., M.L. Lee, Anal. Chem. 67 (1995) 3840.
[40] J.K. Eng, A.L. McCormack, J.R. Yates III, J. Am. Soc. Mass Spectrom. 5 (1994) 976.
[41] J. Fernandez-De-Cossio, J. Gonzalez, Y. Satomi, T. Shima, N. Okumura, V. Besada, L. Betancourt, G. Padron, Y. Shimonishi, T. Takao, Electrophoresis 21 (2000) 1694.
[42] J.P. Kurvinen, P. Rua, O. Sjovall, H. Kallio, Rapid Commun. Mass Spectrom. 15 (20011084).
[43] K. Klagkou, F. Pullen, M. Harrison, A. Organ, A. Firth, G.J. Langley, Rapid Commun. Mass Spectrom. 17 (2003) 2373.
[44] K. Klagkou, F. Pullen, M. Harrison, A. Organ, A. Firth, G.J. Langley, Rapid Commun. Mass Spectrom. 17 (2003) 1163.
[45] N. Huang, M.M. Siegel, H. Muenster, K. Weissenberg, J. Am. Soc. Mass Spectrom. 10 (1999) 1212.
[46] H. Tong, D. Bell, K. Tabei, M.M. Siegel, J. Am. Soc. Mass Spectrom. 10 (1999) 1174.
[47] J.D. Williams, B.E. Weiner, J.R. Ormand, J. Brunner, A.D. Thornquest Jr., D.J. Burinsky, Rapid Commun. Mass Spectrom. 15 (2001) 2446.
[48] E. Goerlach, R. Richmond, Anal. Chem. 71 (1999) 5557.
[49] E. Goerlach, R. Richmond, I. Lewis, Anal. Chem. 70 (1998) 3227.
[50] W. Windig, J.M. Phalp, A.W. Payne, Anal. Chem. 68 (1996) 3602.
[51] W. Windig, W.F. Smith, W.F. Nichols, Anal. Chim. Acta 446 (2001) 467.
[52] Waters, Waters Brochure Library# 720000104EN, 2002.
[53] A.M. Siouffi, R. Phan-Tan-Luu, J. Chromatogr. A 892 (2000) 75.
[54] P.J. Schoenmakers, N. Mackie, R.M. Lopes Marques, Chromatographia 35 (1993) 18.
[55] P.B. Bowman, J.G.D. Marr, D.J. Salvat, B.E. Thompson, J. Pharm. Biomed. Anal. 11 (1993) 1303.
[56] S.H. Kang, X. Gong, E.S. Yeung, Anal. Chem. 72 (2000) 3014.
[57] Y. He, E.S. Yeung, J. Proteome Res. 1 (2002) 273.