CHEMISTRY
A EUROPEAN JOURNAL

DOI: 10.1002/chem.200801181

Expressing Forest Origins in the Chemical Composition of Cooperage Oak Woods and Corresponding Wines by Using FTICR-MS

Regis D. Gougeon,*[a] Marianna Lucio,*[b] Arnaud De Boel,[a] Moritz Frommberger,[b] Norbert Hertkorn,[b] Dominique Peyron,[c] David Chassagne,[a] François Feuillat,[d] Philippe Cayot,[e] Andrée Voilley,[e] Istvan Gebefügi,[b] and Philippe Schmitt-Kopplin*[b]

Abstract: A non-targeted, ultra-high-resolution mass spectrometric, direct analysis of oak-wood extracts from two species (*Quercus robur* L. and *Quercus petraea* Liebl.) from three French forests, and of a wine aged in barrels derived therefrom has been performed to identify families of metabolites that could discriminate both the species and the geographical origin of woods. From 12 T ultra-high-resolution Fourier transform ion cyclotron resonance mass spectra of wood extracts, hundreds of mass signals were identified as possible significant biomarkers of the two species, with phenolic and carbohydrate moieties leading the differentiation between *Q. robur* and *Q. petraea*, respectively, as corroborated by both FTMS and NMR data. For the first time, it is shown that oak woods can also be discriminated on the basis of hundreds of forest-related compounds, and particular emphasis is put on sessile oaks from the Tronçais forest, for which sugars are significantly discriminant. Despite the higher complexity and diversity of wine metabolites, forest-related compounds can also be detected in wines aged in related barrels. It is only by using these non-targeted analyses that such innovative results, which reveal specific chemodiversities of natural materials, can be obtained.

Keywords: analytical methods · chemodiversity · mass spectrometry · NMR spectroscopy · *Quercus* species

Introduction

Initially aimed at serving as suitable wine containers, oak barrels have today become a practical means of modulating the fine sensory characteristics of wine.[1] Several studies have revealed the influence of oak wood on the organoleptic properties of wines matured in oak barrels.[2–6] This influence is considered to be due to variations in the physical and chemical properties of oak, which mainly depend on the geographical origin and the species.[7–9] So far, attempts to establish correlations between the chemical properties of oak wood and the origin or species have relied on targeted analyses of selected compounds. These studies in particular revealed significant species effects; for instance, it is recognised that among the two predominant western European oak species, *Quercus robur* L. (pedunculate oak) has larger ring widths and contains more ellagitannins than *Quercus petraea* Liebl. (sessile oak), which in contrast generally contains more volatile compounds, such as *cis*- and *trans*-β-methyl-γ-octalactones (whisky lactones), eugenol, vanillin and furfural, although discrepancies can be found in the literature.[7–12] A similar trend, but restricted to ellagitanins and whisky lactones, has been generalised for eastern European pedunculate and sessile oaks.[8] When considering more specifically aromatic whisky lactones, French sessile oaks are generally poorer than American white oaks (*Quercus Alba*) and eastern European sessile oaks.[8,13] Besides the species effect, the effects that forests could impose upon the chemical composition of oak wood, and ultimately on wines, have also been investigated.[2,4,7–9,14–16] These studies showed that forest effects on the chemical composition of wood are less pronounced than species effects, and significant discrimination, regardless of the species, could only be made between American and western and eastern European forests, on the basis of their eugenol, 2-phenylethanol, vanillin and syringaldehyde contents.[8] The huge inter-individual variability of the chemical composition of oak trees, even within a given species in a given forest, is actually the major acknowledged reason for the current absence of established significant correlations relating a forest and its oak wood composition, regardless of the species and location.[7,9,10,17] All these studies emphasise the major drawback of such targeted analyses; on the basis of only few chemical parameters, there are often atypical samples that do not fit in the discrimination.[18]

When considering the further chemical composition correlations that can be made between the geographical origin or the species of oaks and wines matured in related barrels, the only acknowledged generalisation is that the American white oak species introduces higher amounts of *cis*-whisky lactone into wines than the European sessile oak species.[1,2] *cis*-Whisky lactone is often mentioned as the major discrimi-

[a] Dr. R. D. Gougeon, A. De Boel, Dr. D. Chassagne
EA 581 EMMA
Institut Universitaire de la Vigne et du Vin—Jules Guyot
Université de Bourgogne, Rue Claude Ladrey
21078 Dijon (France)
Fax: (+33) 380-396-265
E-mail: regis.gougeon@u-bourgogne.fr

[b] M. Lucio, Dr. M. Frommberger, Dr. N. Hertkorn, Dr. I. Gebefügi, Priv.-Doz. Dr. P. Schmitt-Kopplin
Helmholtz Zentrum München
German Research Centre for Environmental Health (GmbH)
Institute for Ecological Chemistry, Ingoldstädter Landstrasse 1
85764 Neuherberg (Germany)
Fax: (+49) 893-187-3358
E-mail: marianna.lucio@helmholtz-muenchen.de
schmitt-kopplin@helmholtz-muenchen.de

[c] Dr. D. Peyron
Equipe de Psychologie Cognitive des Sens Chimiques
UMR CNRS/INRA/uB, Centre Européen des Sciences du Goût
Dijon (France)

[d] Dr. F. Feuillat
Office National des Forêts
Bourgogne (France)

[e] Prof. P. Cayot, Prof. A. Voilley
EA 581 EMMA, ENSBANA Université de Bourgogne
1 Esplanade Erasme
21079 Dijon (France)

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/chem.200801181.

nant compound because its content in wood correlates well with its content in wines aged in respective oak barrels, and also with the coconut, toasty or vanilla sensory descriptors of these wines.[19] In contrast, despite the abundance of heartwood ellagitanins and their solubility in wines, their concentration in oak-aged wines is generally lower than expected.[20] Therefore, in terms of chemical composition, no unambiguous forest effects of general validity on wines have been reported yet and effects on the chemical composition of cooperage oak woods have heavily relied upon the species-based identification of natural forests.[7]

Consequently, correlations between forest classification and wine aged in a barrel made of oaks from this forest are at best feeble. In addition, there is a multi-stage process between the cutting of oaks and the end of the barrel ageing period of wines. First, the wood staves undergo natural seasoning and are then toasted, which is designed to shape barrels. Both of these processes contribute to modulate the chemical composition of the wood[9,16,21] and subsequently of the wine.[22,23] However, although heating does form new compounds as a result of lignin and cellulose degradation, many heartwood constituents are barely or not at all affected by the heating intensity normally used and instead of eliminating the intrinsic variation between wood samples, heating would rather appear to complement it.[4] Secondly, several concurrent processes take place during the ageing period of wine.[1,6,24–28] This has been recently illustrated by Jarauta et al.,[6] who have identified at least seven processes responsible for the evolution of the 79 aroma compounds analysed in wines aged in oak barrels. These authors have confirmed that, in addition to the most studied extraction processes from the barrel, microbiological transformations, weak oxidation reactions enabled by the porosity of the container, condensation reactions and sorption to wood also modulate wine compounds during barrel aging. Another example of the complex mechanisms involved in wine chemistry related to barrel ageing has been provided by Quideau et al., who highlighted the fact that many ellagitannin derivatives would probably exist as a result of nucleophilic substitution reactions with wine-relevant nucleophiles.[29,30]

All these studies have fundamentally contributed to the knowledge of the chemical composition of oak wood as related to its species and to a lesser extent to its origin, and also to its impact on the composition and flavour of barrel-aged wine. However, as shown by Jarauta et al.,[6] most studies have failed to consider oak casks as a physically, chemically and biochemically active system. Oak wood itself is already a complex living system for which environmental conditions, such as the forest ecosystem where it has grown, may modulate its chemical composition as extensively as genetic diversity between species; genetic analyses have actually shown rather minor differentiation between the two species (*Q. robur* L. vs. *Q. petraea L.*).[31–33]

In 1998, a full-scale integrated study (Tonnellerie 2000) that initially involved nine French forests providing twelve lots of 24 trees (5 lots of pedunculate and 7 lots of sessile oak) was designed to evaluate the influence of both the geographic origin and the species of oak on the quality of wines matured in oak barrels.[34] We hypothesised that such sets of wood and wine samples would become unique panels of chemical compositions with little variation, and as such, ideal candidates for a non-targeted analysis of the correlations that could possibly exist between the species and/or forest origin of oak wood and wine aged in barrels made of this wood. It is most important to note that a rigorous review of the literature provides us with only one analytical report[17] and one sensory analysis by wood sniffing[35] on the Tonnellerie 2000 samples. These two studies gave consistent results in which sensory descriptors significantly discriminate sessile woods, in agreement with their significantly higher contents (on average) of whisky lactone. However, tannin content (based on the analysis of the few commonly found molecules) was less discriminant, though it showed the expected differences between species. Above all, these two studies, along with most of the previously related ones, emphasise the poorly investigated field of non-volatile extracts from oak wood, in which only the frequently found phenolic compounds are analysed.

Herein, we report the first non-targeted chemical characterisation approach by using organic structural spectroscopy/spectrometry, based primarily upon ultra-high-resolution Fourier transform ion cyclotron resonance (FTICR) mass spectrometric analysis[36] of cooperage oak wood non-volatile extracts and of a related wine, with the aim of drawing comprehensive chemical pictures that would allow significant correlations between these samples to be established. Within the field of chemistry, this approach not only addresses the as-yet unravelled specific chemodiversities of woods but also addresses possible acute correlations between chemodiversity and metabologeography.[37] It must be noted that recently, a similar non-targeted approach based on the "electronic tongue" analysis has been able to nicely discriminate wines with respect to the origin of oak barrels they were aged in.[38] However, these discriminations were only based on the high cross-selectivity of voltammetric sensors and provided no structural information on any active molecules involved on a molecular level.

Results and Discussion

Wood differentiation: Figure 1 shows an example of full mass spectrum of forest-averaged oak-wood extracts for the two species. Within the range m/z 150 to 1050 that was explored, these spectra exhibit several thousands of signals that represent all ionisable metabolites under the selected experimental conditions (electrospray negative mode). Although aqueous alcoholic solutions do not necessarily provide the best extracting efficiencies for non-volatile compounds, 5727 distinct signals were observed at S/N = 1 for the pedunculate species (**P**), of which 1045 that could be assigned to elemental formulae that contain CHONS. Similarly, 7677 resolved signals were observed for the sessile species (**S**), with 1562 assignments of elemental formulae. A cu-

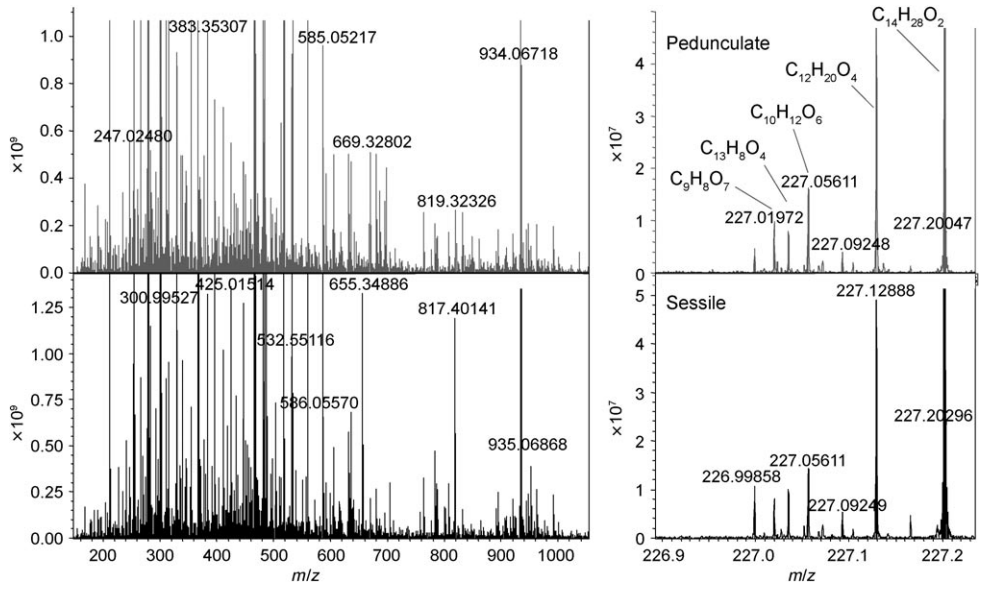

Figure 1. Left: Typical negative ion mass spectrum of the pedunculate (top) and sessile (bottom) wood extract samples (extracted averaged wood sample of each species). Right: Detail of *m*/*z* 226.9–227.2 with elemental composition assignments of the major intensities.

mulative total of 8354 different signals and 1797 assignments of detected non-identical molecular formulae indicate the occurrence of both common and divergent molecules for **P** and **S**.

Hierarchical cluster analysis (HCA) readily identifies two major groups of samples (Figure 2), and shows more uniformity among the **P** samples than among the **S** samples. Clearly, the correct classification of each of the six sets of three repetitions is available from their negative ion mode FTICR mass spectra to assess similarity/dissimilarity. The modes employed and the choice of linkage methods used for clustering greatly affect the numerical outcome of the HCA results. Following careful examination of the available similarity/dissimilarity assessments, Pearson correlation coefficient distance (the straight-line distance between two points in *c*-dimensional space defined by *c* variables) as the similarity descriptor in conjunction with the complete linkage method were found to produce the most distinctive grouping, in which each member within the group is more similar to its fellow members than to any member from outside the group. This is a confirmation that the complete linkage method performs quite well in cases in which objects form naturally distinct "clumps".[39]

HCA does not provide a statistical test of group dissimilarity, but external tests like the Kolmogorov–Smirnov test can be applied for this purpose. This elaboration was done with SAS version 9.1 (SAS Institute Inc., Cary, NC, USA) with the Kolmogorov–Smirnov hypothesis that two groups of observations have identical distributions. By using this test, the difference among **S** subgroups defined by HCA (Figure 2) were determined to be statistically significant for the **DS** and **CS** ($p < 0.0001$) and **TS** and **CS** subgroups ($p < 0.0001$), whereas for subgroups **DS** and **TS**, the asymptotic *p*-value (0.0052) indicates rejection of the null hypothesis that the distributions were also identical for these two subgroups. The partial least square discriminant analysis (PLS-DA) score plot of wood species (Figure 3) provides a representation of how forests from a given species are grouped together. The two predictive components of the PLS-DA model, $R^2(Y) = 99\,\%$ and the prediction accuracy $Q^2(cum) =$

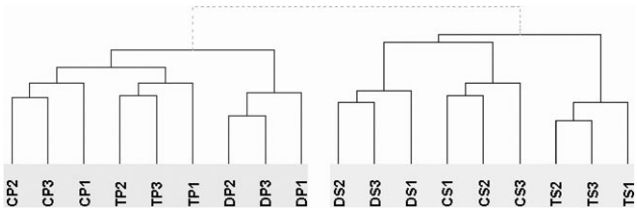

Figure 2. Dendrogram for HCA and classification of 18 (6×3) sets of samples (minimum similarity = 0.44); **T**: Tronçais, **C**: Citeaux, **D**: Darney.

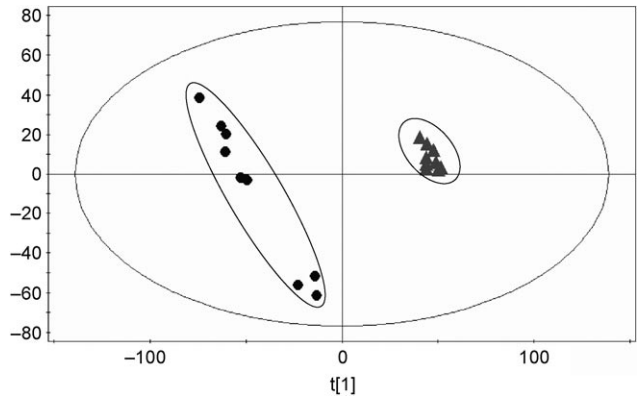

Figure 3. PLS-DA score plot for the first two components indicating the separation between the two species (▲: pedunculate; ●: sessile).

0.96, were obtained though a typical seven-fold cross-validation and guaranteed that this model is satisfactory. In agreement with the cluster analysis (Figure 2), **P** samples exhibit a narrower distribution between the three forests than **S** samples. These findings corroborate the previously observed higher inter-individual variability in whisky lactone contents among **S** oaks in comparison with **P** oaks, in which only traces have been found.[17]

The species effect: The mass spectral signals (m/z values) that drive the differentiation between species were extracted with the highest value of the regression coefficient. For the **S** group, 159 mass signals with a regression coefficient greater than 0.001 were considered to be possible significant biomarkers; in the **P** group, 207 mass signals with a regression coefficient in excess of 0.0004 were selected.

The selected range of m/z 334.95 to 335.30 (Figure 4) illustrates the remarkable resolution of the 12 T FTICR mass spectra. Within this nominal mass, more than a dozen resolved signals were identified in the **S** and **P** samples from the Citeaux forest. The clear and unambiguous differentiation of the species is only feasible at this extent of resolution.

Within the frame of a full-scale metabolomic approach, both the mass-resolving power and the mass accuracy should be precise enough to enable an unambiguous identification of the elemental compositions at the same nominal mass. Even if these two conditions would appear to be fulfilled for most of the mass signals in this study, the lack of experimental databases and the chemical complexity of wood would make the task of identifying all of the corresponding molecules extremely tedious. Yet in particular cases tentative assignment of signals to known wood-related compounds is feasible without the need for other analytical tools. As illustrated by the signal at m/z 335.17114 (averaged value), found only in the mass spectra of **S** oaks (Figure 4a and b), the corresponding $[M-H]^-$ ion $C_{15}H_{27}O_8$ can most likely be assigned from literature data to 3-methyl-4-*O*-β-D-glucopyranosyloctanoic acid, a common precursor of whisky lactone.[40,41] This attribution is confirmed by the higher content of both *trans* and *cis* isomers of whisky lactone in **S** oaks,[42] and in particular in those of the Tonnellerie 2000 experiment.[17] Therefore, whisky lactone and its precursor can logically be considered as biomarkers of **S** oaks.

As already mentioned, errors were made on the original identification of standing trees from the Darney forest; three out of the 20 **P** oaks were actually **S** oaks, and conversely, four out of the 20 **S** oaks were in fact **P** oaks.[34] Because our sets of sawdust samples were prepared in ignorance of these erroneous attributions, we should observe signals specific to the **S** species in the mass spectra of **P** samples from Darney forest, and vice versa. This is illustrated in Figure 4a for the m/z 335.17114 signal attributed to the whisky lactone precursor, which is only detected for the **P** sample from Darney, and not for the other **P** samples.

Note that each of the six samples studied actually represent the average polled sample of 20 wood pieces (as used for one barrel), each piece coming from one distinct tree. In contrast, all of the previously reported studies were based upon detailed analysis of single trees, which obviously favours the detection of singularities but at the expense of excess instrument time. The major consequence of working with "averaged" mass spectra is that singularities at the species or forest level (inter-individual variability) might be attenuated beyond recognition. On the other hand, any differentiation based upon "averaged" spectra will represent more solid evidence of actual tree distinction. To the best of our knowledge, this is the first non-targeted analysis of "averaged" oak metabolites leading to a clear species differentiation between *Q. robur* L. and *Q. petraea* Liebl. and, as shown below, to a forest differentiation.

The 1D ^{1}H NMR analysis confirms this differentiation (Supporting Information, Figure S1). Principle component analysis (PCA) has been found to be a suitable method for the comparison of NMR data from forest-consolidated wood extracts (for both **P** and **S**). In the present study, this analysis led to one statistically significant principle component accounting for 62% of the data variability, which confirmed the chemical characteristics of **P** and **S** wood. This conclusion was supported by the score plot (Figure 5 inset) that shows positive values only for **P** and negative values for all the **S** woods. As indicated by the congruence in the line shapes of the 1D ^{1}H NMR spectra, which indicates molecular environments, the six wood extracts investigated showed considerable similarity at the level of coarse molecular fragments (Supporting Information, Figure S1). However, 1D ^{1}H NMR spectra showed variation in both the NMR integrals of these coarse substructures and in the fine detail of line shapes (Figure 5 bottom). Clearly, the proportion of ^{1}H chemical shifts that weight for the discrimination of **P** extracts is higher (comprised of the $\delta = 0.2$–1.8 and 3.8–5.6 ppm regions and the section downfield from $\delta =$ 6.4 ppm) than the proportion of chemical shifts which discriminate **S** extracts ($\delta = 1.8$–3.8 and 5.6–6.4 ppm regions). The variance in NMR integrals allowed us to quantify the occurrence of substructures, and pattern analysis in multiple 2D NMR spectra aided the structural assignment of classes of molecular environments down to individual molecules, nicely complementing the mass spectral findings. Accordingly, we considered the selection of a single sample sufficient for in-depth NMR characterisation by a suite of two-dimensional NMR experiments, which in combination provide single bond (^1J), geminal (^2J) and vicinal (^3J) connectivity and allow the assignment of molecular fragments across three bonds (Figure 6).

In the aromatic chemical shift region ($\delta(^1H) > 6.2$ ppm), cross-peak positions in COSY and HSQC NMR spectra were shielded in both ^{1}H and ^{13}C frequencies, which indicates a predominance of multiply oxygenated aromatic rings[43] in agreement with the ^{1}H,^{13}C HMBC NMR cross-peaks (Figure 6, lower left-hand dashed box), which show multiple quaternary carbon atoms in the ^{13}C NMR shift range from $\delta(^{13}C) = 106$ to 170 ppm (with maximum cross-peak amplitude between $\delta(^{13}C) = 130$–150 ppm). Further-

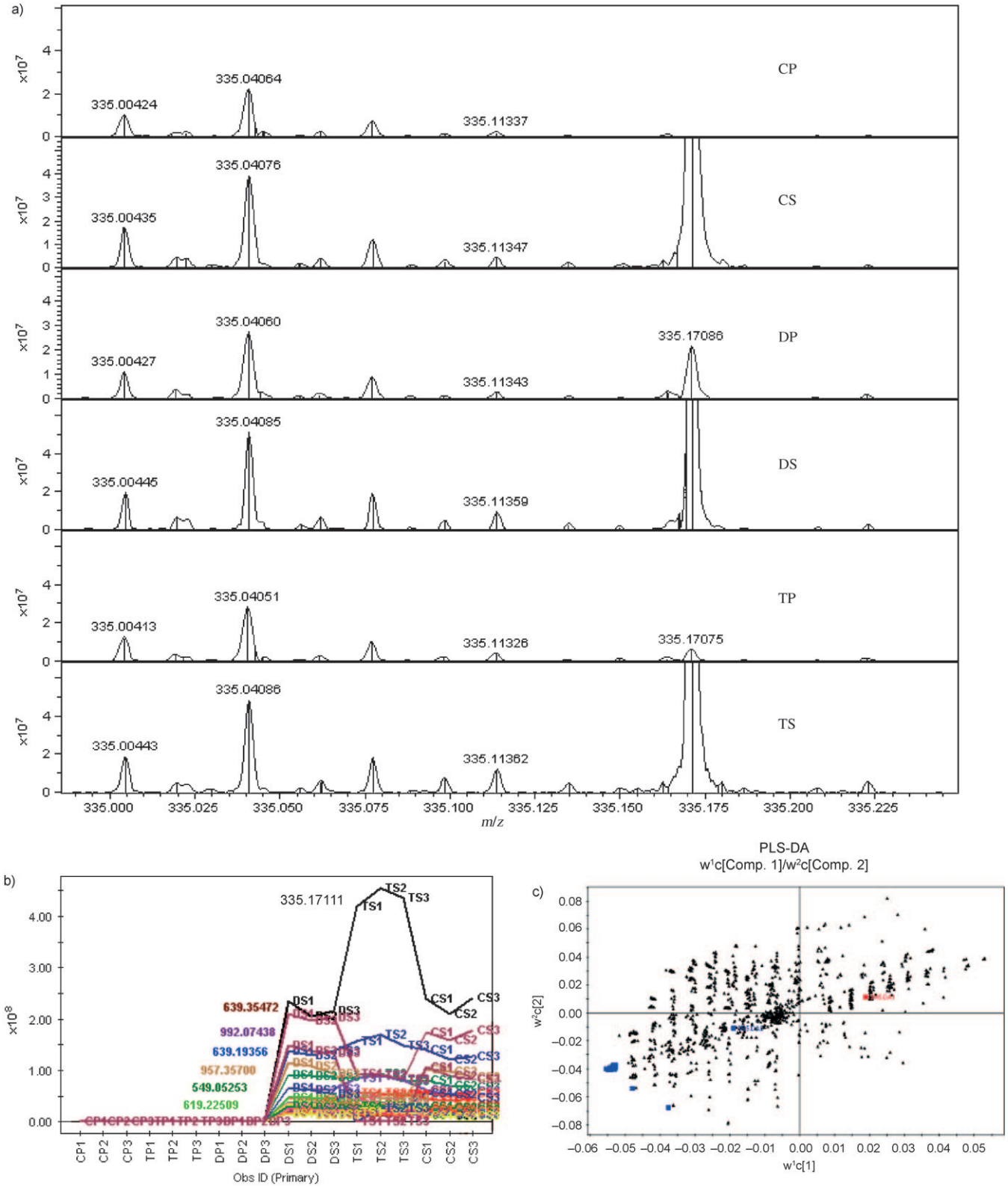

Figure 4. a) Details of the mass spectra for the range *m*/*z* 335.000–335.225 for the two species from the three forests showing the "contamination" of the pedunculate sample from Darney by sessile wood. b) The trend plot delineates the characterization of *m*/*z* 335.17114 for this type of wood as defined in part a); numbers 1, 2 and 3 indicate the three repetitions of each extract. c) Loading plot of masses, with masses peculiar to sessile wood (blue) and masses peculiar to pedunculate wood (red).

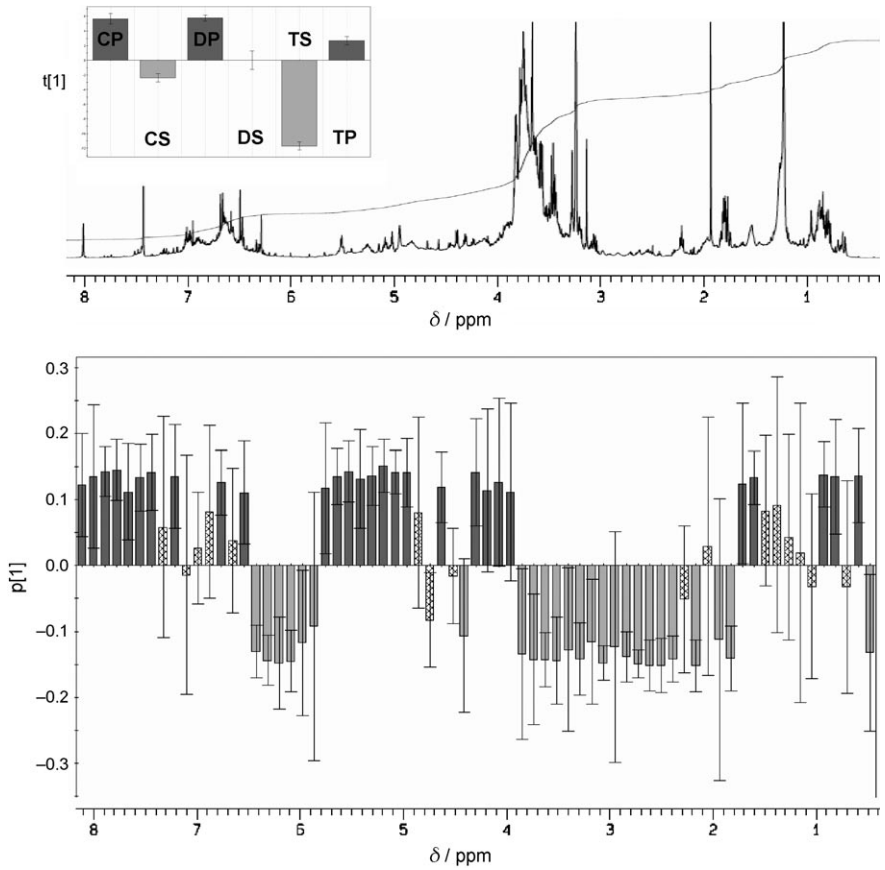

Figure 5. Typical ^{1}H NMR spectrum for the **CS** wood extract and the corresponding PCA of the six lots of methanolic extracts. The score plot (upper left inset) and the loading plot (bottom) of the first principle component from the analysis of the six spectra are also shown. The loading plot highlights discriminant chemical shift weightings for the two species. Non-significant variables are dashed.

more, the minor contribution of aromatic environments to the COSY cross-peak integral (Figure 6 top) as compared with the sizable integrated intensity of aromatic hydrogen obtained from one dimensional ^{1}H NMR spectra (Supporting Information, Figure S1) indicates the occurrence of many isolated aromatic protons. This result convincingly suggests rather extensive degrees of aromatic substitution. All these spectral features are typical of ellagitannins, which are well-established wood constituents,[44,45] and therefore confirm that the ^{1}H NMR spectral region downfield from $\delta = 6.2$ ppm, which discriminates **P** extracts, mostly corresponds to ellagitanins. Similarly, the proton NMR resonances in the $\delta = 5$ to 6 ppm range were not generated from olefinic protons but were attributed to the phenolic ester-type protons because of the ^{1}H,^{1}H COSY, ^{1}H,^{13}C HSQC and ^{1}H,^{13}C HMBC spectral cross-peak positions, which occupied typical shift ranges of phenolic esters rather than those of double bonds. The binding partners as identified from ^{1}H,^{1}H COSY spectral cross-peaks are most likely various carbohydrates, which themselves provide strong signatures in all NMR spectra (e.g., >35 methylene carbon signals (OCH_2) with $\delta(^{13}C) = 60$–66 ppm and methylene-derived HSQC cross-peaks). Again, these features agree with the $\delta = 5$ to 5.6 ppm region of the 1D ^{1}H NMR spectra, which contributes to the discrimination of the **P** species, being correlated to oak tannins, possibly of the galloyl ester type, as identified by LCMS analyses.[46] Conversely, the presence of several cross-peaks between ^{1}H signals in the $\delta = 3$ to 3.8 ppm region and ^{13}C signals in the $\delta = 60$ to 80 ppm region of the ^{1}H,^{13}C HSQC and ^{1}H,^{13}C HMBC spectra (Figure 6 bottom) indicate that carbohydrates probably participate in the discrimination of the **S** species by the $\delta = 2.8$ to 3.8 ppm region of the ^{1}H NMR spectra (Figure 5). Recently, dehydro- and deoxyellagitannins have been identified in toasted oak wood.[47] In general, extended proton spin systems were rather found in the aliphatic section; sizable degrees of branching in purely aliphatic structures ($\delta(^{1}H) < 1.2$ ppm) are also indicated by the positions of ^{1}H,^{13}C HMBC cross-peaks, with carbon chemical shifts of up to 60 ppm (see the upper right-hand dashed box in Figure 6).

The forest effect: The PLS-DA of six times three sets of samples resulted in clear differentiation according to species and to the geographic location of the forests. This is the first time that a molecule-based differentiation according to geographical origin has been demonstrated between oak trees from distinct forests in a given country. Figure 7 shows the 3D score plot of the three times six sets of samples, which indicates good discrimination of the six forests. Most interestingly, the closer correlation between all the **P** groups as compared with all the **S** groups indicates a much higher homogeneity among the former group. From this analysis, it was possible to draft a list of masses characteristic of each of the six forests based on correlation coefficient values. Although selected for this study, **P** oaks are actually scarce in the Tronçais forest (in the centre of France), which is much more renowned for the quality of its sessile oaks for wine ageing.[14] Therefore, these findings not only agree with this fame, which is largely attributed to the higher whisky lactone contents of **S** oaks from Tronçais,[17] but they also provide molecular evidence for this distinction. Indeed, **S** oaks from Tronçais can be discriminated on the basis of more than 194 mass signals (Supporting Information, Table S1), all of them being unambiguously associated with absolute

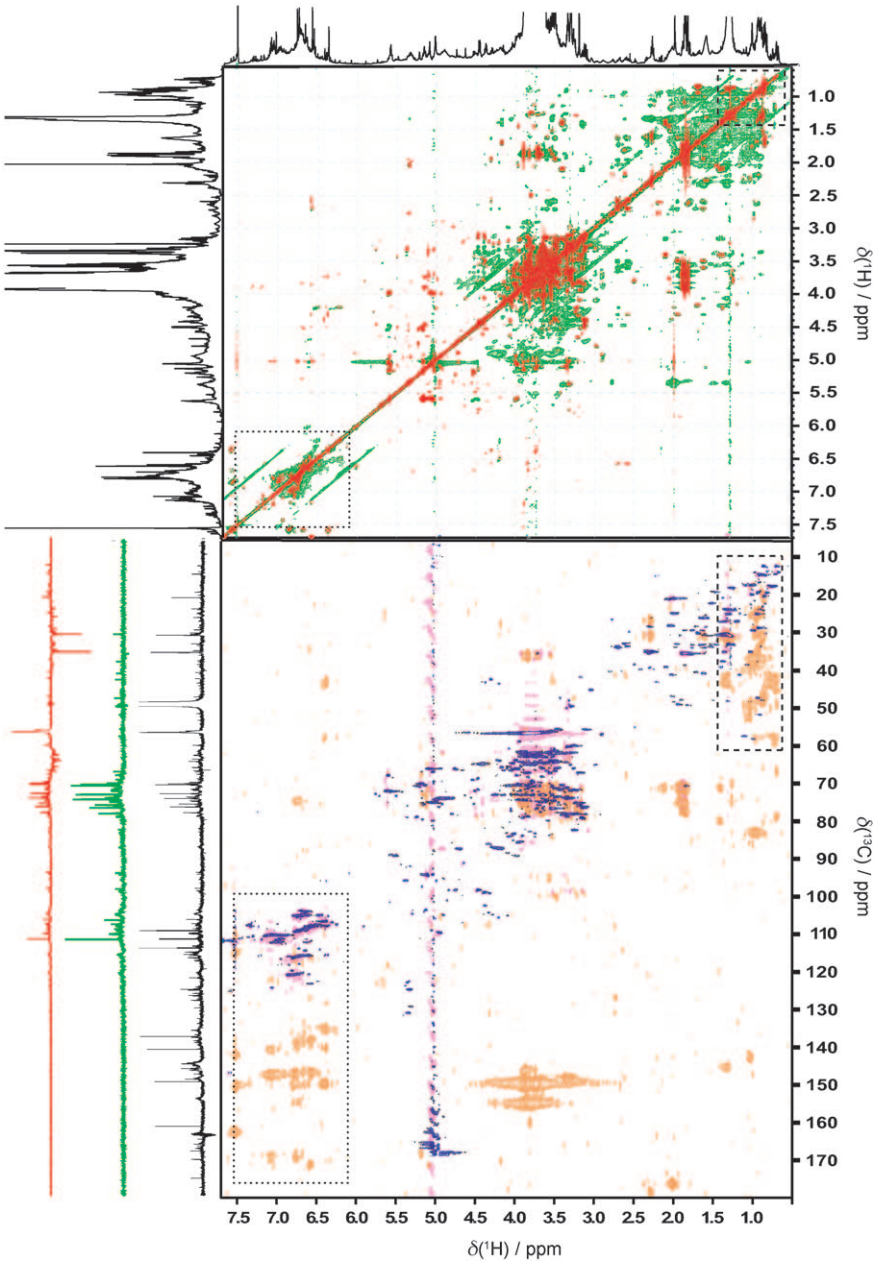

Figure 6. Top: ^{1}H,^{1}H COSY (red) and ^{1}H,^{1}H TOCSY NMR spectra (green) of the **CS** wood extract with ^{1}H NMR projection spectra Bottom: An overlay of ^{1}H,^{13}C HSQC (blue), ^{1}H,^{13}C HMBC (orange) and ^{1}H,^{13}C HSQC-TOCSY (purple) NMR spectra, together with edited ^{13}C NMR projection NMR spectra; red: DEPT-135 (methylene CH_2 down), green: DEPT-90 (methine only) and black: standard ^{13}C NMR spectrum. No appreciable ^{13}C NMR signal intensity was found below $\delta(^{13}C) = 180$ ppm at this S/N ratio. The lower left-hand dashed boxes indicate cross-peaks derived from oxygenated aromatics, whereas the upper right-hand dashed boxes denote cross-peaks from branched aliphatics.

formulae. A particular emphasis should be put on the differentiation of **TS** samples by using sugars as discriminating molecules (Figure 7). Indeed, the PCA and PLS-DA analyses of ^{1}H NMR spectra (Figure 5) already showed that the $\delta = 2.8$ to 3.8 ppm region was weighted for the discrimination of the **S** species, and that this weighting was maximum for the Tronçais forest. Bearing in mind that the 2D NMR spectroscopic experiments supported the assignment of this ^{1}H chemical shift range to carbohydrates, together these results bring insights into a possible sweetness that would particularly characterise oaks from the Tronçais forest.

Wood–wine correlations: Based upon the metabolomic differentiation of wood extracts, the discrimination of selected wines that were aged in barrels made of these particular averaged woods seems very promising. This is illustrated in Figure 8 by the analysis of a Mercurey wine aged in **P** and **S** barrels. The different *m*/*z* distributions in the whole mass range as compared with the wood extract in Figure 1 is clearly visible; the expansion at nominal *m*/*z* 227 shows the likely presence of resveratrol (not present in Figure 1) that traces its origin from the grapevine.

Most interestingly, the expansion at nominal *m*/*z* 335 (Figure 9) demonstrates the higher molecular diversity of the wine compared with the wood extract (Figure 4a) but nevertheless allows to verify the presence of oak wood biomarkers in the wine. Indeed, only wines aged in barrels made of **S** oak woods exhibit the signal attributed to the whisky lactone precursor, which from our above-mentioned results appears to be a potential biomarker of **S** oaks. Analogous relationships apply throughout the entire mass range. Therefore, we exemplify one of the multiple masses likely to differentiate wood origins as a result of a "metabologeography" identification[37] that is further transferable to wine either directly (same biomarker in both wood and wine) or as the result of chemical reactions between wood extracts and wine metabolites.

Conclusion

Many studies of the variability of wood properties have concluded that the largest variations are observed between

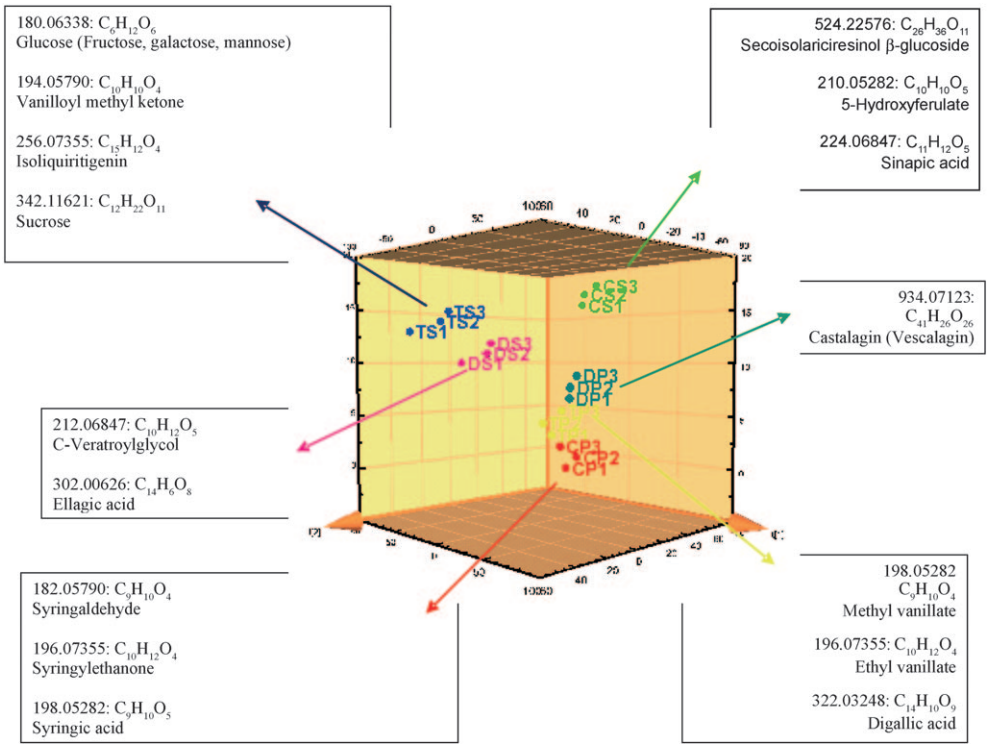

Figure 7. PLS-DA score plot of the 18 sets of samples (Q^2(cum) = 0.80, R^2(Y) = 0.98) grouped by forest, shown with some of the discriminating absolute masses (from negative ESI), the corresponding unique formulas (from neutral mass) and possible associated molecules known to be related to wood.

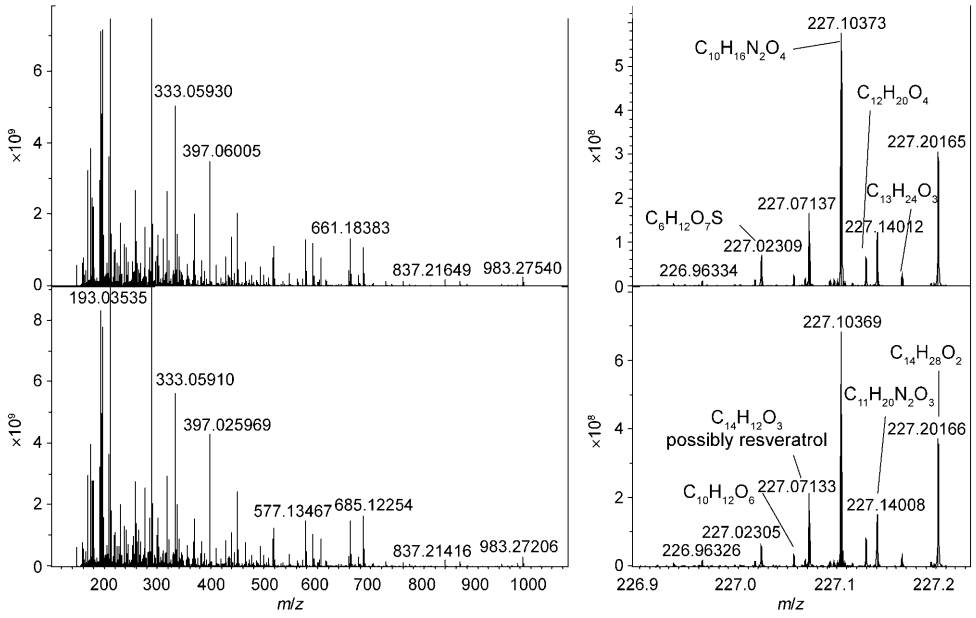

Figure 8. Left: Typical negative ion mass spectrum of Mercurey wine aged for 12 months in pedunculate (top) and sessile (bottom) oak wood barrels from the Tronçais forest. Right: Detail of *m*/*z* 226.90–227.20 (similar to Figure 1 for wood only) with elemental composition assignments for the major intensities (the similar intensity of all signals in *m*/*z* 227 shows no influence due to the wood species for that particular *m*/*z*).

trees within the same forest.[14,48] However, all of these studies have relied upon targeted analyses of initially pre-selected compounds, which in the case of cooperage oaks had been identified as responsible for organoleptic properties. In this study, we have applied an advanced FTICR-MS technique at the highest commercially accessible field strength to assess the possibility of molecularly discriminating a series of oak wood extracts and corresponding wines on a non-targeted basis. The major information provided by such an approach is the relative quantities of all the molecules in any sample that can ionise under the selected experimental conditions. In this context, wood is considered to be a complex biological system that can evolve because of many environmental conditions related to the local ecosystem where it has grown, with the consideration that this multi-parametric variation will express itself through a particular chemical space. Our results demonstrate that ultra-high-resolution FTICR mass spectrometry allows the definition of this chemical space down to a single species in a single

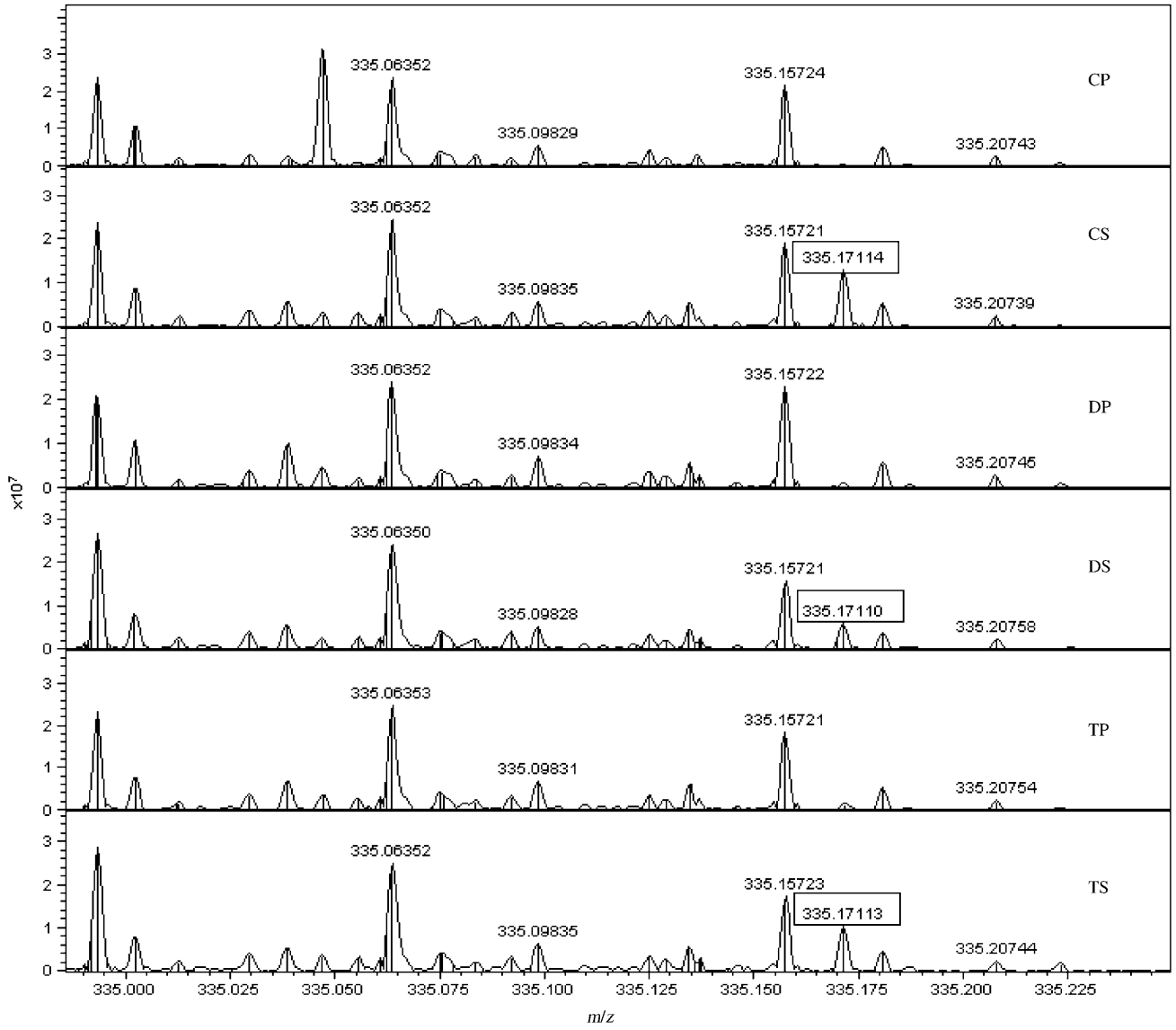

Figure 9. Details of the mass spectra range *m*/*z* 335.000–335.225 range for Mercurey wine aged in barrels made of the two species from the three forests, showing the presence of the *m*/*z* 335.17114 signal only in wine aged in barrels made from sessile oak.

forest. Furthermore, these lists of molecules allow the definition and identification of chemical sub-spaces that could be associated with a forest regardless of the species, or alternatively, selectively associated with a species regardless of the forest.

These results provide insights of considerable novelty in terms of the identification of the chemical composition of oak woods made feasible by ultra-high-resolution FTICR mass spectrometry, which is capable of identifying thousands of distinct molecular compositions directly from mixtures. Even if this full-scale metabolomics approach that includes identification of molecular structures remains at present extremely tedious due to the lack of experimental databases, a promising alternative approach is metabonomics. Here the identification of any single signal (molecular structure) is not necessarily required. Instead, whole sub-spaces are considered and their variations from one sample to another are monitored by the use of advanced processing tools, able to handle very large data sets. We are currently investigating the possibility of applying this metabonomics approach to our set of oak wood samples to assess feasible correlations with the sensory attributes that these woods can transfer to wine.

Finally, we envision general value and applicability in this non-targeted molecular-level traceability, not only for cooperage, but more generally for wine and beverage forensics assessments on European or larger scale levels, or for botanical science and sylviculture to record environmental changes (such as climate modifications over decades) and to improve the nutritional value and sensory properties of agricultural products based upon knowledge of their molecular composition.

Experimental Section

Wood sample collection: The Tonnellerie 2000 experiment[34] was designed to particularly take into account the high inter-individual variability that had already been observed even between trees from the same forest. Therefore, the selected procedure was based on the combination of lots of trees considered to be representative of one species from one forest.

The detailed procedure used to select trees has already been described elsewhere.[34] In brief, twelve lots (five pedunculate and seven sessile) of 24 trees were selected from nine French forests. During the cutting of trees, a disk (about 5 cm thick) was cut one meter up the bole of each tree for further analyses. All wood samples followed a drying procedure (natural seasoning) of one year before the fabrication of barrels. From each disk, a radial strip oriented along the diameter and centred on the outer part of heartwood was kept. For this study, we only considered the three forests in which both the pedunculate and sessile species were represented, that is, Citeaux, Darney and Tronçais. Therefore, we had six lots of 24 strips at our disposal, which had been stored in plastic boxes in the basement of our university building. After a careful examination of the 144 strips, four strips per lot that showed visual traces of mould were excluded. For our study, we therefore had six lots of 20 wood samples (120 samples), which each corresponded to one species from one geographical origin. Note that laboratory morphological analyses realised later, after the original identification of the standing trees, revealed that errors had been made in the assignment of species from the Darney forest; three out of the 20 pedunculate oaks were actually sessile oaks, and conversely, four out of the 20 sessile oaks were actually pedunculate oaks.[34] Our sets of sawdust samples were prepared regardless of these errors, which means that, for instance, 15% of the Darney pedunculate set actually corresponds to the sessile species.

Barrel elaboration: One barrel corresponds to one lot of 24 trees. Each barrel was thus assembled from 24 trees that each contributed $^1/_{24}{}^{th}$ of the toasted surface (body) and $^1/_{24}{}^{th}$ of the untoasted surface (head and bottom). The staves were naturally seasoned for one year, then 48 barrels (12 lots × 4 repeats) were assembled and subsequently medium toasted for 45 min.

Wine elaboration: A first experiment was designed during the 1998 harvest and used a red wine with the appellation "Mercurey rouge 1er cru" made from the Pinot noir grape variety (12 lots × 2 repeats+1 reference stainless steel tank) and a white wine with the appellation "Beaune 1er cru" made from the Chardonnay grape variety (12 lots × 2 repeats+1 reference stainless steel tank). At the end of the wine ageing period (12 months for the red and 14 months for the white), bottling was performed after the two repeats for each lot were blended and thus provided us with 13 bottles of Mercurey and 13 bottles of Beaune.

Wood sample preparation: On each of the 120 wood samples, the outer duramen zone was planed at different locations to obtain coarse shavings a few millimetres thick. Thus, each lot was made of 20 sets of wood shavings equally represented and mixed together. The six lots of wood shavings thus obtained were then ground into powdered samples of less than 250 μm granulometry.

A sample of each sawdust lot (20 mg) was then extracted with ethanol/water (1 mL; 8:2) at RT for 30 min in an ultrasonic bath. Each of the six mixtures was then centrifuged (10 min, 25400 *g*) and further filtered by using 0.2 μm filters. Three repetitions were performed for each of the six lots, which provided us with 18 aqueous alcoholic extracts. Although the aqueous alcoholic solution does not necessarily exhibit the best extracting efficiency for non-volatile compounds, we chose it to minimise the preparation steps prior to injection into the mass spectrometer. For NMR spectral analysis, deuterated solvent was used for extraction and the ethanol extract was analysed after centrifugation.

Wine sample preparation: Wine was sampled directly through the cork of the bottles by using a Hamilton needle. 20 μL of wine was diluted in 1 mL methanol and 50 μL of this solution was used for each experiment and to give the spectral quality presented herein.

FTICR-MS analysis: High-resolution mass spectra for molecular formula assignment were acquired by using a Bruker (Bremen, Germany) APEX Qe Fourier transform ion cyclotron resonance mass spectrometer equipped with a 12 T superconducting magnet and an APOLLO II ESI source set in the negative ionisation mode. Samples were introduced into the microelectrospray source at a flow rate of 120 $\mu L\,h^{-1}$ with a nebuliser gas pressure of 20 psi and a drying gas pressure of 15 psi (200 °C). The spectra were externally calibrated by using clusters of arginine (10 $mg\,L^{-1}$ in methanol), and the accuracy reached values of less than 0.1 ppm in day-to-day measurements. Further internal calibration was performed for each sample by using fatty acids, and the accuracy reached values of less than 0.05 ppm. The spectra were acquired with a time domain of 1 megaword with a mass range of m/z 100 to 2000. The spectra were zero-filled to a processing size of 2 megawords and an average resolution of 250.000 was reached at m/z 200 (100.000 at m/z 600) in full-scan mode. Before Fourier transformation of the time-domain transient, a sine apodization was performed. The ion accumulation time in the ion source was set to 0.2 s for each scan. 1024 scans were acquired for each sample.

NMR spectroscopy: All experiments were performed by using a Bruker DMX 500 spectrometer and a $^{13}C/^{1}H$ dual 5 mm cryogenic probe at 283 K on forest-consolidated wood samples from both species dissolved in 99.95% ^{2}H CD_3OD (184 mg, 225 mL; references for ^{1}H and ^{13}C NMR spectra were $\delta = 3.30$ and 49 ppm; 90°(^{1}H) = 10.1 μs; 90°(^{13}C) = 10.0 μs). 1D ^{1}H NMR spectra were also recorded for each of the six lots of aqueous alcoholic solutions. 1D ^{1}H NMR spectra were recorded by using the first increment of the presat-NOESY sequence (solvent suppression with presaturation and spin-lock, 5.0 s acquisition time, 10.0 s relaxation delay, 320 scans, 1 ms mixing time, 1 Hz exponential line broadening). ^{13}C NMR spectra were acquired by using inverse-gated WALTZ-16 decoupling (13.75 s relaxation delay, 42153 scans for ^{13}C NMR, 75821 scans for DEPT-135 and 32768 for DEPT-90) with an acquisition time of 1.25 s and an exponential line broadening of 1.5 Hz.

$^{1}J(CH)$ used in 1D ^{13}C DEPT and proton-detected 2D NMR spectra was set to 150 Hz. Sensitivity-enhanced, carbon-decoupled $^{1}H,^{13}C$ HSQC NMR spectra were acquired under the following conditions: 90°(^{13}C) decoupling pulse, GARP (70 μs); $F2(^{1}H)$: acquisition time: 291 ms at spectral width of 6009 Hz, $^{1}J(CH) = 150$ Hz, 1.21 s relaxation delay; $F1(^{13}C)$: SW = 22009 Hz (175 ppm); number of scans ($F2$)/$F1$ increments (^{13}C frequency) for $^{1}H,^{13}C$ HSQC experiments: 144/800; for absolute value; $^{1}H,^{13}C$ HMBC (heteronuclear multiple bond correlation): 320/270; $^{1}H,^{1}H$ COSY (correlated spectroscopy): 64/1056; $^{1}H,^{1}H$ TOCSY (total correlated spectroscopy; 70 ms mixing time): 64/938; $^{1}H,^{13}C$ HSQC–TOCSY: 160/513 (70 ms mixing time), respectively. HSQC and DEPT-HSQC spectra were calculated in a 2048 × 512 matrix with exponential line broadening of 2 Hz in $F2$ and a shifted sine bell (π/3) in $F1$. Gradient sequences (1 ms length, 450 μs recovery) were used for all proton-detected spectra.

Analysis of NMR spectra: NMR integrals were measured manually from printed spectra. Bucket analyses[49] was performed on the experimental ^{13}C NMR spectra of six wood extracts; these were separated into 87 equidistant integral segments with 0.1 ppm bandwidth that ranged from $\delta =$ 0.4 to 8.1 ppm.

Statistical analyses: Raw data (from mass spectra) were normalised and then transformed to $\log(X+0.00001)$. The constant 0.00001 was added to provide non-detectable components with a small non-zero value.[50] The transformed variables were then mean-centred and Pareto-scaled and represented as an X matrix. Pareto scaling gives each variable a variance equal to its standard deviation by dividing by the square root of the standard deviation of each column.[51] The sample classification and prior information about the sample were performed by using a HCA unsupervised method. On the other hand, PLS-DA performed by using SIMCA v.11.5,[54] was used to discover characteristic biomarkers.[52] This multivariate procedure provides bioinformatic clues for the selection of a limited number of masses that are most effective in discriminating between different species and forests.

The primary advantage of using targeted profiling as an input in PLS-DA is that the resulting variables are combinations of measured metabolite concentrations. A positive regression coefficient indicates that there is a

greater concentration of the considered metabolites relative to the other samples, whereas a negative value indicates a lower concentration relative to the other samples.[53] As such, these variables are easier to interpret as factors in the underlying classification model. Thus, targeted profiling provides meaningful and interpretable factors describing the input data. PLS-DA is a regression extension of PCA that takes advantage of class information in an attempt to maximize the separation between groups of observations.

The feature selection procedure is comprised of two steps: 1) identification of those masses that best describe each class (a list based on the modelling power of the original variables) and 2) scoring and ranking of the variables in every class-related list according to their ability to discriminate the class they model from all other categories. The ranking and score take place after computation of the minimum number of masses through the formula generator (in-house code written in FORTRAN). The generated formulas were validated by setting sensible chemical constraints (N rule, O/C ratio = 1, H/C ratio = $2n+2$ and element counts: C = 100, O = 80, N = 5, S = 1), and only those masses in conjunction with their automated generated theoretical isotope patterns were taken into consideration.

Acknowledgements

We thank the UNESCO chair "Culture and Wine Tradition" of the University of Burgundy and all of the partners involved in the Tonnellerie 2000 experiment: IUVV, ENSBANA, ONF, ENGREF, INRA, ITV Beaune, DERF, Conseil Régional de Bourgogne, ONIVIN, Professionnels de la Vigne et du Vin de Bourgogne, Syndicat des tonneliers de Bourgogne, and Fédération Française de Tonnellerie.

[1] T. Garde-Cerdan, C. Ancin-Azpilicueta, *Trends Food Sci. Technol.* **2006**, *17*, 438.
[2] A. L. Waterhouse, J. P. Towey, *J. Agric. Food Chem.* **1994**, *42*, 1971.
[3] P. J. Spillman, P. G. Iland, M. A. Sefton, *Aust. J. Grape Wine Res.* **1998**, *4*, 67.
[4] J. R. Mosedale, J.-L. Puech, F. Feuillat, *Am. J. Enol. Vitic.* **1999**, *50*, 503.
[5] L. J. Perez-Prieto, J. M. Lopez-Roca, A. Martinez-Cutillas, F. Pardo-Minguez, E. Gomez-Plaza, *J. Agric. Food Chem.* **2002**, *50*, 3272.
[6] I. Jarauta, J. Cacho, V. Ferreira, *J. Agric. Food Chem.* **2005**, *53*, 4166.
[7] F. Doussot, P. Pardon, J. Dedier, B. De Jeso, *Analusis* **2000**, *28*, 960.
[8] A. Prida, J.-L. Puech, *J. Agric. Food Chem.* **2006**, *54*, 8115.
[9] F. Doussot, B. De Jeso, S. Quideau, P. Pardon, *J. Agric. Food Chem.* **2002**, *50*, 5955.
[10] E. Guchu, M. C. Diaz-Maroto, I. J. Diaz-Maroto, P. Vila-Lameiro, M. S. Perez-Coello, *J. Agric. Food Chem.* **2006**, *54*, 3062.
[11] J. R. Mosedale, P. S. Savill, *Forestry* **1996**, *69*, 47.
[12] P. Chatonnet, D. Dubourdieu, J. N. Boidron, *Am. J. Enol. Vitic.* **1998**, *49*, 79.
[13] J. P. Towey, A. L. Waterhouse, *Am. J. Enol. Vitic.* **1996**, *47*, 163.
[14] J. R. Mosedale, A. Ford, *J. Sci. Food Agric.* **1996**, *70*, 273.
[15] J. Marco, J. Artajona, M. S. Larrechi, F. X. Rius, *Am. J. Enol. Vitic.* **1994**, *45*, 192.
[16] E. Cadahia, S. Varca, L. Munoz, B. Fernandez de Simon, M. C. Garcia-Vallejo, *J. Agric. Food Chem.* **2001**, *49*, 3677.
[17] M. Feuillat, R. Bakour, R. Keller, F. Huber, B. Leaute, J.-L. Puech, *Rev. Oenologues* **2003**, *109*, 19.
[18] J. R. Mosedale, F. Feuillat, R. Baumes, J.-L. Dupouey, R. Keller, J.-L. Puech, *Rev. Fr. Oenol.* **2001**, *187*, 30.
[19] F. Sauvageot, F. Feuillat, *Am. J. Enol. Vitic.* **1999**, *50*, 447.
[20] J.-L. Puech, F. Feuillat, J. R. Mosedale, *Am. J. Enol. Vitic.* **1999**, *50*, 469.
[21] P. Chatonnet, J. N. Boidron, M. Pons, *Connaiss. Vigne Vin* **1989**, *23*, 223.
[22] M. D. Hale, K. McCafferty, E. Larmie, J. Newton, J. S. Swan, *Am. J. Enol. Vitic.* **1999**, *50*, 495.
[23] P. J. Spillman, M. A. Sefton, R. Gawel, *Aust. J. Grape Wine Res.* **2004**, *10*, 216.
[24] D. Chassagne, M. Guilloux-Benatier, H. Alexandre, A. Voilley, *Food Chem.* **2005**, *91*, 39.
[25] C. Saucier, M. Jourdes, Y. Glories, S. Quideau, *J. Agric. Food Chem.* **2006**, *54*, 7349.
[26] G. Ramirez, S. Lubbers, C. Charpentier, M. Feuillat, A. Voilley, D. Chassagne, *J. Agric. Food Chem.* **2001**, *49*, 3893.
[27] V. D. Barrera-Garcia, R. D. Gougeon, A. Voilley, D. Chassagne, *J. Agric. Food Chem.* **2006**, *54*, 3982.
[28] V. D. Barrera-Garcia, R. D. Gougeon, D. Di Majo, D. A. Carmen, A. Voilley, D. Chassagne, *J. Agric. Food Chem.* **2007**, *55*, 7021.
[29] S. Quideau, M. Jourdes, C. Saucier, Y. Glories, P. Pardon, C. Baudry, *Angew. Chem.* **2003**, *115*, 6194; *Angew. Chem. Int. Ed.* **2003**, *42*, 6012.
[30] S. Quideau, M. Jourdes, D. Lefeuvre, D. Montaudon, C. Saucier, Y. Glories, P. Pardon, P. Pourquier, *Chem. Eur. J.* **2005**, *11*, 6503.
[31] J. R. Mosedale, B. Charrier, G. Janin, *Forestry* **1996**, *69*, 111.
[32] C. Scotti-Saintagne, S. Mariette, I. Porth, P. G. Goicoechea, T. Barreneche, C. Bodénès, K. Burg, A. Kremer, *Genetics* **2004**, *168*, 1615.
[33] A. L. Curtu, O. Gailing, L. Leinemann, R. Finkeldey, *Plant Biol.* **2007**, *9*, 116.
[34] F. Feuillat, R. Keller, F. Sauvageot, J.-L. Puech, *Am. J. Enol. Vitic.* **1999**, *50*, 513.
[35] F. Sauvageot, C. Tessier, *Ann. For. Sci.* **2002**, *59*, 171.
[36] N. Hertkorn, M. Meringer, R. Gugisch, C. Ruecker, M. Frommberger, E. M. Perdue, M. Witt, P. Schmitt-Kopplin, *Anal. Bioanal. Chem.* **2007**, *389*, 1311.
[37] J. L. Green, B. J. M. Bohannan, R. J. Whitaker, *Science* **2008**, *320*, 1039.
[38] V. Parra, A. A. Arrieta, J. A. Fernańdez-Escudero, M. Íñiguez, J. A. de Saja, M. L. Rodríguez-Méndez, *Anal. Chim. Acta* **2006**, *563*, 229.
[39] J. E. Taylor, K. D. Hyde, E. B. G. Jones, *J. Biogeogr.* **2000**, *27*, 297.
[40] E. Masson, R. Baumes, C. Le Guerneve, J.-L. Puech, *J. Agric. Food Chem.* **2000**, *48*, 4306.
[41] Y. Hayasaka, K. L. Wilkinson, G. M. Elsey, M. Raunkjaer, M. A. Sefton, *J. Agric. Food Chem.* **2007**, *55*, 9195.
[42] G. Masson, E. Guichard, N. Fournier, J.-L. Puech, *Am. J. Enol. Vitic.* **1995**, *46*, 424.
[43] E. M. Perdue, N. Hertkorn, A. Kettrup, *Anal. Chem.* **2007**, *79*, 1010.
[44] S. Quideau, K. S. Feldman, *Chem. Rev.* **1996**, *96*, 475.
[45] C. Hervé Du Penhoat, V. Michon, S. Peng, C. Viriot, A. Scalbert, D. Gage, *J. Chem. Soc., Perkin Trans. 1* **1991**, 1653.
[46] P. Mämmela, H. Savolainen, L. Lindroos, J. Kangas, T. Vartiainen, *J. Chromatogr. A* **2000**, *891*, 75.
[47] A. Glabasnia, T. Hofmann, *J. Agric. Food Chem.* **2007**, *55*, 4109.
[48] F. Feuillat, R. Keller, *Am. J. Enol. Vitic.* **1997**, *48*, 502.
[49] J. T. Brindle, H. Antti, E. Holmes, G. Tranter, J. K. Nicholson, H. W. L. Bethell, S. Clarke, P. M. Schofield, E. McKilligin, D. E. Mosedale, D. J. Grainger, *Nat. Med.* **2002**, *8*, 1439.
[50] K. Sjödin, L. M. Schroeder, H. H. Eidmann, T. Norin, S. Wold, *Scand. J. Forest Res.* **1989**, *4*, 379.
[51] E. Johansson, N. Kettaneh-Wold, S. Wold, L. Eriksson, *Multi- and Megavariate Data Analysis Principles and Applications*, Umetrics Academy, Umeå, Sweden, **2001**, p. 94.
[52] S. Wold, M. Sjöström, L. Eriksson, *Partial Least Squares (PLS) in Chemistry in Encyclopaedia of Computational Chemistry, Vol. 4* Wiley, New York, **1999**, p. 2006.
[53] R. Rossello′-Mora, M. Lucio, A. Pená, J. Brito-Echeverría, A. López-López, M. Valens-Vadell, M. Frommberger, J. Antón, P. Schmitt-Kopplin, *The ISME Journal* **2008**, *2*, 242.
[54] SIMCA-P v. 11.5, Umetrics, Umea (Sweden).

Received: June 17, 2008
Published online: November 28, 2008

