Anal. Chem. **2002,** *74,* 5393−5404

Analyzing the Physiological Signature of Anabolic Steroids in Cattle Urine Using Pyrolysis/Metastable Atom Bombardment Mass Spectrometry and Pattern Recognition

Marc-Emmanuel Dumas,[†] Laurent Debrauwer,*[,†] Loïc Beyet,[§] Denis Lesage,[§] François André,[‡] Alain Paris,[†] and Jean-Claude Tabet[§]

Laboratoire des Xénobiotiques UMR 1089, INRA, BP3, 31931 Toulouse Cedex 09, France, LABERCA, ENV, BP 50707, 44307 Nantes Cedex 03, France, and Laboratoire de Chimie Structurale Organique et Biologique UMR 7613, Université Pierre & Marie Curie, 75252 Paris Cedex 05, France

Pyrolysis coupled to metastable atom bombardment (MAB) and time-of-flight mass spectrometry (TOFMS) is used for generating mass spectra from bovine urine samples obtained from cattle treated with anabolic steroids. These spectra constitute fingerprints, which can be discriminated by multivariate statistical analysis. Four main conclusions can be drawn from this work: (i) The use of different metastable gases, such as Xe*, Kr*, or N_2*, as an energy-tunable ionization beamline allows control of the internal energy and the dissociation processes of the produced odd electron molecular ions, thus giving rise to complementary mass spectra fingerprints. (ii) A variable transformation depending on the biofluid matrix suitably contracts the frequency distribution of the generated data for low m/z ratios holding information related to endogenous metabolites encountered in urine. (iii) Coupling variable selection to statistical pattern recognition methods results in low error rates (<1%) for predicting MAB mass fingerprints, especially using linear discriminant analysis (LDA). (iv) LDA discriminates controls from treated animals and also correlates to quantitative physiological responses induced by anabolic steroids. This work shows that Py-MAB-TOFMS could be a suitable method for complementary monitoring anabolic use in sports, medicine, and cattle breeding, as well as monitoring many other long-lasting although weak physiological disruptions.

Mass spectrometry (MS) has been widely used for a long time for the analysis of complex chemical or biological systems. For this purpose, a first functional approach consists of a qualitative analysis of metabolites or signals characterizing a biological context[1] as relevant biomarkers. These biomarkers can be considered as binary thresholds characterizing qualitative phenomena[2−4] that are likely observed by semiquantitative methods.[5] A further level of investigation involving descriptive and functional characterization of structures related to the functioning of organisms is sought for biological systems. This requires acute description of different biological situations. In this context, MS studies for monitoring growth of bacteria cultures,[6] typing antigen polymorphism,[7] serotyping viruses,[8] detection of cell cultures infected by viruses,[9] discrimination of bacterial strains resistant or sensitive to antibiotics,[10] or phyloproteomic typing of bacterial strains[11] have been sought. All of these applications are devoted to genomic and proteomic applications.

Alternatively, discriminating physiological variations requires probing metabolic network changes. Analyzing these variations and performing a misclassification assessment when predicting samples[12] covers both functional characterization and acute quantification aspects. Goodacre and his colleagues have developed an approach that combines multivariate treatment to MS[13,14] as well as IR spectroscopy.[15] In a parallel way, Nicholson's group has already developed a similar approach called "metabonomics"[16]

* To whom correspondence should be addressed. Phone: +33 561 285 013. Fax: +33 561 285 244. E-mail: ldebrauw@toulouse.inra.fr.

[†] Laboratoire des Xénobiotiques.

[‡] LABERCA.

[§] Université Pierre & Marie Curie.

(1) Jarman, K. H.; Cebula, S. T.; Saenz, A. J.; Petersen, C. E.; Valentine, N. B.; Kingsley, M. T.; Wahl, K. L. *Anal. Chem.* **2000**, *72*, 1217−23.

(2) Xiang, F.; Anderson, G. A.; Veenstra, T. D.; Lipton, M. S.; Smith, R. D. *Anal. Chem.* **2000**, *72*, 2475−81.

(3) Winkler, M. A.; Uher, J.; Cepa, S. *Anal. Chem.* **1999**, *71*, 3416−9.

(4) Snyder, A. P.; McClennen, W. H.; Dworzanski, J. P.; Meuzelaar, H. L. *Anal. Chem.* **1990**, *62*, 2565−73.

(5) Jimenez, C. R.; Li, K. W.; Dreisewerd, K.; Mansvelder, H. D.; Brussaard, A. B.; Reinhold, B. B.; Van der Schors, R. C.; Karas, M.; Hillenkamp, F.; Burbach, J. P.; Costello, C. E.; Geraerts, W. P. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 9481−6.

(6) Arnold, R. J.; Karty, J. A.; Ellington, A. D.; Reilly, J. P. *Anal. Chem.* **1999**, *71*, 1990−6.

(7) Worrall, T. A.; Schmeckpeper, B. J.; Corvera, J. S.; Cotter, R. J. *Anal. Chem.* **2000**, *72*, 5233−8.

(8) Yao, X.; Freas, A.; Ramirez, J.; Demirev, P. A.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 2836−42.

(9) Madonna, A. J.; Voorhees, K. J.; Hadfield, T. L.; Hilyard, E. J. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 502−11.

(10) Edwards-Jones, V.; Claydon, M. A.; Evason, D. J.; Walker, J.; Fox, A. J.; Gordon, D. B. *J. Med. Microbiol.* **2000**, *49*, 295−300.

(11) Ryzhov, V.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 746−50.

(12) Pineda, F. J.; Lin, J. S.; Fenselau, C.; Demirev, P. A. *Anal. Chem.* **2000**, *72*, 3739−44.

(13) Goodacre, R.; Kell, D. B. *Curr. Opin. Biotechnol.* **1996**, *7*, 20−8.

(14) Goodacre, R.; Neal, M. J.; Kell, D. B. *Zentralbl. Bakteriol.* **1996**, *284*, 516−39.

(15) Goodacre, R.; Timmins, E. M.; Rooney, P. J.; Rowland, J. J.; Kell, D. B. *FEMS Microbiol. Lett.* **1996**, *140*, 233−9.

10.1021/ac025656k CCC: $22.00
Published on Web 09/10/2002

involving NMR spectroscopy of biofluids and multivariate statistics to investigate biological contrasts.[17−19]

Sample identification involves computational methods, such as algorithms for analyte identification by data-dependent MS/MS spectra;[20] database searching;[12,21,22] multivariate statistics;[23,24] or neural networks.[14,15,25] As a matter of fact, multivariate statistical analysis is a classical mean for quantifying risks related to classification and for explanation of such contrasts. Multivariate techniques can be grouped into two families: (i) unsupervised methods explaining total variance, such as clustering methods or principal components analysis (PCA); and (ii) supervised or pattern recognition methods, such as linear discriminant analysis (LDA),[26] multiple logistic regression (MLR),[27] or neural networks (NN)[28] that analyze the class distribution and classify unknown samples after an initial learning step. Other pattern recognition methods use nonparametric classifiers, such as *k*-nearest neighbors (*k*-nn) or learning vector quantization (LVQ).[29]

Concerning the generation of data, most mass spectrometry applications used for studying biological mixtures are dedicated to the detection of microorganisms. The diversity of analytical techniques involved for this purpose is plentiful. Sample uptake and separation methods include pyrolysis,[30,31] GC,[32] LC, and capillary LC.[33] Several ionization/desorption processes have been used, such as FAB,[23,34,35] MALDI,[1,6,10,21] even intact-cell MALDI[36] or electrospray.[2] Analysis of ions is generally achieved by means of quadrupole,[37] time-of-flight (TOF)[21] and Fourier transform ion cyclotron resonance[20,33] analyzers. In fact, GC/EI-MS was one of the first methods involved in metabolic profiling,[32] and it is still employed.[38] At the same period, the choice of pyrolysis for complete analysis of samples was obvious, especially in the case of materials that are difficult to volatilize using conventional ionization modes. Indeed, the first characterization of complex nonvolatile biological samples from bacteria was made by Curie-point pyrolysis mass spectrometry.[30,31] Seeking for features such as repeatability or large mass and dynamic range, pyrolysis-mass spectrometry (Py-MS) appears to be a method of choice for characterizing complex systems of chemical or biological origin. Electron ionization at low energies (25 eV or less)[39] is widely used as an ionization technique for this purpose and is preferred to chemical ionization, even though a recent study has shown a slight superiority of CI against EI (70 eV)[40] for discriminating bacteria using PCA.

However, extensive fragmentation causes loss of spectral information and detrimental complication of mass spectra interpretation. Alternatively, the recent metastable atom bombardment (MAB) ionization mode is known to provide selective ionization as well as controlled fragmentation processes.[41] The potential of this ionization mode is fully expressed when fingerprinting is performed on biological fluids.[42−44]

In this study, we use Py-MAB-TOFMS as a metabonomic tool, providing a suitable variable generator for assessing weak physiological variations induced by normal or subnormal anabolic treatment conditions in cattle. Anabolic steroids are responsible for morphological modifications, such as muscle accretion and lower fat deposit, which are explained by biochemical variations at the general metabolism level.[45] This work shows how indirect metabolic variations induced by physiological response to hormonal treatment can be evidenced through urine monitoring without searching hormone residues.

EXPERIMENTAL SECTION

Animals and Preparation of Urine Samples and Chemical Standards. Hereford breeding steers ($n = 20$), 1 year old, were equally distributed between a control (group C) and groups treated with one (group 1), two (group 2), and four (group 4) Revalor implants (Hoechst-Roussel Vet, Sommerville, NJ). These implants contain 140 mg of trenbolone acetate (TBA) and 24 mg of 17β-estradiol (E_2). Groups 1 and 4 were implanted on day 0, whereas group 2 was implanted once at day 0 and once again on the 45th day. Urine was collected on the 90th day after the first implantation. Urine samples were aliquoted twice, and freeze-dried. Each lyophilizate was dissolved at 50 mg/mL in water, and 0.2 μL of solution was deposited inside a quartz capillary.

(16) Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* **1999**, *29*, 1181−9.
(17) Gavaghan, C. L.; Holmes, E.; Lenz, E.; Wilson, I. D.; Nicholson, J. K. *FEBS Lett.* **2000**, *484*, 169−74.
(18) Holmes, E.; Nicholls, A. W.; Lindon, J. C.; Connor, S. C.; Connelly, J. C.; Haselden, J. N.; Damment, S. J.; Spraul, M.; Neidig, P.; Nicholson, J. K. *Chem. Res. Toxicol.* **2000**, *13*, 471−8.
(19) Holmes, E.; Nicholson, J. K.; Tranter, G. *Chem. Res. Toxicol.* **2001**, *14*, 182−91.
(20) Li, L.; Masselon, C. D.; Anderson, G. A.; Pasa-Tolic, L.; Lee, S. W.; Shen, Y.; Zhao, R.; Lipton, M. S.; Conrads, T. P.; Tolic, N.; Smith, R. D. *Anal. Chem.* **2001**, *73*, 3312−22.
(21) Demirev, P. A.; Ho, Y. P.; Ryzhov, V.; Fenselau, C. *Anal. Chem.* **1999**, *71*, 2732−8.
(22) Demirev, P. A.; Lin, J. S.; Pineda, F. J.; Fenselau, C. *Anal. Chem.* **2001**, *73*, 4566−73.
(23) Platt, J. A.; Uy, O. M.; Heller, D. N.; Cotter, R. J.; Fenselau, C. *Anal. Chem.* **1988**, *60*, 1415−9.
(24) Vaidyanathan, S.; Rowland, J. J.; Kell, D. B.; Goodacre, R. *Anal. Chem.* **2001**, *73*, 4134−44.
(25) Goodfellow, M.; Freeman, R.; Sisson, P. R. *Zentralbl. Bakteriol.* **1997**, *285*, 133−56.
(26) Fisher, R. A. *Ann. Eugenics* **1936**, *7*, 179−88.
(27) Venables, W. N.; Ripley, B. D. *Modern Applied Statistics with S-PLUS*, 3rd ed.; Springer-Verlag: New York, 1999.
(28) McCulloch, W. S.; Pitts, W. *Bull. Math. Biophys.* **1943**, *5*, 115−33.
(29) Kohonen, T. *Proc. IEEE* **1990**, *78*, 1464−80.
(30) Meuzelaar, H. L.; Kistemaker, P. G. *Anal. Chem.* **1973**, *45*, 587−90.
(31) Schulten, H. R.; Beckey, H. D.; Meuzelaar, H. L.; Boerboom, A. J. *Anal. Chem.* **1973**, *45*, 191−5.
(32) Thompson, J. A.; Markey, S. P. *Anal. Chem.* **1975**, *47*, 1313−21.
(33) Shen, Y.; Tolic, N.; Zhao, R.; Pasa-Tolic, L.; Li, L.; Berger, S. J.; Harkewicz, R.; Anderson, G. A.; Belov, M. E.; Smith, R. D. *Anal. Chem.* **2001**, *73*, 3011−21.
(34) Heller, D. N.; Cotter, R. J.; Fenselau, C.; Uy, O. M. *Anal. Chem.* **1987**, *59*, 2806−9.
(35) Heller, D. N.; Murphy, C. M.; Cotter, R. J.; Fenselau, C.; Uy, O. M. *Anal. Chem.* **1988**, *60*, 2787−91.
(36) Evason, D. J.; Claydon, M. A.; Gordon, D. B. *J. Am. Soc. Mass Spectrom.* **2001**, *12*, 49−54.
(37) Griffin, T. J.; Gygi, S. P.; Rist, B.; Aebersold, R.; Loboda, A.; Jilkine, A.; Ens, W.; Standing, K. G. *Anal. Chem.* **2001**, *73*, 978−86.
(38) Peres, C.; Viallon, C.; Berdague, J. L. *Anal. Chem.* **2001**, *73*, 1030−6.
(39) Mellon, F.; Self, R.; Startin, J. R. *Mass Spectrometry of Natural Substances in Food*; Royal Society of Chemistry: Cambridge, 2000, Chapter 11.
(40) Beverly, M. B.; Basile, F.; Voorhees, K. J.; Hadfield, T. L. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 747−58.
(41) Faubert, D.; Paul, G. J. C.; Giroux, J.; Bertrand, M. J. *Int. J. Mass Spectrom. Ion Processes* **1993**, *124*, 69−77.
(42) Wilkes, J. G.; Heinze, T. M.; Freeman, J. P.; Rafii, F.; Cyr, M.; Mousselmal, M.; Bertrand, M. J.; Voorhees, K. J. *Proceedings of the 45th ASMS Conference on Mass Spectrometry and Allied Topics*, Palm Springs, CA, June 1−5, 1997; p 1358.
(43) Wilkes, J. G.; Letarte, S.; Glover, K. L.; Holcomb, M.; Rafii, F.; Bertrand, M. J. *Proceedings of the 47th ASMS Conference on Mass Spectrometry and Allied Topics*, Dallas, TX, June 13−17, 1999; p 859.
(44) Letarte, S.; Mousselma, M.; Faubert, D.; Bertrand, M. J. *Proceedings of the 46th ASMS Conference on Mass Spectrometry and Allied Topics*, Orlando, FL, May 31−June 4, 1998; p 897.
(45) Meyer, H. H. *Apmis* **2001**, *109*, 1−8.

Table 1. MAB Ionization Conditions and Metastable State Abundances of Different Gases

gas	discharge current (mA)	gun pressure (mbar)	energy (eV)	lifetime (s)	approximate population (%)
xenon	4−6	60	9.45	7.8×10^{-2}	<7
			8.32	1.5×10^{2}[a]	>93
krypton	6−9	60	10.56	4.9×10^{-1}	<10
			9.92	8.5×10^{1}[a]	>90
nitrogen	10−13	80	11.88	2×10^{-4}[a]	<15
			9.02	$1-5 \times 10^{-4}$	>85
			8.67	$1-1.5 \times 10^{-4}$	
			8.52	1.4[a]	

[a] Major metastable states involved in ionization process.

Representative reference compounds, such as urea, creatine, creatinine, trimethylamine-*N*-oxide, and hippuric acid (Sigma-Aldrich, Saint-Quentin Fallavier, France), were prepared in H_2O (5 μg/μL), and 0.2 μL of solution was deposited inside a quartz capillary.

Mass Spectrometry. MAB mass spectra were obtained on a TOF mass spectrometer (Dephy Technologies Inc., Montreal, Canada) fitted with a Pyroprobe 2000 pyrolyzer (CDS Analytical, Oxford, PA). Pyrolysis was achieved by ramping the probe temperature by 20 °C/ms from ambient to 1100−1200 °C, with a final hold time of 180−300 s. The probe was specially modified for enabling helium flow (1−2 mL/min) through the quartz capillary and then enhancing transfer of pyrolysis products into the MAB source. The metastable species N_2*, Xe*, and Kr* (8.5, 8.3, and 9.9 eV, assuming that the major and longest lifetime metastable atoms are responsible for ionization) were produced by corona discharge (−350 V), the current intensity and the gas pressure of which were optimized for each gas (Table 1).

As previously reported by Faubert et al.,[41] MAB ionization is reached in two steps: (i) metastable atoms are produced through a rare gas flow, such as helium, neon, argon, krypton, xenon, or nitrogen, (ii) Penning ionization reactions of types I and II occur between organic compounds and metastable species in the collision cell considered as the analyte source, seemingly

$$A^* + BC \xrightarrow{I} A + BC^{+\bullet} + e^- \xrightarrow{II}$$

$$[\, A + B^{+\bullet} + C + e^- \text{ or } A + B^+ + C^\bullet + e^- \,]$$

These reactions can occur, provided the excitation energy of the metastable atom is higher than the ionization potential of the compound. A reaction of type I involves an electrophilic reaction of the metastable A* with the analyte BC, resulting in a return of the metastable atom to its ground state (A) and ionization of the analyte as an odd-electron $BC^{+\bullet}$ species. More precisely, the molecular orbital of BC transfers one electron to a vacant orbital of the excited atom A*, leading to a simultaneous ejection of one electron from the outer shell of A*. When the amount of excitation energy transferred to the analyte is high enough, one or more bonds may break within $BC^{+\bullet}$, as indicated in reaction II.

Energy-tunable ionization control is reached by using selectively a N_2*, Xe*, or Kr* metastable beam as the ionizing agent. Ions were collected in the 55−800 Th range. Orthogonal extraction for time-of-flight acquisition was run at 32 kHz sampling frequency. Mass spectra were prepared by pyrogram averaging and subtraction of background registered before and after the pyrolysis step.

Multivariate Statistical Analysis. Multivariate statistical analyses were performed using Splus 2000 (v2.0, Mathsoft Inc., Seattle, WA) including MASS, Nnet, Class, Rpart (http://lib.stat.cmu.edu/DOS/S), and Multidim (http://www.lsp.ups-tlse.fr/Carlier/Logiciel.html) libraries. Variables were built from m/z ratios expressed in percentage of base peak, and therefore, spectra corresponded to vectors that characterize samples.[14] Reliability of the ionization mode was investigated by computing variation coefficients (VC) corresponding to the standard deviation (SD) to the mean (μ) ratio for each nominal m/z variable on repeated injections of the same urine sample ($n = 36$). Fingerprints are expressed in proportion of the base peak, and variables were transformed with $y = x^{1-k}$ to stabilize for unequal variances, k being given by the value of the slope obtained by linear regression between logarithmic values of SD and μ for every m/z variable, considering the different m/z variables as a whole.

Multivariate statistical analysis was performed on male groups collected on the 90th day, distributed between groups C, 1, 2, and 4. Because of the singular properties of the variance−covariance matrix (rank problem due to the lower number of individuals compared to the number of variables, high correlation between some specific variables), we needed to reduce its dimension. This was achieved by an efficient variable filtration algorithm as described recently.[46] Briefly, the principle is to select informative but nonredundant variables coming from a simple analysis of variance (one-way ANOVA) with a cutoff P value. In a second step, the correlation between the most informative variable, which is selected, and a dummy variable explaining the group structuring of individuals is performed, as published elsewhere.[47] Then, all variables that are highly correlated to this selected variable are rejected. The most informative remaining variables are taken in a further step to calculate the correlation with a vectorial space, which is orthogonal to this already built from the previous selection steps. A descending stepwise introduction of the variables on their partial correlation to the group distribution, with a rejection of variables that are linearly correlated to the previously selected vectorial subspace (parametrized by β) is performed step by step. The β value corresponds to the flexibility parameter for the correlation between candidate variables and a vectorial space orthogonal to this that has been selected. Such an algorithm provides a sequence of selected variables. The n first variables issued from this sequence are then used to perform statistical pattern recognition. This algorithm is achieved with parameters (P, β) being equal to (10^{-2}, 1) to obtain an optimal discrimination.[46]

In addition, to give a better insight of such a necessary variable selection, the hierarchical ascending clustering of all variables with complete linkage aggregations was analyzed. This variable classification was performed using the metric $1 - r_{ij}$ as a metric giving the distance between two variables (r_{ij} defining the correlation between variables i and j) on the basis of the similarity of their profiles, which are expressed in the space defined by all other variables.[46]

(46) Dumas, M. E.; Canlet, C.; André, F.; Vercauteren, J.; Paris, A. *Anal. Chem.* **2002**, *74*, 2261−2273.

(47) Carlier, A. In *Analyse Discriminante sur Variables Qualitatives*; Celeux, G., Nakache, J. P., Eds.; Polytechnica: Paris, 1994; Chapter 5.

Linear discriminant analysis (LDA) classifies individuals on the basis of a between-group variance maximization along independent (orthogonal) linear combinations of variables. Projecting individuals on the resulting 2D-factorial maps generates graphical representations of the canonical space of discrimination.

Classical multiperceptron neural networks (NN) were built with a single hidden layer.[28] The size of the hidden layer was fixed to the mean of the number of inputs (variables) and the number of outputs (groups). Logistic output units and least-squares fitting with unique initial random weights on the $[-0.7, +0.7]$ interval were used for parameter optimization along 100 iterations. Another pattern recognition method used as a benchmark in this work was multiple linear regression (MLR) that fitted multinomial log−linear models, the parameters of which were computed by neural networks.[48]

The method of *k*-nearest neighbors (*k*-nn) estimates group density by allocating individuals to be classified to the group of their *k*-nearest neighbors.[48] Here, *k*-nn has been parametrized with *k* equal to 1 and the minimum vote for decision being equal to 1 to disable the rejection option. Learning vector quantization (LVQ) iteratively estimates the boundaries between groups by updating the so-called codebook vectors.[29] The initialization step was performed with an equal prior probability with 3 codebook vectors/class from the training set classes. Although giving higher error rates, the OLVQ1 procedure was applied for a rapid convergence. Finally, the LVQ3 procedure was used to get a better prediction after convergence was obtained by OLVQ1.[29]

Ten-fold cross validation (CV) error rates assessed the respective performance of LDA, *k*-nearest neighbors (*k*-nn), learning vector quantization (LVQ), multiple logistic regression parametrized by neural networks (MLR), and neural networks (NN), as recommended previously.[27] In this cross validation procedure, the dataset was randomly divided into 10 sets independent from the group structure. Each part of the dataset was left out for the calibration and then was predicted from this calibration. The total number of prediction errors of the 10 sets that were discarded step by step is given in percent (error rates).

RESULTS AND DISCUSSION

Analyzing weak variations of metabolic processes resulting from physiological perturbations in the context of a conventional hormonal treatment using anabolic steroids was assessed in a case validation of the Py-MAB-TOFMS fingerprinting procedure. Results presented hereafter detail such an approach in an effort to underline quantitative physiological variations by statistical classification of spectra obtained from the urine of animals submitted to an experimental design in which physiological disruptions are clearly controlled.

Analysis of Ionization Selectivity and Fingerprints. From a typical bovine urinary sample, several mass spectra were recorded in order to investigate the ionization selectivity controlled by different metastable reagent gases (Figure 1). Using N_2* as the ionization reagent, Py-MAB-TOFMS analysis of several reference compounds corresponding to urinary metabolites, that is, creatine, creatinine, hippurate, trimethylamine oxide, and urea, was conducted. Some of them, particularly hippurate, underwent dissociation processes and led to mass spectra displaying several fragment ions (not shown). Therefore, MAB fingerprints of urine can be considered as superimposed spectra of the different low-molecular-weight endogenous metabolites that were filtered by kidney glomerulus before urinary excretion. The main differences observed when increasing ionization energies (Table 1) tend to show that high-mass ions are more abundant in the Xe* fingerprint (Figure 1a, 8.3 eV) than in the Kr* fingerprint (Figure 1b, 9.9 eV). In addition, the total ion current (TIC) is multiplied by a 3.4 factor when using Kr*, TIC values varying from 1.3×10^3 for Xe* (Figure 1a) to 4.4×10^3 for Kr* (Figure 1b). This effect can be attributed to different ionization yields in a first attempt. Thus, the qualitative as well as quantitative variations of the fingerprints obtained from different energy-tunable gases can be considered as the result of two main features of MAB ionization: (i) selective ionization of metabolites involved in a Penning reaction of type I, and (ii) fragmentation processes specifically encountered in a Penning reaction of type II.[41] Faubert et al.[41] observed that the fragmentation extent of anisole 2-butanone under EI (70 eV) was similar or situated between the most energetic MAB ionization gases, that is, Ne* and He*, meaning that similar mean energy transfer was involved in the three cases. The lower fragmentation extent observed at lower EI energies with respect to MAB ionization was attributed to a less efficient energy transfer of electrons, as compared to that of metastable atoms. Using low-energy MAB ionization gases such as N_2* or Xe*, the production of molecular ions is favored, whereas the fragmentation extent is expected to be lower. Thus, for a given compound, the effect of ionization selectivity vs. the fragmentation extent can be readily identified using the appropriate MAB gas beam. In our case, since our samples were not pure reference compounds but complex biofluids, these two components were undecipherable, and differences observed in urinary fingerprints were the result of both the distribution of ionization potentials and the fragmentation energies of each component of the urine present in the fingerprint. In other words, the gain in TIC observed with Kr* can be attributed to differential ionization selectivity, whereas the shift of the fingerprint toward higher masses may arise from a lower fragmentation extent obtained using Xe*.

Aiming at producing variables as close as possible to the chemical entities of interest, that is, the endogenous metabolites, low ionization energy gases are preferred in order to reduce the fragmentation extent of molecules (Table 1). Although Xe* holds the lowest ionization energy among ionizing gases, the presence of Xe isotopic cluster on the obtained mass spectra could be of detrimental effect on the statistical analyses. Therefore, N_2* appeared to be a suitable ionizing agent, since $N_2^{+\bullet}$ ions at m/z 28 are not collected (Figure 2).

Another aspect to be investigated was the potentiality of fingerprints to hold some discriminant information characterizing the different groups of animals (C, 1, 2, 4). Using N_2* metastable gas, two urine samples issued from different treatments (groups C and 4) were analyzed by Py-MAB-TOFMS. After averaging and background-subtracting on appropriate portions of the pyrogram, sample variability led to qualitatively different fingerprints (Figure 2). The control urine sample was mainly characterized by m/z 121, 108, 105, and 60 ions (Figure 2a), whereas the 4-implant urine

(48) Ripley, B. D. *Pattern Recognition and Neural Networks*; Cambridge University Press: Cambridge, 1996.

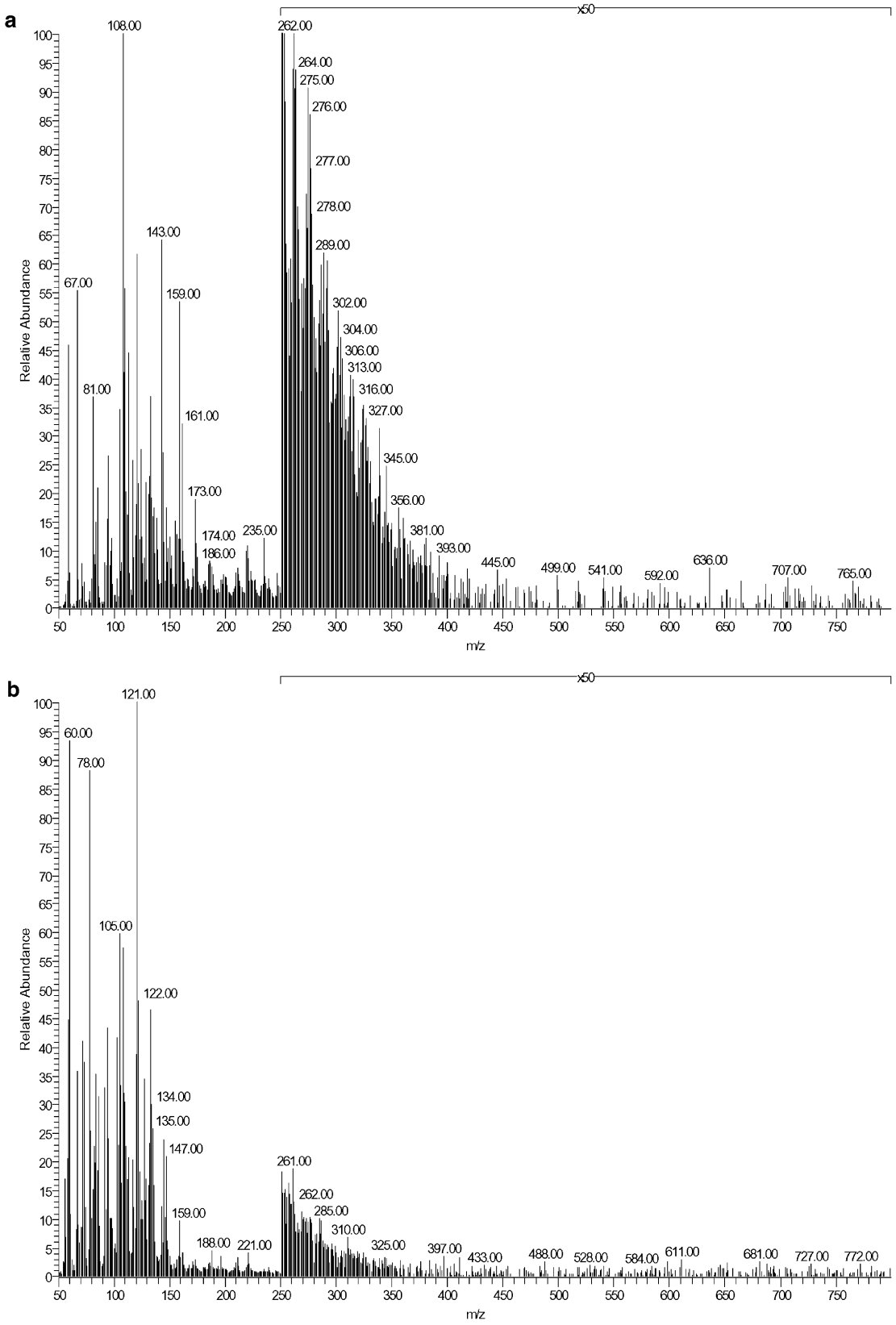

Figure 1. Py-MAB-TOFMS fingerprints of the same urine sample using different ionization agents: (a) Xe* (8.32 eV); TIC, 1.29 × 10^3; and (b) Kr* (9.92 eV); TIC, 4.37 × 10^3.

sample (Figure 2b) displayed m/z 121, 59, 108, 133, 60, and 105 ions as major peaks. The method enabled us to compare animals at a qualitative level by revealing the main characteristic ions a posteriori. Nevertheless, it was not ruled out that minor peaks could also hold variations characteristic of the hormonal treatment used. Moreover, one could also argue that some of the differences observed between spectra remain uncertain as a result of animal specificity or because of instrumental variations.

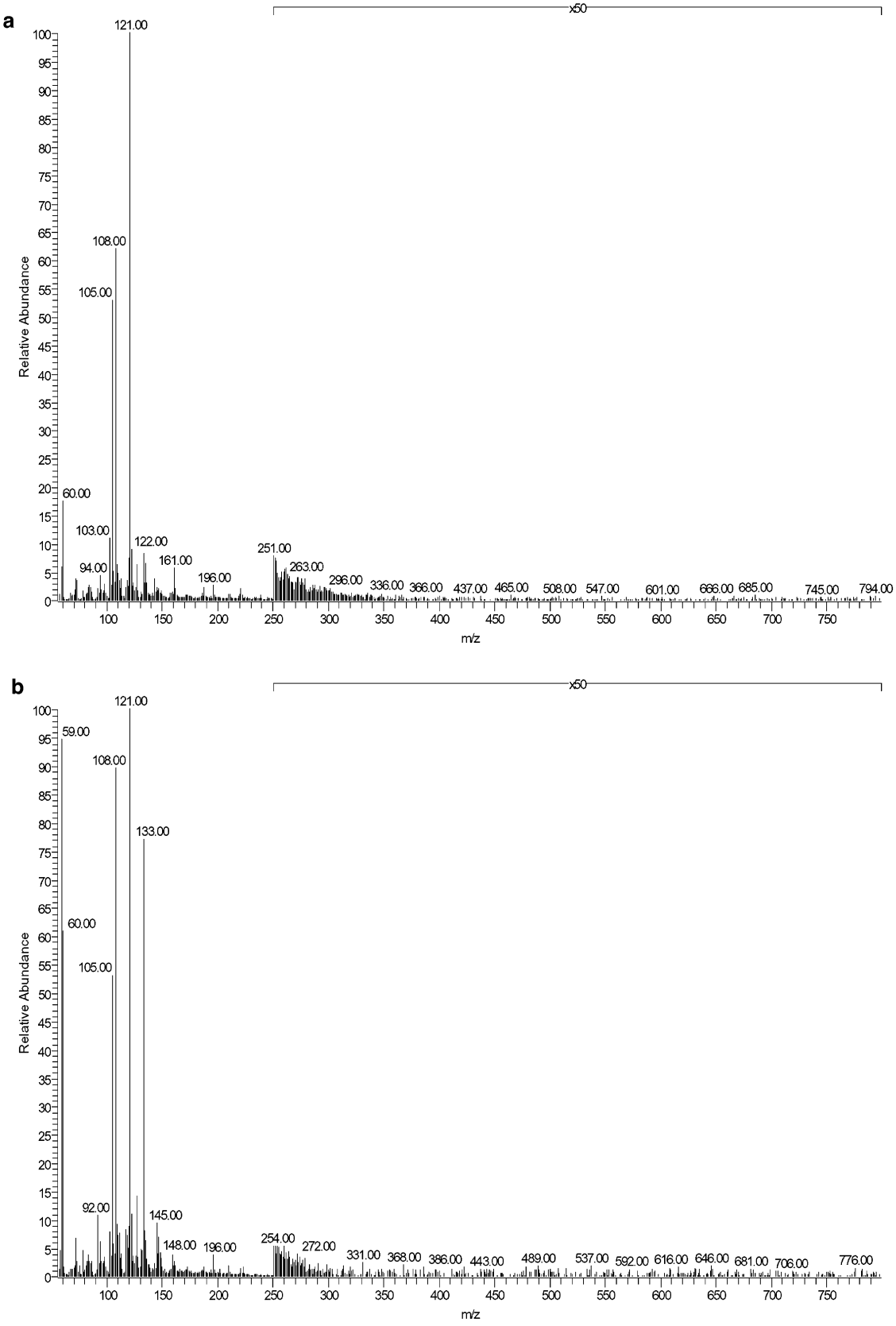

Figure 2. N_2 (8.5 eV) MS fingerprints of urine samples. The major peaks are given by decreasing value: (a) control steer showing *m*/*z* 121, 108, 105, and 60 as major peaks; (b) treated steer implanted with four implants containing TBA and E_2 displaying *m*/*z* 121, 59, 108, 133, 60, and 105 as major ions.

Reproducibility and Reliability. To quantify confidence intervals of the fingerprinting device, some causes of irreproducibility were tested. Both the deposit location on the quartz capillary and the quantity of matrix deposited directly affect the pyrogram

repeatability, but not the average spectrum, meaning that desorption of compounds occurs under roughly similar conditions. Yet, for enhancing the pyrogram reproducibility as well as the detection sensitivity, a helium flow (2 mL/min) drives the pyrolysate from the quartz capillary into the MAB ionization volume. This leads to an increase in the total ionic current by almost 1 order of magnitude (not shown). Limited degradation was ensured by quickly increasing the pyrolyzer temperature under inert atmosphere conditions, using small sample amounts (0.2 μL of a solution at 50 mg/mL), and fast removal of pyrolysis products to reduce secondary reactions.[39] In fact, the helium flow we used holds two main characteristics: an inert atmosphere and a rapid exporting of the desorbed compounds. Such an instrumental development and deposit reproducibility in quantity and topology were sought to protect pyrolysates against extensive and uncontrolled secondary fragmentation inside the pyrolysis quartz tubes. Since repeated mass fingerprints are clearly similar (not shown), a quantitative study of the instrumental reliability and reproducibility was undertaken by the mean of a statistical analysis.

For checking the reliability of the mass variables, 36 mass spectra of the same urine sample were recorded. Dispersion parameters, including mean (μ), standard deviation (SD), and variation coefficient (VC), were computed for each m/z ratio (Figure 3). The class frequency distribution of VC shows a bimodal distribution, revealing two VC populations (Figure 3a). One-half of the values were situated below a VC threshold value of 0.9−1, which is a relatively high value. The VC calculated for every nominal mass variable clearly shows two different regions with a significant threshold around 350 Th (Figure 3b). Low nominal masses present low VCs, and then can be considered as quite repeatable and reliable variables, whereas variables over 350 Th display higher VCs and seem to correspond to rather noisy ones. Moreover, lower m/z ratios show condensed VCs, as compared to higher m/z ratios, with dispersed ones. Indeed, as a quotient between SD and μ, the VC is particularly sensitive to low μ values. Then the relationship between μ and SD values was investigated for every m/z ratio (Figure 3c). Here, SD values are linked to μ values by a bilogarithmic model. Because of the algebraic link, a mathematical correction was used to stabilize variances. Therefore, as is usually done, a classical variable transform can be achieved by using the slope k of the regression line between μ and the SD values calculated for every m/z ratio as $y = x^{1-k}$. Consequently, after such a variable transformation, whatever the m/z ratio considered, the SD values become independent from the μ values.

Therefore, with such transformed variables, the frequency distribution is more contracted (Figure 3d): ~300 variables hold VC values comprised between 0 and 0.15. These variables with reduced VC values correspond to m/z ratios lower than 350 Th (Figure 3e). The obvious conclusion is that transformed variables corresponding to low m/z ratios are more reproducible. To disregard the effect of low mean values on a VC, a specific analysis of the SD obtained on transformed variables was performed, considering the m/z ratio values (Figure 3f). Two characteristic patterns were observed: (i) the first one displayed variables with low SD values in general (around 0.02) that correspond to low m/z ratios, (ii) the second one, for variables over 350−400 Th, revealed a constant SD value around 0.05.

In the MAB fingerprints of urine, fragment ions over 350 Th are either seldom detected or show really low intensities (Figure 2). As a result, we can hypothesize that only 300 ions comprised between 50 and 350 Th are produced from the biological matrix. This could also explain why the first 300 m/z ratios display low VC and SD values, positively resulting from a structured signal corresponding to well-detected analytes. Because of the influence of the matrix background noise, the higher m/z values display higher SD and VC values. Such high nominal mass variables with high VC have a reduced probability to reveal some relevant information for fingerprint discrimination, seemingly due to the high background noise compared to low abundance variables. So, low-molecular-weight analytes really determine the fingerprint pattern and are, therefore, quantitatively reliable.

The discrete ionization energy provided by the quantum transfer during collision with metastable atoms in MAB ionization mode differs strikingly from the average energy involved in EI and other classical ionization modes. This could explain such a suitable repeatability of the MAB mass spectra.[41] Low and repeatable energies involved in MAB ionization allows us to obtain reproducible ionization and fragmentation processes in Penning reactions, even in the case of complex biological matrixes, as evidenced previously.[42,49] As a result, Py-MAB-TOFMS performed on metabolites provides an alternative to genomic and proteomic procedures that are mostly semiquantitative in the case of MALDI,[5] although labeling-based quantitative proteomics by MALDI-Q-Tof[37] and LC/ESI-FT-ICR[50] MS have recently been developed. Furthermore, one of the main advantages of Py-MAB-TOFMS lies in its ability to minimize the sample preparation.

Hierarchical Analysis of Correlations between Fragmentation Processes and Physiological Origin. Although fragmentation processes can be readily controlled under MAB ionization conditions, their occurrence cannot totally be ruled out. Moreover, isotopic clustering can also be responsible for variations of the measured signal on a given m/z ratio. This justifies a study of the redundancy between m/z ratios by hierarchical classification of variables.[46] Variables are aggregated by correlation values (Figure 4). In both classifications, molecular and fragment ions that are selected correspond to low values of the metric $1 - r_{ij}$. This is explained by a high correlation between these selected variables and their neighboring ones, the latter being not selected to perform discriminant analysis. Variables displaying metric $1 - r_{ij}$ values higher than 0.25 were not selected. On the right side of the dendrogram shown in Figure 4a, most of the variables seem to correspond to noninformative variables, except those corresponding to the m/z 73 and 59 ions, which can be attributed to the $M^{+\bullet}$ and $(M - 16)^+$ ions of the trimethylamine oxide, respectively. The m/z values of these noninformative variables are higher than 350 and should correspond to spectrometric noise, as mentioned above. On the basis of the mass spectra obtained from reference compounds (creatine, creatinine, hippurate), the

(49) Wilkes, J. G.; Holcomb, M.; Rafii, F.; Letarte, S.; Bertrand, M. J.; Colby, S. *Proceedings of the 46th ASMS Conference on Mass Spectrometry and Allied Topics*, Orlando, FL, May 31−June 4, 1998; p 1119.

(50) Conrads, T. P.; Alving, K.; Veenstra, T. D.; Belov, M. E.; Anderson, G. A.; Anderson, D. J.; Lipton, M. S.; Pasa-Tolic, L.; Udseth, H. R.; Chrisler, W. B.; Thrall, B. D.; Smith, R. D. *Anal. Chem.* **2001**, *73*, 2132−9.

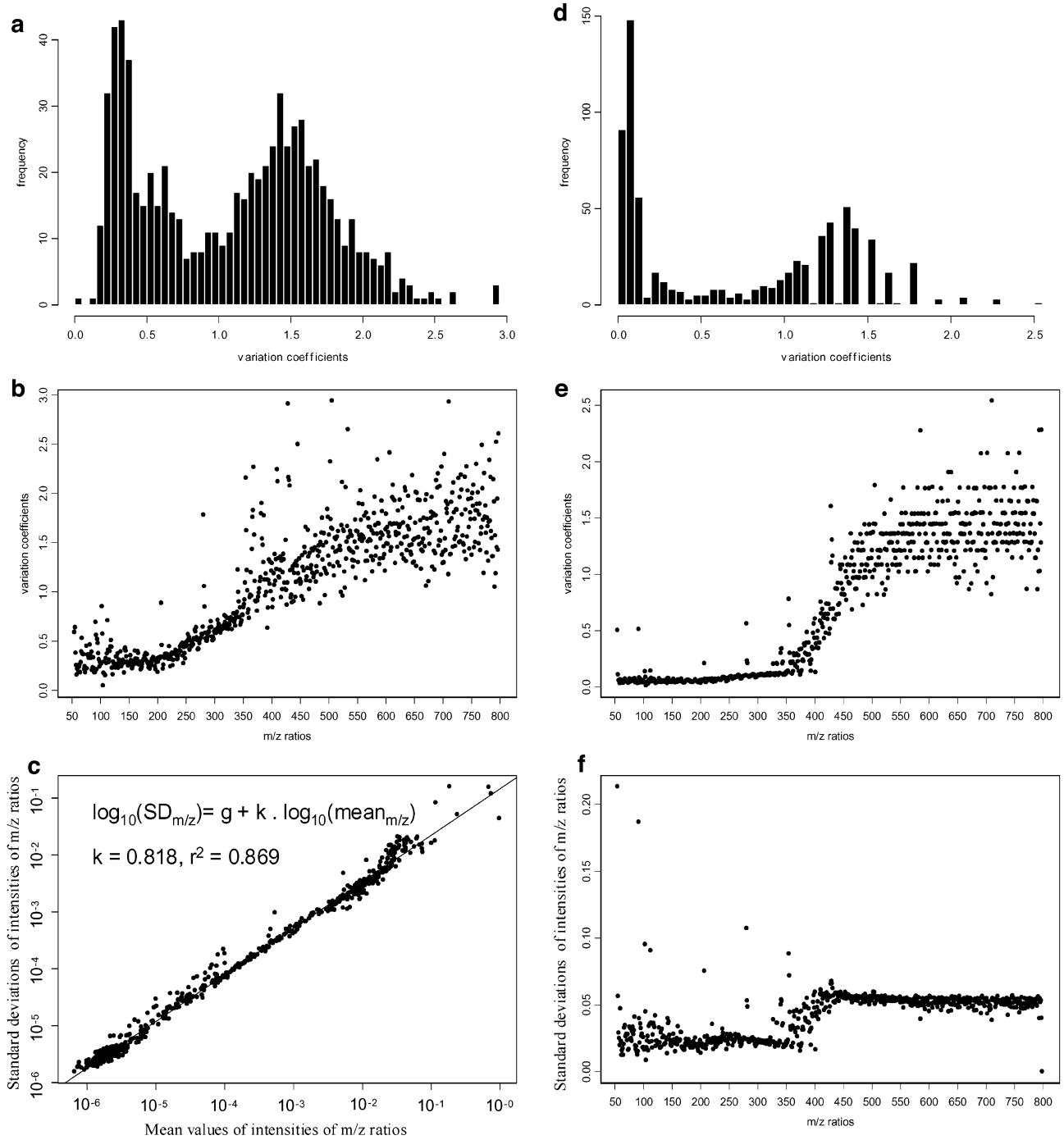

Figure 3. Reliability study for *m*/*z* variables issued from Py-MAB-TOFMS measurements: (a) class frequency distribution of variation coefficients, (b) variation coefficient values for every mass variable, (c) regression between SD and μ values for every nominal *m*/*z* ratio given in a bilogarithmic scale leading to the variable transform, (d) class frequency distribution of variation coefficients for transformed variables, (e) variation coefficient values for every transformed variable, and (f) standard deviation values for every transformed variable.

selected ions belong to a reduced number of clusters or metaclusters.

When noisy variables ($m/z > 350$) were discarded for the variance study, the selected variables were regularly distributed inside nine different clusters or metaclusters (Figure 4b). The first variables aggregated in the 55−350 Th range at the bottom of the dendrogram are highly correlated, whereas the last clusters at the top of the dendrogram are highly anticorrelated. Values around 1 correspond to noncorrelated variables. In analysis of NMR fingerprints,[46] this kind of hierarchical classification of variables is known to underline two features: (i) a high positive correlation is interpreted as a 2D-NMR molecular connectivity, thanks to NMR assignments; (ii) low positive or negative correlation could reveal some physiological or metabolic correlation. From this information, we can infer that (i) most of the highly positive correlations observed in this dendrogram may describe redundancy issued from fragmentation or pyrolysis steps and (ii) weaker positive or negative correlations could be more probably linked to a physiological coordination, which appears to be not well-determined using this clustering method.

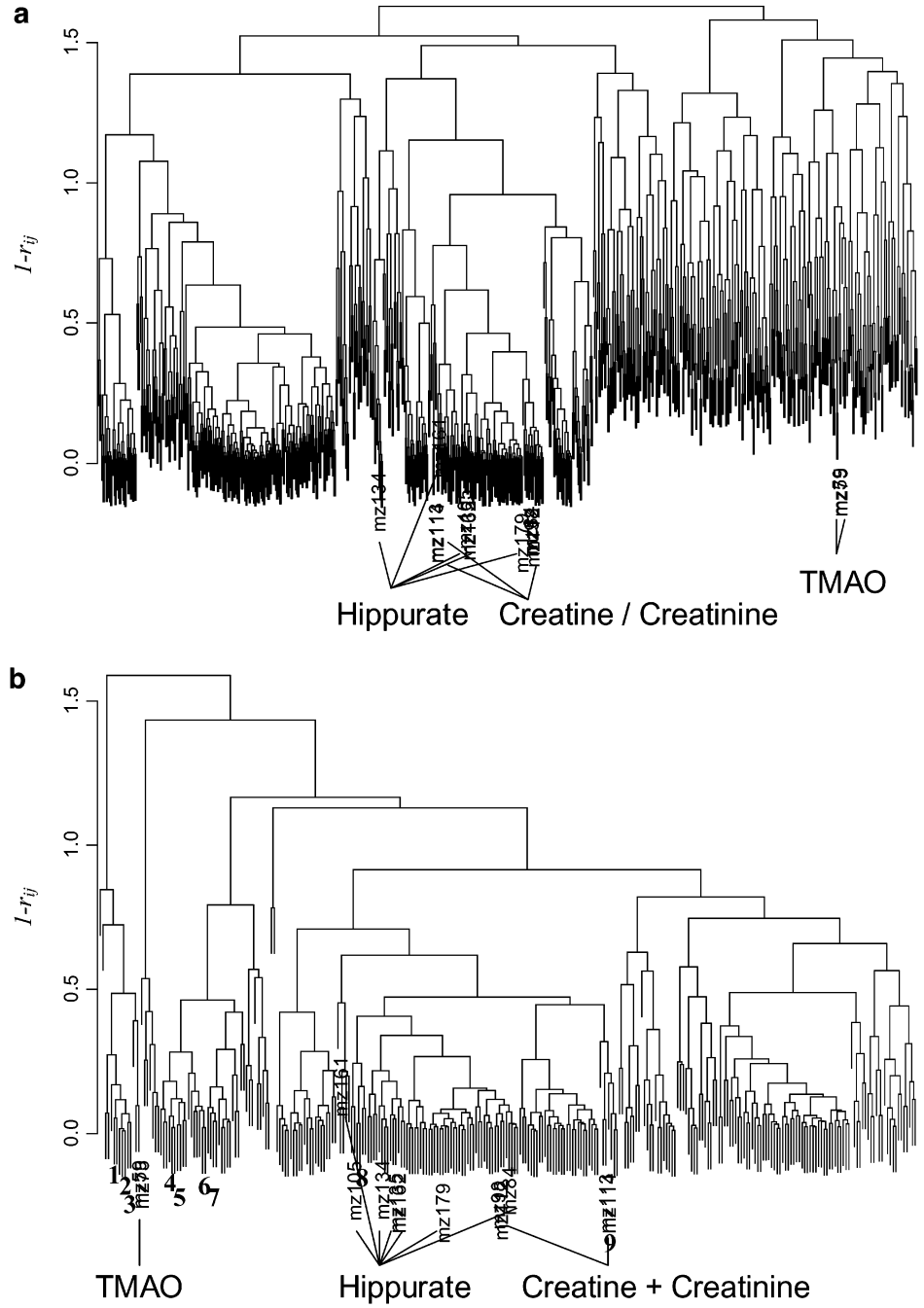

Figure 4. Variable clustering analysis of N_2* MAB-TOFMS fingerprints: (a) dendrogram of 745 variables transformed using $y = \arcsin(x^{1/2})$; a typical noisy area is displayed on the right side of the tree corresponding to *m*/*z* ratios over 350 Th with no evident structured aggregation of those variables; (b) dendrogram of the 296 variables transformed using $y = x^{1-k}$ with $k = 0.818$; labels from 1 to 9 correspond to selected variables with $(P, \beta, n) = (10^{-2}, 1, 9)$. Note the regular distribution of selected variables along the clusters and metaclusters showing the efficiency of filtration of redundant variables. In the two dendrograms, the *m*/*z* ratios corresponding to molecular or fragment ions of TMAO, hippurate, creatine, and creatinine analyzed separately as referenced compounds are shown.

Efficiency of Pattern Recognition Segmentation Applied to N_2* Py-MAB-TOFMS Fingerprints. Statistical performances of pattern recognition methods were usually assessed by computing error rates using cross-validation procedures. To avoid noise introduction and weak performance of classification, variables were selected by ANOVA on the basis of their response to the hormonal treatment and their ability to introduce significant but nonredundant information into the system. We used an algorithm based on partial correlations between variables that are significantly affected by hormonal treatment ($P \leq 10^{-2}$) in order to select them. Those selected variables were used to train pattern recognition methods. From our dataset, the number of variables taken from the selection sequence strongly influenced performance results (Table 2). For instance, NN error rates changed strikingly when adding a supplementary variable without any clear tendency, even with more than 10 variables. Optimal parameter sets corresponded to those minimizing the error rate. A significant difference in terms of efficiency was observed between LDA regularly reaching 0%

Table 2. Pattern Recognition Error Rates by a 10-Fold Cross Validation on MAB Fingerprints with N_2[a]

pattern recognition method	Error rates (%) calculated in function of the number of selected variables														
	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
LDA	10.0	6.7	3.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
k-nn	10.0	13.3	10.0	6.7	10.0	13.3	10.0	13.3	13.3	13.3	13.3	13.3	10.0	10.0	10.0
MLR	13.3	20.0	10.0	16.7	13.3	16.7	16.7	16.7	16.7	20.0	20.0	16.7	16.7	10.0	10.0
NN	20.0	23.3	13.3	33.3	20.0	10.0	20.0	6.7	10.0	20.0	13.3	10.0	16.7	10.0	13.3
LVQ	23.3	16.7	16.7	26.7	20.0	13.3	16.7	13.3	20.0	20.0	23.3	20.0	23.3	13.3	20.0

[a] Variables selected with optimal parameters $(P, \beta) = (10^{-2}, 1)$.

error rates, and *k*-nn, MLR, NN, and LVQ that are less efficient. One could sort the efficiency of these methods as following: LDA > *k*-nn > MLR > NN > LVQ, the latter method being the worst to classify test or unknown samples. In addition, the $y = x^{1-k}$ variable transform giving a variance stabilization between mass ratio variables is the most efficient to give better performances by CV, all other variable transform procedures being less performing (not shown).

Statistical Analysis of N_2* Py-MAB-TOFMS Fingerprints by LDA. Statistical filtering of this set of quantitative mass spectra allowed us to point out a few meaningful variables. Among pattern recognition methods, LDA was evidenced as the most efficient method in terms of sample classification (Table 2). With minimal prediction error rates, we optimized between-group variance segmentation by LDA for graphical exploration of groups projected on factorial maps (Figure 5 and Table 2). This representation was reached using nine variables extracted from the 296 initial ones in the m/z ratio range between 55 and 350 Th (Table 2). As evidenced on the dendrogram, these were regularly and nonredundantly distributed within the clusters (Figure 4b).

The first linear discriminant, corresponding to 56% of the between-group variability segregated control animals from treated ones ($P < 10^{-4}$) (Figure 5). Canonical correlations describe links between variables and discriminant axes (Table 3). Since all canonical correlations to LD1 are negative, the first axis corresponds to a size factor. Then, LD1 expresses the fact that relative abundance of m/z ratios is increased by the global anabolic response. More precisely, the variables m/z 133 ($r = -0.99$), 132 ($r = -0.87$), and 146 ($r = -0.83$) show the highest absolute correlation ranks, and can therefore be considered to be related to the anabolic response (Table 3). The effect of the anabolic treatment was tested by ANOVA on these variables as highly significant (m/z 133, $P_{\text{treatment}} = 2 \times 10^{-15}$; m/z 132, $P_{\text{treatment}} = 5 \times 10^{-8}$; m/z 146, $P_{\text{treatment}} = 3.3 \times 10^{-6}$). A one-way ANOVA and the subsequent mean comparison using the Student−Neuman−Keuls (SNK) test were performed on the 4 groups of animals. Variables m/z 133 and 146 display level D in controls, which is significantly lower than levels A, B, and C measured when the animals were treated by one, two, or four implants, respectively. In the treated groups, these variables describe a proportional response to the treatment that is not the case for m/z 132 that do not underline a simple proportional response linked to the number of implants used (Table 3).

LD2 axis, accounting for 23% of the between-group variability ($P < 10^{-4}$), was also a size factor, revealing a dose-proportional response to the anabolic treatment (Figure 5a). This dose response is described by a set of variables, such as m/z 81 ($r = 0.81$), 69 ($r = 0.76$), and 130 ($r = 0.64$) showing highly positive correlations with this second discriminant function. These variables were mainly encountered in urine collected on steers treated by low doses of anabolics, but not in urine coming from animals treated with higher doses (Table 3). ANOVA and SNK tests yielded similar results for these variables (m/z 81, $P_{\text{dose}} = 2.3 \times 10^{-5}$; m/z 69, $P_{\text{dose}} = 4.3 \times 10^{-4}$; m/z 130, $P_{\text{dose}} = 7.7 \times 10^{-8}$) and also for variables responsible for the LD1 effect (m/z 132, $P_{\text{dose}} = 2.6 \times 10^{-15}$; m/z 146, $P_{\text{dose}} = 8.9 \times 10^{-12}$).

The last 21% of variability on LD3 segregated treated animals belonging to groups 1 and 4 from those of group 2 ($P < 10^{-4}$) (Figure 5b). What made group 2 different is the second hormonal stimulation on the 45th day by administration of another implant, whereas groups 1 and 4 were stimulated just once on the first day of the experiment. This discriminant axis reveals the multiplicity of stimulation of the anabolic treatment. The double stimulation response is mainly explained by m/z 113 ($r = 0.77$; $P_{\text{dose}} = 5.1 \times 10^{-6}$), m/z 109 ($r = 0.43$; $P_{\text{dose}} = 4.6 \times 10^{-8}$), 81 ($r = 0.33$; $P_{\text{dose}} = 2.3 \times 10^{-5}$) and 196 ($r = 0.32$; $P_{\text{dose}} = 5.2 \times 10^{-4}$) (Table 3). SNK tests revealed such a contrast in a complementary way: for animals treated with two implants, the variable m/z 113 was significantly raised (level A), as compared to all the other groups (level B). This contrast was also evidenced to a lesser extent for m/z 109 and 196 for animals receiving one or two implants. Three different physiological hypotheses can account for this long-lasting physiological effect: (i) an independent dose effect in anabolic treatment, (ii) impregnation by the first implant enhancing the effect of the second one administered on the 45th day, and (iii) implant delivery kinetics providing higher concentrations of steroids in plasma on the 90th day of implantation for animals treated with two implants.

Accuracy of the variable filtration step and efficiency of the pattern recognition device allow us to consider that m/z 133 is a molecular ion or a fragment ion of special interest as far as the physiological response of steers to 17β-estradiol and trenbolone acetate treatment is concerned. This was already presumed in a qualitative analysis of fingerprints (Figure 2). Nevertheless, correlations to discriminant axes (Table 3) show that the other mass variables selected in the filtration step have also significant and complementary contributions for building discriminant axes. This is particularly true for weaker effects, such as the dose response or the response linked to the multiplicity of hormonal stimulation, which was not underlined by major peaks in fingerprints (Figure 2). From the nine original ions significantly affected by the experimental design and extracted by the variable filtration algorithm, LDA attributed m/z 133, 132, and 146 to the signature

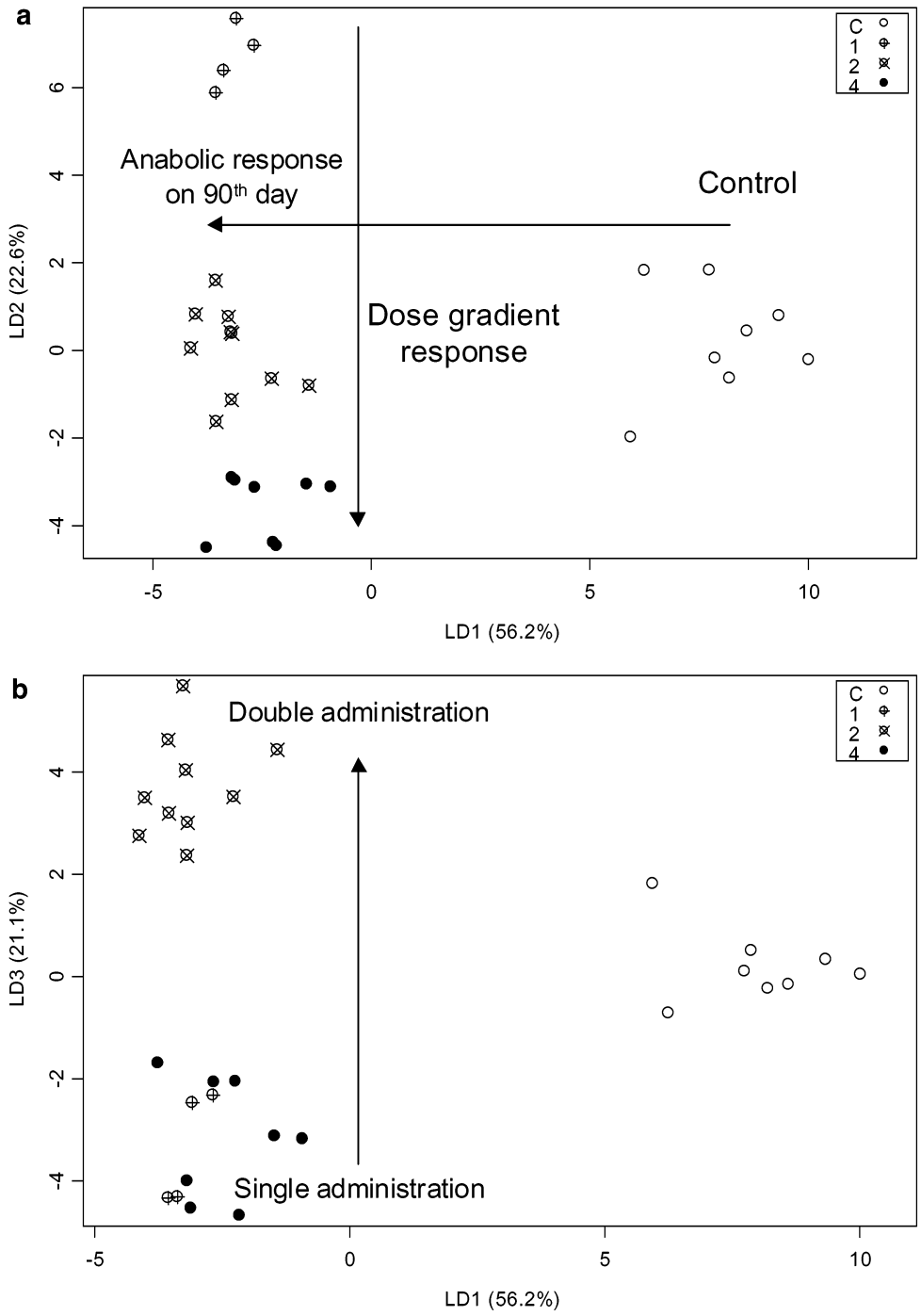

Figure 5. LDA 2D-factorial plots: (a) LD1 × LD2 and (b) LD1 × LD3. Nine variables issued from filtration were used to perform classification of individuals.

of the anabolic treatment (control vs treated animals), m/z 81, 69, and 130 to the dose-proportional response to the anabolic treatment and m/z 113, 109, 81, and 196 to the segregation between groups of animals treated by a single implantation from those receiving a second implant on the 45th day.

A set of representative standard reference compounds (urea, creatine, creatinine, trimethylamine-*N*-oxide, and hippuric acid) was checked under N_2* ionization conditions (data not shown). It is noteworthy that the m/z 113 ion was observed to be the main ionic species from the analysis of both creatinine ($M^{+\bullet}$, m/z 113) and creatine ($[M - H_2O]^{+\bullet}$, m/z 113). We could then hypothesize from these results that the creatinine and creatine metabolic pathway could be affected by anabolic treatment, which was previously reported.[51]

Factorial decomposition involved in LDA reveals the complexity of this simple biological contrast. The three linear discriminant functions explain well a differential segmentation of the MAB fingerprints by decreasing order of heterogeneity. These factors have obvious physiological interpretations: (i) response to anabolic treatment, (ii) dose response, and (iii) a hormonal effect linked to a second stimulation effect. Similar effects have also been underlined on this same set of urinary samples using quantitative

Table 3. Canonical Correlations, ANOVA, and SNK Mean Grouping Test of Discriminant Metabolites from LDA Performed on Nine *m/z* Ions To Reveal the Dose Effect on Males after a 90-Day Period of Implantation[a,b]

variables (m/z ratio)	canonical correlation (r)						ANOVA					
	LD1		LD2		LD3				SNK[c] test			
	r	$\|r\|$ rank	r	$\|r\|$ rank	r	$\|r\|$ rank	$P_{treatment}$	P_{dose}	C	1	2	4
132	−0.87	2	0.43	7	0.23	6	4.98×10^{-2}	2.55×10^{-15}	D	A	B	C
133	−0.99	1	0.09	9	0.06	8	2.00×10^{-15}	8.77×10^{-15}	C	A	A	B
113	−0.63	7	0.11	8	0.77	1	8.22×10^{-3}	5.18×10^{-6}	B	B	A	B
130	−0.77	4	0.63	3	−0.01	9	1.50×10^{-4}	7.74×10^{-8}	D	A	B	C
109	−0.75	5	0.50	6	0.43	2	2.80×10^{-4}	4.56×10^{-8}	B	A	A	B
196	−0.74	6	0.59	4	0.32	4	6.14×10^{-3}	5.19×10^{-4}	B	A	A	B
69	−0.61	8	0.76	2	0.22	7	5.47×10^{-2}	4.26×10^{-4}	B	A	AB	B
81	−0.48	9	0.81	1	0.33	3	7.28×10^{-2}	2.29×10^{-5}	C	A	B	C
146	−0.83	3	0.51	5	0.24	5	3.31×10^{-6}	8.91×10^{-12}	D	A	B	C

[a] Effects of treatment and dosage are tested by analysis of variance. [b] Bold characters correspond to main correlations or probabilities. [c] Student-Newman-Keuls (SNK) means that are noted by the same letters are not significantly different for an α risk of 5%. Letters A, B, C, and D are sorted by decreasing mean order.

variables generated by another method, that is, $^{1}H-^{13}C$ HMBC/NMR (unpublished results). These results point out the power of such a statistical analysis of fingerprints, able to evidence similar treatment modalities, thanks to the experimental design, whatever the spectrometric fingerprinting device used.

CONCLUSION

The new Py-MAB-TOFMS technology provides repeatable mass fingerprints on biofluids. This property seems to be linked to the discrete and tunable energy of the metastable beam of rare gases necessary to produce fingerprints. With performance improvements and sample uptake automation,[52] Py-MAB-TOFMS technology appears to be very encouraging for generation of metabonomic databases necessary to perform screening of drugs, xenobiotic use or to detect pathologies or physiological disruptions.[16] Indeed, conditions we have employed were proved to be relevant and reliable enough to perform valuable chemometric analyses. Quantum energy transfer involved in the MAB ionization process enables a discrete and tunable control of the internal energy of molecular ions, thus leading to reproducible dissociation processes and reproducible fingerprints. Such reproducibility studies and a posteriori variable transforms seem to be useful and should be performed before analyzing each new dataset or at least each new sample matrix. Statistical pattern recognition algorithms successfully use the potential mass spectra information inherently present in the fingerprint. Thanks to statistical interpretation of the metabolic variations induced by steroids, a clear description of the physiological disruption can be obtained. Nevertheless, structural identification of the ions responsible for such discrimination has to be performed to produce comprehensive structural data. A classical approach to obtain this structural information is deconvolution of mixture spectra.[53] However, this could be also achieved by performing Py-MAB/MS/MS experiments on characteristic samples. Because of performances and interpretations similar to NMR ones (unpublished results), the underlying question we can address is related to the statistical correlation between datasets generated with different variable generators, such as NMR and Py-MAB-TOFMS or between datasets obtained by Py-MAB-TOFMS using different ionization gases. This global correlation between different datasets could be at the origin of some converging clues for obtaining a structural identification of the relevant metabolites involved in metabonomic discriminations.

(51) Van Pilsum, J. F.; Ungar, F. *Arch. Biochem. Biophys.* **1968**, *124*, 372−9.
(52) Wilkes, J. G.; Sutherland, J.; Rafii, F.; Voorhees, K. J.; Warden, C.; Hall, R.; Freeman, J. P.; Lay, J. O., Jr. *Proceedings of the 44th ASMS Conference on Mass Spectrometry and Allied Topics*, Portland, OR, May 12−16, 1996; p 589.
(53) Halket, J. M.; Przyborowska, A.; Stein, S. E.; Mallard, W. G.; Down, S.; Chalmers, R. A. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 279−84.

ACKNOWLEDGMENT

Dr. D. Maume and Dr. B. Le Bizec are thanked for helpful discussions, instrumental advice and providing samples. Dr. P. Martin, from Dephy Technologies Inc., is thanked for Py-MAB-TOF experimental supply. This work was supported by a grant from the European Commission (B6-7920/98/000823).

Received for review March 26, 2002. Accepted August 7, 2002.

AC025656K