Published in final edited form as:

Anal Chem. 2007 May 1; 79(9): 3355–3362. doi:10.1021/ac0622781.

Untargeted Analysis of Mass Spectrometry Data for Elucidation of Metabolites and Function of Enzymes

Raymundo Sanchez-Ponce and F. Peter Guengerich

Department of Biochemistry and Center in Molecular Toxicology, Vanderbilt University School of Medicine, Nashville, Tennessee 37232-0146

Abstract

A Matlab®-based computer program termed “Discovery of General Endo- and Xenobiotics” (DoGEX) was developed, which uses wavelets and morphological analysis to process liquid chromatography-mass spectrometry (LC-MS) data. The output of the program is a list of integration areas as a function of retention time and molecular mass. A feature of the computer program is spectral filtering to facilitate the detection of chromatographic peaks with a particular isotopic ratio. The program DoGEX was used to automatically select oxidation products formed from felodipine (i.e. two chlorine atoms) and bromocriptine (one bromine atom) with cytochrome P450 3A4. The recognized isotope ratio can be changed to permit a natural or artificial mixture of isotopes to be monitored for selections. This computer program can be used to analyze LC-MS data for untargeted metabolic profiling experiments, e.g. to assign endogenous functions to newly characterized cytochrome P450 enzymes. In a representative example, an incubation of testosterone, NADPH, and a 1:1 $^{16}\text{O}_2$: $^{18}\text{O}_2$ mixture yielded products with M and M+2 ions resembling bromine doublets. Another use of the program is the subtraction of one set of t_R , m/z data from another, e.g. in comparisons of changes in patterns during enzyme reactions.

INTRODUCTION

The elucidation of the functions of gene products is a major challenge in biology.¹ More is known about genomes than about the functions of their products, i.e. proteins. Historically, trial and error methods have been used to establish gene function, but new approaches are needed.² Due to its inherent selectivity and sensitivity, liquid chromatography (LC)-mass spectrometry (MS) can play a key role in the elucidation of enzyme functions. The use of LC-MS instrumentation for the study of biological samples produces a multidimensional data matrix that is a function of the retention time (t_R) and the mass spectra of components in the sample. This information can be used in metabolic studies to determine the function of an enzyme.³

In “targeted” metabolic profiling, the changes in the concentrations of one or several pre-selected compounds are evaluated across multiple samples. For example, an analysis may involve a comparison of a known compound or pathway of wild-type vs. “knock-out” transgenic organisms. In “untargeted” metabolic profiling, a typical research goal is to

Address correspondence to: Prof. F. Peter Guengerich Department of Biochemistry Vanderbilt University School of Medicine 638 Robinson Research Building 23rd and Pierce Avenues Nashville, Tennessee 37232-0146 Telephone: (615) 322-2261 FAX: (615) 322-3141 f.guengerich@vanderbilt.edu.

Supporting Information Available: The effect of the interpolation interval on chromatographic peak shape and data density, selection of the wavelet transformation algorithm, effect of varying t_R and two-dimensional LC-MS data overlapping, alignment of multiple single ion chromatograms using zero padding symlets and Daubechies wavelets with their ϕ (father) and ψ (mother) functions, and selection of the decomposition level to remove noise in LC-MS data is available as Supporting Information. Current ordering information is found on any masthead page.

discover chemical differences between transgenic vs. wild-type specimens.^{4,5} The vast amount of data generated by LC-MS instrumentation during untargeted metabolic profiling studies requires the use of bioinformatics tools to detect very small differences among samples.³ Most of the approaches that have been used to analyze data involve either subtraction methods^{4,6-8} or principal component analysis (PCA).^{9,10}

An application of LC-MS and bioinformatics in functional genomics is the discovery of cytochrome P450 (P450) functions in the metabolism of endobiotics (i.e. compounds normally found in the body). P450 enzymes are involved in the metabolism of a wide variety of substrates with very different chemical structures, including endobiotic and xenobiotics (e.g. drugs, carcinogens).¹¹ The most common reactions produced by P450s are hydroxylations and dealkylations. Of the 57 P450 genes in humans that potentially express P450 enzymes, roughly 13 have unknown function (in terms of both endobiotics and xenobiotics).^{12,13}

One strategy that can be used in the elucidation of function is isotopic labeling. In one approach, a ¹³C substrate is added and the subsequent labeling patterns are observed,¹⁴ but this is really a targeted approach. A more direct example of this type of procedure is the ³²S/³⁴S labeling of sulfated lipids of mycobacteria.¹⁵ In an application of this technique applied in our work here, a gas mixture containing a 1:1 ratio of ¹⁶O₂ and ¹⁸O₂ is incubated with a crude fraction from the human tissue plus a heterologously expressed P450, recombinant NADPH-P450 reductase, and an NADPH-generating system for the elucidation of potential P450 candidates that undergo primarily monooxygenation reactions; the products will have an isotopic signature dependent on the insertion of ¹⁶O₂ and ¹⁸O₂, as in the application of H₂¹⁸O/H₂¹⁶O strategies for peptide labeling.¹⁶ Each oxygenated metabolite will have an isotopic abundance (50% M, 50% M+2) independent of the molecular mass of the substrate. For selection of peaks with a desired ratio, a digital filter can be used to scan the data and match the similarity of the theoretical spectral shape of the doublets against the actual data.

Because most of the substrates of the known P450s are hydrophobic, a general reversed-phase LC-MS method with positive and negative ion electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI) can be used to analyze the samples obtained from incubations that use the isotopic oxygen mixture. With the large variety of LC-MS instrumentation available, a need exists for computer programs capable of handling data generated by different LC-MS platforms. We report software that can accommodate these functions.

EXPERIMENTAL SECTION

P450 Incubations and Analysis of Products

Bromocriptine, felodipine, and testosterone were selected as well-established P450 3A4 substrates.^{12,17,18} Testosterone (1.0 mM) or felodipine (50 μM) was incubated with “bicistronic” membranes containing recombinant (human) P450 3A4 and NADPH-P450 reductase (100 pmol each), obtained from *Escherichia coli* cells.¹⁹ All incubations were conducted for 30 min at 37 °C in 1.0 mL of 0.10 M phosphate buffer (pH 7.4). For the isotopic labeling experiments, a 1:1 (v/v) mixture of ¹⁶O₂/¹⁸O₂ (Cambridge Isotopes, Andover, MA) was used. The incubation mixtures, in modified 20-mL Thunberg tubes, were made partially anaerobic by three alternate cycles of vacuum and Ar purging^{20,21} and removed from the manifold under vacuum. The ¹⁸O₂/¹⁶O₂ mixture was then added from a premixed cylinder (Cambridge Isotopes), under pressure, via a short needle (a 2-stage valve was used on the cylinder). The NADPH-generating system included 100 μL of 100 mM glucose 6-phosphate, 50 μL of 10 mM NADP⁺, and 2 μL of yeast glucose 6-phosphate

dehydrogenase (1 mg mL^{-1})²² and was used to initiate reactions. The products and substrate were extracted with 1.0 mL of CH_2Cl_2 , and 0.50 mL of the organic phase was collected. The organic phase was evaporated to dryness under an N_2 stream, and the samples were dissolved in 100 μL of $\text{CH}_3\text{OH}/\text{H}_2\text{O}$ (1:4, v/v) for LC-MS analysis.

LC-MS data was generated using an LCQ-DecaXP ion trap mass spectrometer (ThermoFinnigan, San Jose, CA) equipped with an APCI ion source and an Agilent 1100 LC system (Agilent Technologies, Palo Alto, CA). The scan range was set from m/z 100 to 750 and from m/z 300 to 450 in the testosterone and felodipine experiments, respectively. The source current was set at 10 μA with a vaporizer temperature of 250 $^\circ\text{C}$. The capillary voltage was 46 V and its temperature was set at 250 $^\circ\text{C}$. The sheath gas flow was set to 67 and the auxiliary gas flow was set to 55 (arbitrary units). A YMC ODS-AQ octadecylsilyl LC column (2 mm $\times$ 250 mm, 5 μm) was used with a flow rate of 200 $\mu\text{L min}^{-1}$. The mobile phase A consisted of 95% H_2O , 5% CH_3CN (v/v), and 5 mM $\text{NH}_4\text{CH}_3\text{CO}_2$, and mobile phase B consisted of 5% H_2O , 95% CH_3CN (v/v), and 5 mM $\text{NH}_4\text{CH}_3\text{CO}_2$. The gradient used was 100% mobile phase A for 1 min, with a linear increase from 0 to 100% B over 3 min, followed by 100% B for 10 min. The column was allowed to equilibrate for an additional 10 min prior to the next injection.

Algorithm Development

Matlab 7.1® (Mathworks, Natick, MA), with image processing, wavelets, and bioinformatics toolboxes, was used to develop a computer program termed Discovery of General Endo- and Xenobiotics (DoGEX) (Figure 1). The program is available upon request.

The RAW files obtained in the LC-MS experiments were converted into an mzXML format using a Dell Dimension XPSGen5 computer (Dell, Round Rock, TX) with a dual processor.^{23,24} From this format the files can be transferred into TXT files using an mzXML decoder, in order to compress the data and eliminate extraneous information. A Macintosh computer (dual 2 GHz PowerPC G5, Apple Computer, Cupertino, CA) with 4 GB of SDRAM was used to perform the data analysis in Matlab 7.1.

The DoGEX program is composed of several routines used to process the LC-MS data. In the first step the data is aligned using a linear interpolation function. The interpolation has the effect of reducing the data density in the chromatographic dimension while preserving the same number of spectral data points. Several interpolation intervals were tested; we found that the use of 0.2 min provided a good compromise between reducing the data density while preserving all useful chromatographic information. The effect of the interpolation intervals on the file size and chromatographic peak shape is considered in the Supporting Information, with a chromatographic peak that is at least three times the signal to noise ratio (Figure S-1A). With a two min interpolation interval, a 91% data reduction is achieved but the chromatographic peak is lost (Figure S-1B). While the data reduction is more when a 0.2 min interval (Figure S-1C) is used, it is clear that a 0.02 min interval (Figure S-1D) preserved more chromatographic peak features. With a 0.002 min interpolating interval (Figure S-1E), the data was increased by 337% due to time oversampling.

The next processing step requires wavelets to remove noise from the chromatographic and spectral data.²⁵⁻²⁷ Wavelets may be used to represent a signal as a function of time. The following equation defines a signal, $f(t)$, as the sum of father (ϕ) and mother (ψ) wavelets, each of which is a function of time.²⁸⁻³⁰

$$f(t) = \sum_k s_{jk} \phi_{jk}(t) + \sum_{j=1}^J \sum_k d_{jk} \psi_{jk}(t) \quad (1)$$

The wavelet functions ϕ and ψ used here have orthogonal bases, and d_{jk} and s_{jk} are wavelets coefficients obtained by convolution. The sub-indices j and k are the dilation and translation indices, respectively. The equation that describes the solutions to the wavelet function ψ is:

$$\psi_{(a,b)}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (2)$$

where a and b are the scaling and translation parameters of a wavelet, respectively. We used a two-dimensional discrete stationary wavelet transform with two levels of decomposition and the symlet type-one wavelet to remove noise from the chromatographic signal in the experimental data. The stationary wavelet algorithm was used to remove chromatographic noise because this decomposition does not have any aliasing effect on the signal.

Utilization of the system is shown in Figure S-2 of the Supporting Information. The signal of a single ion chromatogram is shown (Figure S-2A). The effects of the use of the stationary (Figure S-2B) and the undecimated wavelets (Figure S-2C) algorithms are also shown. The level of decomposition for each analysis was kept equal to two, and in both cases a symlet wavelet was used. In both cases, the detail coefficients were set equal to zero and only the approximations were kept. The stationary wavelet was superior to the decimated wavelet in preserving the peak shape when the amount of data points is kept the same as in the original signal. Because wavelets work by decomposing the signal into approximation and details, it is possible to separate the noise (high frequency) from the chromatographic signal (low frequency). All coefficients of the high frequency are set equal to zero; the inverse of the stationary wavelet transform is then used to recreate the signal with less noise than the original signal.

Baseline Correction

The total ion chromatogram in an LC-MS run is the sum of all the individual ion intensities for each scan. Each ion signal will tend to drift to higher intensity values due to sample accumulation in the detector and changes in the ionization efficiency at the ion source, due in turn to changes in the mobile phase composition. Each ion chromatogram was baseline corrected using spline interpolation. The spline regression method is superior to a linear regression because the splines can adapt more readily to the chromatographic shape of the single ion chromatograms. The parameters used in the spline regression were 2 min for the step size window and 1 min for the shifting window. In some cases the spline correction can generate negative values, which were corrected by using a non-negative constraint that sets any negative intensity values to zero. In order to conserve disk space and memory, the entire intensity matrices were converted to unsigned 32- or 64-bit integers depending on the maximum intensity value observed.

Peak Detection

After chromatographic alignment, noise removal, and baseline correction, the DoGEX programs use an edge detection algorithm to detect chromatographic peaks in the data matrix. In a mathematical sense, the intensity data matrix is similar to a gray image that in turn allows use of imaging processing techniques with mass spectrometry data. The Matlab 7.1 image processing toolbox contains an edge detection function that is applied to detect chromatographic peaks in the data. Matrix convolution with a Roberts filter³¹ with no

thinning, followed by image dilation and hole filling, was used to detect all the chromatographic peaks and to suppress false chromatographic peaks in all the samples. The convolution with a Roberts filter can detect chromatographic peaks by using a derivative estimation to find gradients when an arbitrary threshold is exceeded. For this analysis a sensitivity factor of 1×10^6 intensity was employed. For the first peak detection step, the area integration threshold was 1×10^6 min-intensity. However, all of these parameters can be customized depending on the type of analysis to be done.

Chromatographic Peak Alignment

The DoGEX program can handle varying retention times (t_R) of the same chromatographic peak (m/z) across multiple samples by using two-dimensional LC-MS data overlapping. In each of 12 different samples, the *Streptomyces coelicolor* 4-butylcycloheptylprodiginine component has a slightly different retention time. However, by using two-dimensional LC-MS data overlapping it is possible to group the same chromatographic peak (m/z), even with slightly different retention times (Figure S-5 of Supporting Information). To handle very irreproducible retention time differences, zero padding (Figure S-6 of Supporting Information) or time warping algorithms^{32,33} could be implemented into the program.

Isotopic Signature Identification

Prior to the spectral filtering step, the local spectrum of each individual chromatographic peak is averaged. The spectral averaging has the effect of reducing the chromatographic noise while improving the quality of the local chromatographic spectra.

After averaging the spectra, DoGEX can display a two-dimensional view of the intensity matrix using a coloring scheme based on a user defined target ratio. The spectral filtering is designed to detect chromatographic peaks that have a user-specified ratio of the intensities ($m/z_1/(m/z_2)$). The intensity matrix is zero padded before performing this calculation and a threshold is applied to reduce the number of false positives in the data. A Jet color scheme is used, with a 64-color map that ranges from yellow to green (shown here) or any of several other color combinations. The user of the computer program can choose to obtain an integration results table that contains the integration areas of chromatographic peaks found in each individual LC-MS file. This matrix can also be studied in Matlab using PCA.

RESULTS AND DISCUSSION

Development of DoGEX Algorithm

In the first step of the DoGEX computer program, the program aligns each scan to an arbitrary value (Figure 1). For this alignment, a start and an end time are selected. Several intervals were tested, and 0.2 min provided a good compromise for reducing the data points in the chromatographic dimension without removing too much information (Figure S-1, Supporting Information). This data reduction and alignment are very important because under-sampling can decimate small chromatographic peaks.

Several families of wavelets were tested for noise removal in LC-MS data, and the symlet wavelet was chosen to perform the undecimated wavelet decomposition because it removed noise without affecting the chromatographic peak profiles (Figure S-2, Supporting Information). Other wavelets performed equally well in the noise removal step; however, the symlet wavelet was preferred over the Daubechies wavelet³⁴ because its shape resembles a chromatographic shape more closely (i.e., Figure S-5, Supporting Information; the ϕ function of the symlets resembles more chromatographic peaks than the ϕ function of the Daubechies wavelets). Two levels of decomposition were sufficient to remove most of the chromatographic noise in the LC-MS data sets without removing chromatographic peaks.

Higher noise removal levels eliminated some of the chromatographic signal and distorted the peak shape. In an example of the effect of using five levels to decompose with the symlets and Daubechies wavelets (Figure S-6, Supporting Information), the Daubechies method causes negative dips in the chromatogram as an artifact of the noise removal while in the symlet wavelet decomposition the t_R of the peak changes. More than two levels of decomposition are not recommended, in order to avoid artifacts due to the denoising step using wavelets. Baseline correction is applied after the noise in the data has been removed. The noise removal step is very important because it facilitates the gradient detection step used for automatic chromatographic peak detection.

In the next step (Figure 1), an isotopic filter is applied to filter out chromatographic peaks that do not have a desired spectral signature. The spectral filtering is achieved by calculating the ratio of intensities of spectral peaks that are separated by two mass units. This filter can be customized to any given ratio and to any given mass unit separation, e.g. bromine (Figure 2). The green color indicates regions where the ratio of the chromatographic peaks has a value close to the target selected, i.e. 0.95 in this case. We have observed that the filter is not effective when peaks have a very weak response, e.g. in the case of felodipine (*vide infra*). A typical area integration threshold was 1×10^6 min-intensity for the ThermoFinnigan LTQ instrument under these conditions.

Use of Software to Detect Halogens in LC-MS data

The filter was set to allow for detection of only brominated compounds, e.g. bromocriptine and its products (Figure 2). However, felodipine (with two chlorine atoms) also appears to have an isotopic ratio of 0.95 although the theoretical ratio of this molecule is 1.5. This result can be explained by considering that the mobile phase is the major contribution to the spectra for compounds that have a very weak response in electrospray ionization.

The application of a chlorine filter to LC-MS data obtained from an incubation of felodipine with P450 3A4 (and the cofactors NADPH and O_2) is shown in Figure 3. Felodipine was oxidized and the digital filter for the LC-MS data was set equal to 1.5. When the same sample was analyzed using APCI (negative ion mode), the chlorine filter allowed an easy detection of the monooxygenated, doubly-oxygenated, and dehydrogenated felodipine products.^{17,35}

Use of Software with Oxygen Isotopes

Isotopically-labeled testosterone hydroxylation products, prepared with a $^{18}O_2$: $^{16}O_2$ mixture, were analyzed to test the ability of the spectral filter to detect low concentrations of products under typical experimental conditions (Figure 4). The automatic peak detection algorithm of the DoGEX program was also utilized. Each colored line in Figure 5 represents a selected ion chromatogram with a chromatographic peak. Several types of gradient detection filters were tested, and the best results were obtained with the Roberts filter. Spectral filtering for an isotopic ratio of 1.8 is shown in Figure 6. In spectral filtering experiments with testosterone, the DoGEX program easily revealed chromatographic peaks with a predefined ratio. The products detected by DoGEX were identified as 15β -hydroxy-, 6β -hydrox-, and 2β -hydroxy-hydroxytestosterone by direct comparison with authentic standards (Figure 6). The remaining product (t_R 18.7 min) was assigned as 1β -hydroxytestosterone by comparison with P450 3A4 experiments done previously in this laboratory.^{36,37} The data also show several false positives near the substrate testosterone (t_R 23.0 min) that come from MS fragments and gas phase ion adducts resulting from overemphasized sensitivity settings in the program. This isotopic labeling approach, developed for hydroxylations, will not function for oxidative dealkylation and

rearrangement reactions that do not retain oxygen atoms, but in those cases PCA and other multivariate chemometrics techniques can be used to analyze the data.

CONCLUSIONS

The untargeted metabolic approach developed here has potential to elucidate functions of new enzymes. The DoGEX software is a very flexible computer program that can also be used to perform multivariate analysis such as PCA to look for unbiased differences among many samples. DoGEX can generate a chromatographic report of all the integration of all the chromatographic peaks identified in each sample. This new matrix can be interrogated using the statistical toolbox available in Matlab®. Although a number of LC-MS programs have been used to implement various aspects of isotope analysis^{16,38-40} and metabolomic analysis of LC-MS data,^{3,4,6-10} none of the existing software programs appear to be capable of analysis of isotope tagging, baseline correction, *and* qualitative analysis of multiple sets of m/z . t_R data that the program described here (DoGEX) is. Moreover, the program is able to accommodate collision-induced dissociation spectra, which can be obtained by running data-dependent experiments to help in the elucidation of the structure of candidates. The fragmentation patterns of the substrates and the metabolites should be similar, except for the presence of the hydroxyl group (or loss of an alkyl group in dealkylated products), to allow the matching of substrates and metabolites and establish new metabolic reactions for the P450s.

The DoGEX program does not employ autocorrelation (CODA)^{41,42} or spectral slicing (XCMS)⁷ to detect peaks but rather uses an approach similar to imaging edge detection. We preferred this method to others reported in the literature because we do not make any assumption about chromatographic peak shapes prior to the peak detection step. In DoGEX, a full LC/MS matrix was convoluted using a digital filter (i.e. Roberts) for detecting the gradients in LC/MS data. These gradients indicate the localization of potential chromatographic peaks in LC/MS data. We believe that this is a useful additional approach to the existing methods in the literature for peak detection in LC/MS data. Zero padding was used to align multiple LC/MS data files (Figure S-4, Supporting Information). The zero padding protocol is a first step to compensate for non-linear retention time behavior and newer versions of DoGEX will implement more advanced alignment tools. We are aware of other peak matching protocols, e.g. those employed by XCMS which matches peaks by generating average chromatograms define groups of peaks. A limitation of this approach is that peak differences cannot be larger than the retention time between two contiguous peaks. For these reasons, we have aligned the peaks as a preprocessing step and then define group of peaks by two-dimensional overlapping. Additionally, DoGEX provides an alternative to identify chromatographic peaks in LC/MS data that have a specific isotopic label.⁴³

We demonstrated that spectral filtering of LC-MS data is a valuable tool for observing the products formed by an enzymatic reaction when the substrates have a characteristic spectral signature or when the products can be isotopically labeled during the ongoing reaction. Assigning endogenous functions to newly characterized P450s in humans¹³ and microorganisms⁴⁴⁻⁴⁶ is a challenge that can be addressed with the new bioinformatic strategies developed in this work. As demonstrated here, the software can also be utilized in the interrogation of LC-MS data for minor reaction products generated from compounds with distinctive natural (and artificial) multiple-isotope atoms (e.g., Cl and Br).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank D. T. Duncan for writing a program for the extraction of raw LC-MS data into text files, D. Lamb and M. R. Waterman for the bacterial extracts used to derive 4-butylcycloheptylprodigine data, and D. L. Hachey for helpful discussions and comments on the manuscript. Financial support was provided in part from National Institutes of Health grants R37 CA090426, T32 ES007028, and P30 ES000267.

LITERATURE CITED

1. Hughes TR, Robinson MD, Mitsakakis N, Johnston M. *Curr. Opin. Microbiol.* 2004; 7:546–554. [PubMed: 15451511]
2. Jansen R, Gerstein M. *Curr. Opin. Microbiol.* 2004; 7:535–545. [PubMed: 15451510]
3. Saito N, Robert M, Kita mura S, Baran R, Soga T, Mori H, Nishioka T, Tomita M. *J. Proteome Res.* 2006; 5:1979–1987. [PubMed: 16889420]
4. Saghatelian A, Cravatt BF. *Curr. Opin. Chem. Biol.* 2005; 9:62–68. [PubMed: 15701455]
5. Porter SE, Stoll DR, Rutan SC, Carr PW, Cohen JD. *Anal. Chem.* 2006; 78:5559–5569. [PubMed: 16878896]
6. Katz JE, Dumlao DS, Clarke S, Hau J. J. *Am. Soc. Mass Spectrom.* 2004; 15:580–584. [PubMed: 15047062]
7. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. *Anal. Chem.* 2006; 78:779–787. [PubMed: 16448051]
8. Saghatelian A, Trauger SA, Want EJ, Hawkins EG, Siuzdak G, Cravatt BF. *Biochemistry.* 2004; 43:14332–14339. [PubMed: 15533037]
9. Katajamaa M, Oresic M. *BMC Bioinformatics.* 2005; 6:179. [PubMed: 16026613]
10. Smilde AK, van der Werf MJ, Bijlsma S, van der Werff-van der Vat BJ, Jellema RH. *Anal. Chem.* 2005; 77:6729–6736. [PubMed: 16223263]
11. Ortiz de Montellano, PR., editor. *Cytochrome P450: Structure, Mechanism, and Biochemistry.* 3rd ed.. Kluwer Academic/Plenum Publishers; New York: 2005.
12. Guengerich, FP. *Cytochrome P450: Structure, Mechanism, and Biochemistry.* 3rd ed.. Ortiz de Montellano, PR., editor. Kluwer Academic/Plenum Publishers; New York: 2005. p. 377-530.
13. Guengerich FP, Wu Z-L, Bartleson CJ. *Biochem. Biophys. Res. Commun.* 2005; 338:465–469. [PubMed: 16126164]
14. Dalluge JJ, Liao H, Gokarn R, Jessen H. *Anal. Chem.* 2005; 77:6737–6740. [PubMed: 16223264]
15. Mougous JD, Leavell MD, Senaratne RH, Leigh CD, Williams SJ, Riley LW, Leary JA, Bertozzi CR. *Proc. Natl. Acad. Sci. U.S.A.* 2002; 99:17037–17042. [PubMed: 12482950]
16. Qian WJ, Monroe ME, Liu T, Jacobs JM, Anderson GA, Shen Y, Moore RJ, Anderson DJ, Zhang R, Calvano SE, Lowry SF, Xiao W, Moldawer LL, Davis RW, Tompkins RG, Camp DG 2nd, Smith RD. *Mol. Cell. Proteomics.* 2005; 4:700–709. [PubMed: 15753121]
17. Guengerich FP, Brian WR, Iwasaki M, Sari M-A, Bäärnhielm C, Berntsson P. *J. Med. Chem.* 1991; 34:1838–1844. [PubMed: 2061924]
18. Isin EM, Guengerich FP. *J. Biol. Chem.* 2006; 281:9127–9136. [PubMed: 16467307]
19. Parikh A, Gillam EMJ, Guengerich FP. *Nat. Biotechnol.* 1997; 15:784–788. [PubMed: 9255795]
20. Burleigh BD Jr, Foust GP, Williams CH Jr. *Anal. Biochem.* 1969; 27:536–544. [PubMed: 5767207]
21. Guengerich FP, Johnson WW. *Biochemistry.* 1997; 36:14741–14750. [PubMed: 9398194]
22. Guengerich, FP. *Principles and Methods of Toxicology.* 4th ed.. Hayes, AW., editor. Taylor & Francis; Philadelphia: 2001. p. 1625-1687.
23. Lin SM, Zhu LH, Winter AQ, Sasinowski M, Kibbe WA. *Expert Rev. Proteomics.* 2005; 2:839–845. [PubMed: 16307524]
24. Pedrioli PG, Eng JK, Hubley R, Vogelzang M, Deutsch EW, Raught B, Pratt B, Nilsson E, Angeletti RH, Apweiler R, Cheung K, Costello CE, Hermjakob H, Huang S, Julian RK, Kapp E, McComb ME, Oliver SG, Omenn G, Paton NW, Simpson R, Smith R, Taylor CF, Zhu W, Aebersold R. *Nat. Biotechnol.* 2004; 22:1459–1466. [PubMed: 15529173]

25. Walczak B, Massart DL. *Trac-Trends Anal. Chem.* 1997; 16:451–463.
26. Andersson FO, Kaiser R, Jacobsson SP. *J. Pharmaceut. Biomed. Anal.* 2004; 34:531–541.
27. Dremine IA. *Physics Atomic Nuclei.* 2005; 68:508–520.
28. Graps A. *IEEE Computat. Sci. Engin.* 1995; 2:50–61.
29. Jetter K, Depczynski U, Molt K, Niemoller A. *Analyt. Chim. Acta.* 2000; 420:169–180.
30. Perrin C, Walczak B, Massart DL. *Anal. Chem.* 2001; 73:4903–4917. [PubMed: 11681466]
31. Roberts, LG. *Machine Perception of Three-dimensional Solids.* MIT Press; Cambridge, MA: 1965.
32. Jaitly N, Monroe ME, Petyuk VA, Clauss TR, Adkins JN, Smith RD. *Anal. Chem.* 2006; 78:7397–7409. [PubMed: 17073405]
33. Wang CP, Isenhour TL. *Anal. Chem.* 1987; 59:649–654.
34. Daubechies I. *Commun. Pure Appl. Math.* 1988; 41:909–996.
35. Bäärnhielm C, Backman A, Hoffmann KJ, Weidolf L. *Drug. Metab. Dispos.* 1986; 14:613–618. [PubMed: 2876870]
36. Krauser JA, Voehler M, Tseng L-H, Schefer AB, Godejohann M, Guengerich FP. *Eur. J. Biochem.* 2004; 271:3962–3969. [PubMed: 15373842]
37. Krauser JA, Guengerich FP. *J. Biol. Chem.* 2005; 280:19496–19506. [PubMed: 15772082]
38. Rockwood AL, Van Orden SL, Smith RD. *Anal. Chem.* 1995; 67:2699–2704.
39. Shouakar-Stash O, Frape SK, Drimmie RJ. *Anal. Chem.* 2005; 77:4027–4033. [PubMed: 15987107]
40. MacCoss MJ, Wu CC, Matthews DE, Yates JR 3rd. *Anal. Chem.* 2005; 77:7646–7653. [PubMed: 16316172]
41. Windig W. *Chemom. Intell. Lab. Syst.* 2005; 77:206–214.
42. Windig W, Phalp JM, Payne AW. *Anal. Chem.* 1996; 68:3602–3606.
43. Yao XD, Freas A, Ramirez J, Demirev PA, Fenselau C. *Anal. Chem.* 2001; 73:2836–2842. [PubMed: 11467524]
44. Park SY, Yamane K, Adachi S, Shiro Y, Weiss KE, Maves SA, Sligar SG. *J. Inorg. Biochem.* 2002; 91:491–501. [PubMed: 12237217]
45. Puchkaev AV, Ortiz de Montellano PR. *Arch. Biochem. Biophys.* 2005; 434:169–177. [PubMed: 15629120]
46. Lamb DC, Guengerich FP, Kelly SL, Waterman MR. *Expert Opin. Drug Metab. Toxicol.* 2006; 2:27–40. [PubMed: 16863466]

Figure 1. DoGEX data processing flow chart. Each intensity matrix is processed with the first five steps; the spectral filtering is optional.

Figure 2.

DoGEX bromine spectral filtering. *A*, Raw data obtained from an injection of a mixture of bromocriptine (1 pmol) and felodipine (10 pmol). Bromocriptine (m/z 654.1, 656.1, ESI positive mode) contains a single bromine and felodipine (m/z 384.1, 383.1, 382.1, ESI negative mode) contains two chlorine atoms. *B*, The green colored regions show which chromatographic peaks have a target ratio close to 0.95, while the yellow area indicates all the values outside of the threshold value. *C*, Expansion of Part B. *D*, Expansion of Part C and conversion to a “3-dimensional” profile, demonstrating the pattern of the two ions separated by 2 amu.

Figure 3. Aerial view of two-dimensional LC-MS data showing the effectiveness of a chlorine filter to automatically select the chromatographic peaks with a predetermined isotopic ratio (set to 1.5). The P450 3A4 oxidation products of felodipine are shown.³⁵ The chlorine spectral filtering shows the localization in the (t_R , m/z) map of the substrate and the incubation products.

Figure 4.

LC-MS data subtraction between an incubation of testosterone with P450 3A4 and a control experiment (devoid of NADPH). The four peaks in the subtracted LC-MS data at m/z 305.1 and eluting at t_R 17.0, 17.8, 18.7, and 19.9 min represent the hydroxylation products of testosterone, (15 β -, 6 β -, 1 β -, and 2 β -hydroxy) respectively (see also Figures 5 and 6).

Figure 5.

Automatic peak detection by means of a gradient detection algorithm. An arbitrary gradient threshold was selected to detect single ion chromatograms present in a total ion chromatogram. Chromatographic peak selection by DoGEX, from an incubation of testosterone with P450 3A4, NADPH, and a 1:1 mixture of $^{16}\text{O}_2$ and $^{18}\text{O}_2$. Each colored line represents a selected ion chromatogram that contains a chromatographic peak. The four peaks at m/z 305.1 and eluting at t_R 17.0, 17.8, 18.7, and 19.9 min represent the hydroxylation products of testosterone (15 β -, 6 β -, 1 β -, and 2 β -hydroxy, respectively).

Figure 6.

LC-APCI-MS data from analysis of a P450 3A4 incubation with testosterone, NADPH, and a 1:1 mixture of $^{16}\text{O}_2$ and $^{18}\text{O}_2$ (Figure 5). The chromatographic peaks near t_R 23 min represent contaminating fragmentation products and adducts ions derived from the substrate, visible at the sensitivity settings used. The four peaks at m/z 305.1 and eluting at t_R 17.0, 17.8, 18.7, and 19.9 min represent the hydroxylation products of testosterone (15 β -, 6 β -, 1 β -, and 2 β -hydroxy), identified in separate experiments by coincident elution with authentic standards or other comparisons.³⁶