Anal. Chem. **2009,** *81,* 2884–2894

^{1}H NMR, GC−EI-TOFMS, and Data Set Correlation for Fruit Metabolomics: Application to Spatial Metabolite Analysis in Melon

Benoit Biais,*,†,§ J. William Allwood,*,∥,⊥ Catherine Deborde,†,§ Yun Xu,∥ Mickael Maucourt,‡,§ Bertrand Beauvoit,‡ Warwick B. Dunn,⊥ Daniel Jacob,#,▽ Royston Goodacre,∥,⊥ Dominique Rolin,†,‡ and Annick Moing†,§

INRA, UMR619 Fruit Biology, BP81, F-33140 Villenave d'Ornon, France, Université de Bordeaux, UMR619 Fruit Biology, BP81, F-33140 Villenave d'Ornon, France, Plateforme Métabolome-Fluxome−Génomique Fonctionnelle Bordeaux, IFR 103 BVI, BP81, F-33140 Villenave d'Ornon, France, School of Chemistry, Manchester Interdisciplinary Biocentre, The University of Manchester, 131 Princess Street, Manchester M1 7DN, U.K., Manchester Centre for Integrative Systems Biology, Manchester Interdisciplinary Biocentre, The University of Manchester, 131 Princess Street, Manchester M1 7DN, U.K., Centre de Bioinformatique de Bordeaux−Génomique Fonctionnelle Bordeaux, Université de Bordeaux, 146 rue Léo Saignat, F-33076 Bordeaux, France, and INRA, UMA1251, IFR103 BVI, BP 81, F-33140 Villenave d'Ornon, France

A metabolomics approach combining ^{1}H NMR and gas chromatography−electrospray ionization time-of-flight mass spectrometry (GC−EI-TOFMS) profiling was employed to characterize melon (*Cucumis melo* L.) fruit. In a first step, quantitative ^{1}H NMR of polar extracts and principal component analyses (PCA) of the corresponding data highlighted the major metabolites in fruit flesh, including sugars, organic acids, and amino acids. In a second step, the spatial localization of metabolites was investigated using both analytical techniques. Direct ^{1}H NMR profiling of juice or GC−EI-TOFMS profiling of tissue extracts collected from different locations in the fruit flesh provided information on advantages and drawbacks of each technique for the analysis of a sugar-rich matrix such as fruit. ^{1}H NMR and GC−EI-TOFMS data sets were compared using independently performed PCA and multiblock hierarchical PCA (HPCA), respectively. In addition a correlation-based multiblock HPCA was used for direct comparison of both analytical data sets. These data analyses revealed several gradients of metabolites in fruit flesh which can be related with differences in metabolism and indicated the suitability of multiblock HPCA for correlation of data from two (or potentially more) metabolomics platforms.

The interest for plant metabolomics[1] grows constantly because of its potential applications in plant functional genomics,[2] food science,[3] and human nutrition[4] and has been successfully applied to fleshy fruits.[5−8] Another goal in metabolomics is high-throughput screening of crude samples with little or no sample preparation.[9] For the latter, metabolomics is often used to obtain metabolic fingerprints at specific times of whole tissues or organs and therefore can provide information on plant physiology through temporal and/or spatial localization of metabolites as performed for *Arabidopsis* rosettes,[10] potato tubers,[11] and strawberry flowers.[12] However, plant metabolome complexity limits the ability to collect global metabolomics data using one single technology; thus, several techniques are usually combined to cover primary and secondary metabolites, i.e., gas chromatography/mass spectrometry (GC/MS), liquid chromatography/mass spectrometry (LC/MS), or NMR; several different extraction procedures may also be applied to increase coverage.[1,13,14] Thus far, for the study of primary metabolites, one of two main technologies (^{1}H NMR

* To whom correspondence should be addressed. Fax: +33 (0) 557 122 541 (B.B.); +44 (0)161 306 4556 (J.W.A.). E-mail: benoit.biais@bordeaux.inra.fr (B.B.); william.allwood@manchester.ac.uk (J.W.A.).

† INRA, UMR619 Fruit Biology.

‡ Université de Bordeaux, UMR619 Fruit Biology.

§ Plateforme Métabolome-Fluxome−Génomique Fonctionnelle Bordeaux.

∥ School of Chemistry, The University of Manchester.

⊥ Manchester Centre for Integrative Systems Biology, The University of Manchester.

Centre de Bioinformatique de Bordeaux−Génomique Fonctionnelle Bordeaux, Université de Bordeaux.

▽ INRA, UMA1251.

(1) Hall, R. D. *New Phytol.* **2006**, *169*, 453–468.
(2) Hall, R. D.; Beale, M.; Fiehn, O.; Hardy, N.; Sumner, L. W.; Bino, R. *Plant Cell* **2002**, *14*, 1437–1440.
(3) Wishart, D. *Trends Food Sci. Technol.* **2008**, *19*, 482–493.
(4) Hall, R. D.; Brouwer, I. D.; Fitzgerald, M. A. *Physiol. Plant.* **2008**, *132*, 162–175.
(5) Moco, S.; Capanoglu, E.; Tikunov, Y.; Bino, R.; Boyacioglu, D.; Hall, R. D.; Vervoort, J.; De Vos, R. C. H. *J. Exp. Bot.* **2007**, *58*, 4131–4146.
(6) Mounet, F.; Lemaire-Chamley, M.; Maucourt, M.; Cabasson, C.; Giraudel, J.-L.; Deborde, C.; Lessire, R.; Gallusci, P.; Bertrand, A.; Gaudillère, M.; Rothan, C.; Rolin, D.; Moing, A. *Metabolomics* **2007**, *3*, 273–288.
(7) Roessner-Tunali, U.; Hegemann, B.; Lytovchenko, A.; Carrari, F.; Bruedigam, C. *Plant Physiol.* **2003**, *133*, 84–99.
(8) Pereira, G. E.; Gaudillère, J. P.; van Leeuwen, C.; Hilbert, G.; Maucourt, M.; Deborde, C.; Moing, A.; Rolin, D. *Anal. Chim. Acta* **2006**, *563*, 346–352.
(9) Johnson, H. E.; Lloyd, A. J.; Mur, L. A. J.; Smith, A. R.; Causton, D. R. *Metabolomics* **2007**, *3*, 517–530.
(10) Gibon, Y.; Usadel, B.; Blaesing, O.; Kamlage, B.; Hoehne, M.; Trethewey, R.; Stitt, M. *Genome Biol.* **2006**, *7*, R76.
(11) Shepherd, T.; Dobson, G.; Verall, S. R.; Conner, S.; Griffiths, D. W.; McNicol, J. W.; Davies, H. V.; Stewart, D. *Metabolomics* **2007**, *3*, 475–488.
(12) Hanhineva, K.; Rogachev, I.; Hokko, H.; Mintz-Oron, S.; Venger, I.; Kärenlampi, S.; Aharoni, A. *Phytochemistry* **2008**, *69*, 2463–2481.
(13) Allwood, J. W.; Ellis, D. I.; Goodacre, R. *Physiol. Plant.* **2008**, *132*, 117–135.
(14) Sumner, L. W.; Mendes, P.; Dixon, R. A. *Phytochemistry* **2003**, *62*, 817–836.

10.1021/ac9001996 CCC: $40.75
Published on Web 03/19/2009

and GC/MS) has been employed, on polar extracts.[15] ^{1}H NMR technology has extensively been employed as a high-throughput technique for nontargeted fingerprinting with little or no sample preparation but has also been applied for targeted profiling and the absolute quantification of metabolites when used with an electronic reference.[16] The main disadvantage of ^{1}H NMR is its relatively low sensitivity.[17] GC−EI-TOFMS is much more sensitive than ^{1}H NMR and appeared to be perfectly suited for the detection of volatile metabolites.[13,14] The main drawbacks of GC−EI-TOFMS include that nonvolatile metabolites require derivatization[18] and that, in this case, the quantification is relative to a single internal standard and not absolute. Although the two platforms offer differing advantages for the study of primary metabolism, surprisingly, they have previously rarely been combined within a single study.

For ^{1}H NMR as well as for GC−EI-TOFMS (gas chromatography−electrospray ionization time-of-flight mass spectrometry), data preprocessing and processing are important steps for the generation of relevant biological information from metabolomics.[19,20] ^{1}H NMR classical data preprocessing and processing consists generally in line broadening, phase correction, alignment at δ 0.00 ppm using a reference signal (e.g., sodium trimethylsilyl [2,2,3,3-$^2H_4$] propionate (TSP)), baseline correction, and data reduction (bucketing) for nontargeted fingerprinting or peak area integration for absolute quantification prior to multi- and univariate statistical analyses. For GC−EI-TOFMS, data processing classically employs chromatogram alignment and baseline correction (when necessary), peak picking, deconvolution, relative quantification, and finally array construction, prior to multi- and univariate data analyses.[1,19−21] Although the use of more refined algorithms such as genetic or neural network based algorithms is increasing, principal component analysis (PCA) remains widely used for multivariate analysis of NMR or GC/MS profiling data. PCA models on the natural variance within a data set, and so if the process of interest contributes the majority of the variance within the total explained variance then it is an extremely apt method to identify underlying metabolite variables which contribute to that variance. However, in situations where the "normal" variance is greater than that explained by the process of interest then the more sophisticated genetic or neural network based algorithms may identify the underlying variables of significance which are not associated with the greatest variance (the "normal" condition), although care must be taken not to over train/fit such algorithm-based models.

A metabolomics approach by ^{1}H NMR and GC−EI-TOFMS profiling was developed in order to characterize melon (*Cucumis melo* L.) fruit. The first step was to describe the global metabolite composition of melon fruit flesh extracts. The second step was to assess the spatial localization of the major polar metabolites in the melon flesh through ^{1}H NMR and GC−EI-TOFMS profiling and their combined chemometric analysis. This strategy showed (i) advantages, drawbacks, and complementarity of the two analytical techniques for fleshy fruits metabolomics and (ii) the interest of multiblock hierarchical PCA for the correlation of corresponding ^{1}H NMR and GC/MS data sets. The metabolite profiles and gradients, revealed by the combined ^{1}H NMR and GC−EI-TOFMS approach, offer promising prospects for the future study of fruit physiology and food quality through multiplatform-based metabolite analyses.

METHODS AND MATERIALS

Sample Collection and Handling. Three *C. melo* cultivars were studied, Cézanne, Escrito, and Hugo. Fruits were harvested and processed within 2 h and stored at −80 °C. Fruit cultivation and processing is detailed in the Supplementary Methods 1 section in the Supporting Information.

Chemicals. Methanol-d_4 (99.8%) was purchased from Eurisotop (Gif sur Yvette, France). TSP (98%) was purchased from Aldrich (Saint Quentin Fallavier, France). Succinic-d_4 acid, glycine-d_5, and malonic-d_2 acid standard metabolites (all of 99% purity or greater), all solvents (HPLC grade), *O*-methylhydroxylamine chloride, *N*-acetyl-*N*(trimethylsilyl)-trifluoroacetamide, pyridine, and *n*-alkanes (C10, C12, C15, C19, C22) were purchased from Sigma-Aldrich Ltd. (Gillingham, U.K.). All other chemicals were of reagent grade.

Extraction and ^{1}H NMR Analysis of Polar Metabolites of Ground Flesh Samples. The polar metabolites were extracted from ground melon samples as previously described[16] with slight modifications (Supplementary Methods 2 section in the Supporting Information). The lyophilized titrated extracts were stored in darkness under vacuum at room temperature, before ^{1}H NMR analysis was completed within a week. Quantitative ^{1}H NMR spectra were recorded at 500.162 MHz and 300 K on a Bruker Avance spectrometer (Wissembourg, France) using a 5 mm inverse probe and an electronic reference for quantification as described previously.[6]

Spatial Localization of Metabolites in Melon. *Sampling of Melon Flesh Pieces for Spatial Study.* For each cultivar (cv.), one melon was cut in half lengthways and two slices (one each for ^{1}H NMR and GC/MS) were made (thickness: 1 cm, longitudinal slice). On the remaining part of the fruit, two 1 cm thick half-equatorial slices (one each for ^{1}H NMR and GC/MS) were cut. The slices for ^{1}H NMR were processed immediately, and those for GC/MS were deep frozen in liquid N_2 and stored at −80 °C. The slices were divided into 22 small sections (approximately 7 mm × 7 mm), cut from the skin to the center of the fruit, and numbered as shown in Figure 1A. The sections 1, 5, 10, 13, and 18 were gathered under the name of epicarp (epi), sections 2, 3, 6, 7, 11, 14, 15, 19, and 20 as the outer mesocarp (out meso), and sections 4, 8, 9, 12, 16, 17, 21, and 22 as the inner mesocarp (in meso).

Juice Collection and Global ^{1}H NMR Analyses. Each flesh cube was immediately crushed with a garlic press, aliquots of 150 μL were rapidly deep frozen in liquid nitrogen and stored at −80 °C

(15) Krishnan, P.; Kruger, N. J.; Ratcliffe, R. G. *J. Exp. Bot.* **2005**, *56*, 255–265.
(16) Moing, A.; Maucourt, M.; Renaud, C.; Gaudillère, M.; Brouquisse, R.; Lebouteiller, B.; Gousset-Dupont, A.; Vidal, J.; Granot, D.; Denoyes-Rothan, B.; Lerceteau-Köhler, E.; Rolin, D. *Funct. Plant Biol.* **2004**, *31*, 889–902.
(17) Elipe, S.; Victoria, M. *Anal. Chim. Acta* **2003**, *497*, 1–25.
(18) Roessner, U.; Wagner, C.; Kopka, J.; Trethewey, R. N.; Willmitzer, L. *Plant J.* **2000**, *23*, 131–142.
(19) Trygg, J.; Gullberg, J.; Johansson, A.; Jonsson, P.; Moritz, T. *Plant Metabolomics* **2006**, *57*, 117–128.
(20) Jonsson, P.; Stenlund, H.; Moritz, T.; Trygg, J.; Sjöström, M.; Verheij, E. R.; Lindberg, J.; Schuppe-Koistinen, I.; Antii, H. *Metabolomics* **2007**, *2* (3), 135–143.
(21) Fiehn, O.; Kopka, J.; Dormann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. *Nat. Biotechnol.* **2000**, *18*, 1157–1161.

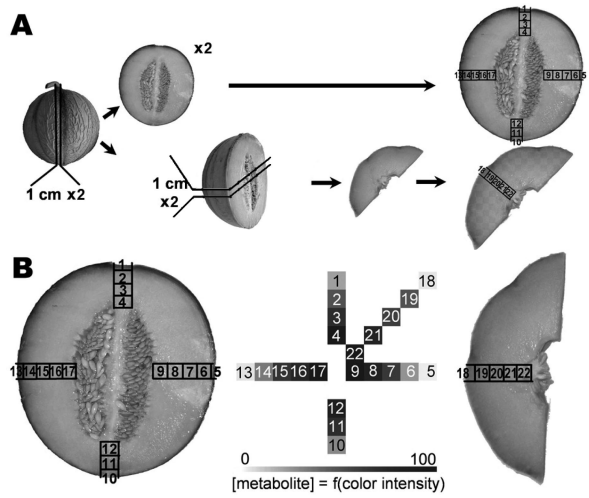

Figure 1. Melon sample collection and data representation. (A) Preparation of melon slices and localization of collected flesh pieces. (B) Schematic representation of concentration gradients depending on the location in the slice.

for later ^{1}H NMR analyses, and the remaining juice was employed for measurements of pH, refractive index, and osmolarity (Supplementary Methods 3 section in the Supporting Information). An appropriate extraction and ^{1}H NMR method for direct analysis of juice required development (Supplementary Methods 4 section in the Supporting Information). The assignments of metabolites in the ^{1}H NMR spectra were made by comparing the proton chemical shifts with literature values,[6,16,22] by comparison with spectra of authentic compounds recorded under identical solvent conditions (own local database), and by spiking the samples. Representative ^{1}H NMR spectra of each cv. and flesh spatial position were converted into JCAMP-DX format and have been deposited, with associated metadata, into the Metabolomics Repository of Bordeaux MeRy-B (http://www.cbib.u-bordeaux2.fr/MERYB/public/PublicREF.php?REF=M08001).

GC−EI-TOFMS Metabolite Extraction and Analysis. After sample grinding in a little liquid nitrogen using a mortar and pestle, 100 mg aliquots (±2 mg) were prepared for later tissue extraction. The extraction procedure for GC−EI-TOFMS analysis essentially followed that of Fiehn et al.[21] Keeping on ice (where possible), samples were extracted in 1 mL of chloroform/methanol/water (1:2.5:1) with shaking at 3 °C for 15 min. The extracts were centrifuged at 14 500*g* for 3 min in a desktop cryofuge at 3 °C to remove debris, and the supernatant was decanted to a clean 15 mL falcon tube. This extraction process was repeated on the same sample material once further. To the 2 mL of pooled extract supernatant, 1 mL of water was added resulting in phase separation which was aided by further centrifugation at 14 500*g* and 3 °C for 3 min. Next, 800 μL of the extract's polar phase (~26.66 mg of material) was removed to a clean 2 mL microcentrifuge tube to which 100 μL of an internal standard solution was added and mixed well. These extracts were finally taken to complete dryness by speed vacuum concentration using an Eppendorf concentrator 5301 set on function 1 at 30 °C for 8 h and stored at −80 °C. The internal standard solution was prepared by dissolving 10 mg of each glycine-d_5, succinic-d_4, and malonic-d_2 acids in 10 mL of water; this stock was further diluted one part to five parts water to provide a working internal standard solution. Prior to derivatization, samples were removed from −80 °C storage and placed in a speed vacuum concentrator for 1 h to remove residual condensation and water. The dried samples were derivatized as follows: 40 μL of 20 mg/mL *O*-methylhydroxylamine solution was added and heated at 40 °C for 90 min followed by addition of 40 μL of *N*-acetyl-*N*-(trimethylsilyl)trifluoroacetamide and heating at 40 °C for 90 min. To each sample derivate, 20 μL of a retention index solution (consisting of 30 mg (±3 mg) of each docosane, nonadecane, decane, dodecane, and pentadecane (98+% purity, Sigma-Aldrich, Ltd. U.K.) dissolved in extra dry hexane (Acros Organics Ltd. U.K.) and diluted 2 parts to 8 parts pyridine to provide a working stock) was added.

All samples were analyzed by GC−EI-TOFMS on an Agilent 6890N gas chromatograph (Stockport, U.K.) coupled to a Leco Pegasus III mass spectrometer (St. Joseph, U.S.A.).[23] This methodology represented a starting point for the analysis of the extremely sugar complex melon matrix; previous GC/MS methodologies had focused purely upon the analysis of volatile compounds in melon fruit and not upon liquid samples requiring chemical derivatization to make them amenable to GC. GC separations were performed using a Supelco DB-50 column (Gillingham, U.K.; 30 m × 250 μm × 0.25 μm) using the following oven conditions: 70 °C (4 min hold time) followed by a 28 °C/min temperature ramp to 290 °C and a final hold time of 1 min. The inlet temperature was 270 °C, an injection volume of 5 μL and a split ratio of 1 in 45 was used with a helium carrier gas at a flow rate of 1 mL/min^{-1}. The transfer line and source temperatures were 250 and 240 °C, respectively. Data was collected in the mass range of 30−600 Da at an acquisition rate of 10 Hz. Data acquisition and raw data processing (chromatographic deconvolution) was performed using ChromaTof v2.15, and all data was exported in ASCII format to Microsoft Excel to produce a data matrix of sample versus metabolite peak with associated peak areas, prior to further data analysis. Peak area data was corrected for derivatization error using the succininc-d_4 acid internal standard. Weigh error was minimal (±2%) and so was not corrected for. Assessments of extraction and analytical reproducibility were made via comparisons of extraction replicates of the same biological material and repeat sample injections (analytical replicates) made within the same run. Chemical identification of metabolite peaks was performed by searching of three mass spectral libraries: NIST/EPA/NIH02 (http://www.nist.gov/srd/nist1.htm), MPI-Golm prepared mass spectral/RI library (http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/gmd.htm1), and an authors prepared mass spectral/RI library. Preliminary identification was confirmed by a mass spectral match >700 and identification was confirmed by analysis of the pure metabolite (purchased from Sigma-Aldrich (Gillingham, U.K.) or ACROS Chemicals (Loughborough, U.K.)) with the same analytical conditions and showing the same RI (±10). Representative GC−EI-TOFMS chromatograms of each cv. and flesh spatial position were converted into Net CDF format and have been

(22) Fan, T. W.-M. *Prog. Nucl. Magn. Reson. Spectros.* **1996**, *28*, 161–219.

(23) O'Hagan, S.; Dunn, W. B.; Brown, M.; Knowles, J. D.; Kell, D. B. *Anal. Chem.* **2005**, *77*, 290–303.

deposited, with associated metadata, into the Metabolomics Repository of Bordeaux MeRy-B (http://www.cbib.u-bordeaux2.fr/MERYB/public/PublicREF.php?REF=M08001).

Data Statistical Analysis and Visualization. All the pH, osmolarity, and ^{1}H NMR data were submitted to *N*-way ANOVA (analysis of variance) and Tukey's test using SAS Software v8.01[24] to reveal significant differences between means. For ^{1}H NMR, PCA was used to highlight differences between cvs. and gradients (and/or specific localization) of metabolites depending on the tissue and position in the fruit. PCA was performed with the R software v2.5.1 (http://www.r-project.org/) and the FactoMiner package with mean-centered data scaled to unit variance.

The GC−EI-TOFMS data was first submitted to classical PCA (on mean-centered data scaled to unit variance) using the MATLAB R2006 software package (The MathWorks Inc. Natick, U.S.A.) as described previously.[25] Since GC−EI-TOFMS data is very information-rich and due to the variable changes within tissue types for some metabolite species, classical PCA may not be able to highlight and describe the interesting trends that one wishes to observe, at least not in the first few principal components (PCs). Standard PCA routines are much more applicable to data sets consisting of a small number of sample groups each containing large numbers of robust and reproducible replicate samples. Such sample sets generally employ tissue pooling to make up each biological replicate, whereas in this study each sample is collected from an independent fruit position, and so the sample set can be grouped according to relative fruit position (epicarp, outer and inner mesocarp), but the heterogeneity within these groups of samples is huge when compared to that seen between replicates within groups of pooled tissue samples.

The lack of success with the conventional PCA approach does not necessarily mean that the GC−EI-TOFMS platform cannot be used to detect the spatial distribution of metabolites. As discussed above, the reason could be the high complexity of the data set, and the trend we are seeking may be overwhelmed by other irrelevant information. Similar situations also occur in the field of statistical process control (SPC) where researchers want to detect the occurrence of abnormality as early as possible. In the SPC scenario, abnormal signals are frequently overwhelmed by "normal" signals. An effective approach to improve the sensitivity of detecting abnormality is to employ several chemical and/or physical sensors to monitor the process and to combine their signals by using multiblock PCA models. Multiblock PCA models[26,27] are extensions of standard PCA which aim to combine several different but potentially connected data sets (called "blocks"), with emphasis upon modeling the "common trend" between blocks. The common trend of the different blocks are revealed in the "super scores" plot, the distribution of the samples of each individual block are shown in their respective "block scores", and like classical PCA, the contribution of variables to the trend shown in the blocks scores plot is shown in their "block loadings" plot. In SPC applications, if all the sensors employed (considering each sensor as an individual block within the multiblock PCA model) can detect abnormal signals, even when they are extremely weak, such signals will be much more easily highlighted with a multiblock PCA model than several independent standard PCA models (based upon one model per each sensor). This would especially be the case if abnormality happened to be detected by all of the sensors, i.e., a common trend across all the blocks. Multiblock PCA specifically looks for the common trend (abnormality in the case of SPC), if the trend is present between the different blocks then it will modeled upon; however, in the case of a standard PCA routine only the pattern of "normal" signals would be illustrated and not the underlying trend toward abnormality. In this study, the same approach can be employed, especially if the gradient distribution of metabolites is not the main trend within the GC−EI-TOFMS data set and standard PCA failed to reveal it. If such a trend should exist within all three melon cultivars, the three cultivars can be considered as three different blocks, and thus a multiblock PCA model can be constructed that would stand a better chance of revealing the trend within its first few PCs than a standard PCA routine.

In this study, a multiblock PCA model called hierarchical PCA (HPCA) was used. In addition to HPCA, *N*-way ANOVA was used to identify potentially significant metabolites differing in their spatial distribution. Since there are a large number of multiple comparisons, the threshold of the *p*-value was down adjusted according to Benjamini and Hochberg procedure[28] by setting the false discovery rate (FDR) to 0.05. We also attempted to combine the GC−EI-TOFMS and ^{1}H NMR data together by building a six block HPCA model to find which metabolites showed similar trends across the different analytical data sets. All HPCA and *N*-way ANOVA were performed using the Multiblock toolbox for Matlab (available from http://www.models.kvl.dk/source/MBToolbox/) and the statistics tool box, respectively, for the MATLAB R2006 software package (The MathWorks Inc.). Prior to HPCA, both GC/MS and ^{1}H NMR data sets were scaled by mean centering to unit variance.

Multi Experiment Viewer (MeV) v4.0 software[29] was used to obtain "heat-map-like" graphics (for both ^{1}H NMR and GC−EI-TOFMS) where the color intensity depends on the metabolite concentration. Data were presented in a cross where each branch represents one part of the longitudinal slice. Data for the equatorial slice were added in a diagonal branch (see the example graphic in Figure 1B).

RESULTS AND DISCUSSION

Establishing a Baseline Metabolome for Melon Flesh. To establish a baseline metabolic profile of melon fruit flesh, ethanolic extraction and ^{1}H NMR analysis was performed. Twenty-seven polar metabolites were identified according to the chemical shifts given in Supporting Information Table S1, including three carbohydrates, two oligosaccharides, two polyols, three organic acids, three amino acids, and two quaternary amines. One representative ^{1}H NMR spectrum is presented in Supporting Information Figure S1A; each identified resonance was integrated,

(24) SAS Institute. *SAS/STAT User's Guide*, version 6, 4th ed.; SAS Institute Inc.: Cary, NC, 1990.
(25) Allwood, J. W.; Ellis, D. I.; Heald, J. K.; Goodacre, R.; Mur, L. A. J. *Plant J.* **2006**, *46*, 351–368.
(26) Smilde, A. K.; Westerhuis, J. A.; Jong, S. *J. Chemom.* **2003**, *17*, 323–337.
(27) Westerhuis, J. A.; Kourti, T.; Macgregor, J. F. *J. Chemom.* **1998**, *12*, 301–321.
(28) Benjamini, Y.; Hochberg, Y. *J. R. Stat. Soc., Ser. B* **1995**, *57*, 289–300.
(29) Saeed, A.; Sharov, V.; White, J.; Li, J.; Liang, W.; Bhagabati, N.; Braisted, J.; Klapa, M.; Currier, T.; Thiagarajan, M.; Sturn, A.; Snuffin, M.; Rezantsev, A.; Popov, D.; Ryltsov, A.; Kostukovich, E.; Borisovsky, I.; Liu, Z.; Vinsavich, A.; Trush, V.; Quackenbush, J. *BioTechniques* **2003**, *34*, 374–378.

and the corresponding metabolite was quantified (Supporting Information Table S2). The absolute metabolite concentrations were submitted to PCA which revealed cv. differences (Supporting Information Figure S1B). The concentrations of individual metabolites, for the three cvs., are presented in Supporting Information Table S2. These results are in general agreement with other data on *C. melo* L.[30−34] They have highlighted that (i) the extracts have particularly high sugar concentrations and (ii) there is at least a 3 order range in concentration for all the polar metabolites detected in the ^{1}H NMR spectra.

The use of this hot ethanolic extraction has brought proofs of robustness and reproducibility for ^{1}H NMR analysis.[16] The extraction concentrated several minor metabolites permitting their detection by ^{1}H NMR. Sample extracts are prepared in a buffered solution, thus avoiding the shift of metabolite peaks due to pH or ionic strength variations between samples. A similar extraction was attempted to produce samples suitable for GC−EI-TOFMS, only fresh material was used and not freeze-dried. In a preliminary experiment, it was noted that freeze-dried melon extracts contained such high sugar levels that chemical derivatization was inhibited or chromatography was heavily overloaded with monosaccharides. Similarly, the amended ethanolic method[16] resulted in very high concentrations of sugars leading to the same difficulties. In future studies this problem may be averted by use of a two-stage analysis, where first a concentrated sample is analyzed for sugar content, a second sample will be prepared by subjecting the same extract to a solid-phase extraction (SPE) for the removal of sugars[35] and analyzed for the lower concentration sample components.

Spatial Localization of Metabolites in Melon. *^{1}H NMR and GC−EI-TOFMS Metabolic Profiling of Melon Juice and Flesh.* Classical analyses to quickly assess fruit quality include pH, osmolarity, and refractive index measurements. These parameters were measured for each flesh piece (Figure 1A) from their squeezed juice (Supporting Information Table S3) revealing some significant gradients between the epicarp and the inner mesocarp. The osmolarity and refractive index showed parallel changes with a clear gradient from the epicarp to the inner mesocarp. In order to gain metabolic insights into the components of pH, osmolarity, and total soluble solids, metabolic profiles of melon juice from each flesh section (Figure 1A) were prepared in methanol-d_4 using ^{1}H NMR spectroscopy. Representative parts of the spectrum of each cultivar are shown in Figure 2A; 15 metabolites were identified (according to chemical shift, Supporting Information Table S1) and 14 were quantified (Supporting Information Table S4). The major metabolites in juice were sugars (sucrose, glucose) and amino acids (alanine, valine, threonine, aspartic acid, and GABA). As the samples were not buffered, it was impossible to distinguish accurately the glutamine and glutamic acid resonances; thus they were described as "glx" metabolite. Citric acid was the major organic acid detected, although malic and acetic acids were also detected. Ethanol was detected in several samples. Finally, two metabolites which could not be identified were named as unknowns (unkD3.1 and unkD1.3).

As the samples contained approximately 30% water (v/v), a water presaturation sequence was used to remove the water peak at 4.8 ppm. Unfortunately, the β-glucose doublet was also affected by the presaturation since its area was decreased between a proton 1D pulse sequence and a Noesy−Presat pulse sequence (NOESYGPPR1D: a presaturation experiment for water suppression incorporating the first increment of the NOESY pulse sequence and a spoil gradient) (data not shown). Thus, the glucose triplet pattern at 3.20 ppm (Figure 2A), which was not affected, was used for the absolute quantification of β-glucose. Both sucrose (5.41 ppm) and α-glucose (5.18 ppm) doublets were generally not affected by the water presaturation, although in some cases the α-glucose doublet appeared to be decreased. The concentration of α-glucose was determined using the calculated β-glucose concentration and the natural abundance ratio of α-glucose and β-glucose (36% α, 64% β, assessed with nonpresaturated acquisitions, data not shown).

The juice-based direct profiling method in methanol-d_4 was advantageous due to the speed and ease of sample processing. The ^{1}H NMR spectra were acquired from preserved samples, since the addition of 70% methanol-d_4 directly into the juice blocks enzymatic activities. Moreover, it allowed for the detection and quantification of endogenous ethanol, an interesting indicator of metabolic changes.[36] However, drawbacks were the lower sensitivity of the technique and a longer shimming time before spectrum acquisition due to the methanol-d_4/water mixture (70/30, v/v). Both the baseline and spatial ^{1}H NMR method gave valuable information on the metabolite composition of melon fruit flesh, but the two data sets may not be compared directly: some differences may be attributed to individual fruit variability and/or the extraction process.

Metabolic profiles of polar extracts from melon flesh sections (Figure 1A) were generated by GC−EI-TOFMS postderivatization with *O*-methylhydroxylamine and *N*-acetyl-*N*-(trimethylsilyl)trifluoroacetamide. Deconvolution of the chromatographic profiles produced an output consisting of 105 metabolite features, 58 of which had been assigned metabolite identifications (by matching of retention index and mass spectra to authentic chemical standards) and a further 12 of which had been assigned as unknown sugars based upon mass spectra. The table of unambiguously identified metabolite features is available in the Supporting Information (Table S5). The major metabolites in the melon flesh extracts were sugars (sucrose, glucose, fructose, trehalose) and amino acids (alanine, β-alanine, valine, threonine, aspartic acid, glycine, lysine, leucine, GABA, serine, and tyrosine, several of which revealed tissue specific gradients not detected by ^{1}H NMR). Similar to ^{1}H NMR, citric acid was the most concentrated organic acid detected, followed by malic acid; however, the improved sensitivity of GC−EI-TOFMS led to the detection of a series of lower concentration organic and fatty acids (including fumaric, succinic, glyceric, stearic, hexadecanoic, glucuronic, galacturonic, myristic, shikimic, and glutam-

(30) Stepansky, A.; Kovalski, I.; Schaffer, A. A.; Perl-Treves, R. *Genet. Resour. Crop Evol.* **1999**, *46*, 53–62.
(31) Gao, Z.; Petreikov, M.; Zamski, E.; Schaffer, A. A. *Physiol. Plant.* **1999**, *106*, 1–8.
(32) Wang, Y.; Wyllie, S. G.; Leach, D. N. *J. Agric. Food Chem.* **1996**, *44*, 210–216.
(33) Chachin, K.; Iwata, T. *Bull. Univ. Osaka Prefect., Ser. B* **1988**, *40*, 27–35.
(34) Hashinaga, F.; Koga, T.; Ishida, K. *Bull. Fac. Agric., Kagoshima Univ.* **1984**, *34*, 29–37.
(35) Suzuki, H.; Achnine, L.; Xu, R.; Matsuda, S. P. T.; Dixon, R. A. *Plant J.* **2002**, *32*, 1033–1048.
(36) Tadege, M.; Dupuis, I.; Kuhlemeier, C. *Trends. Plant Sci.* **1999**, *4*, 320–325.

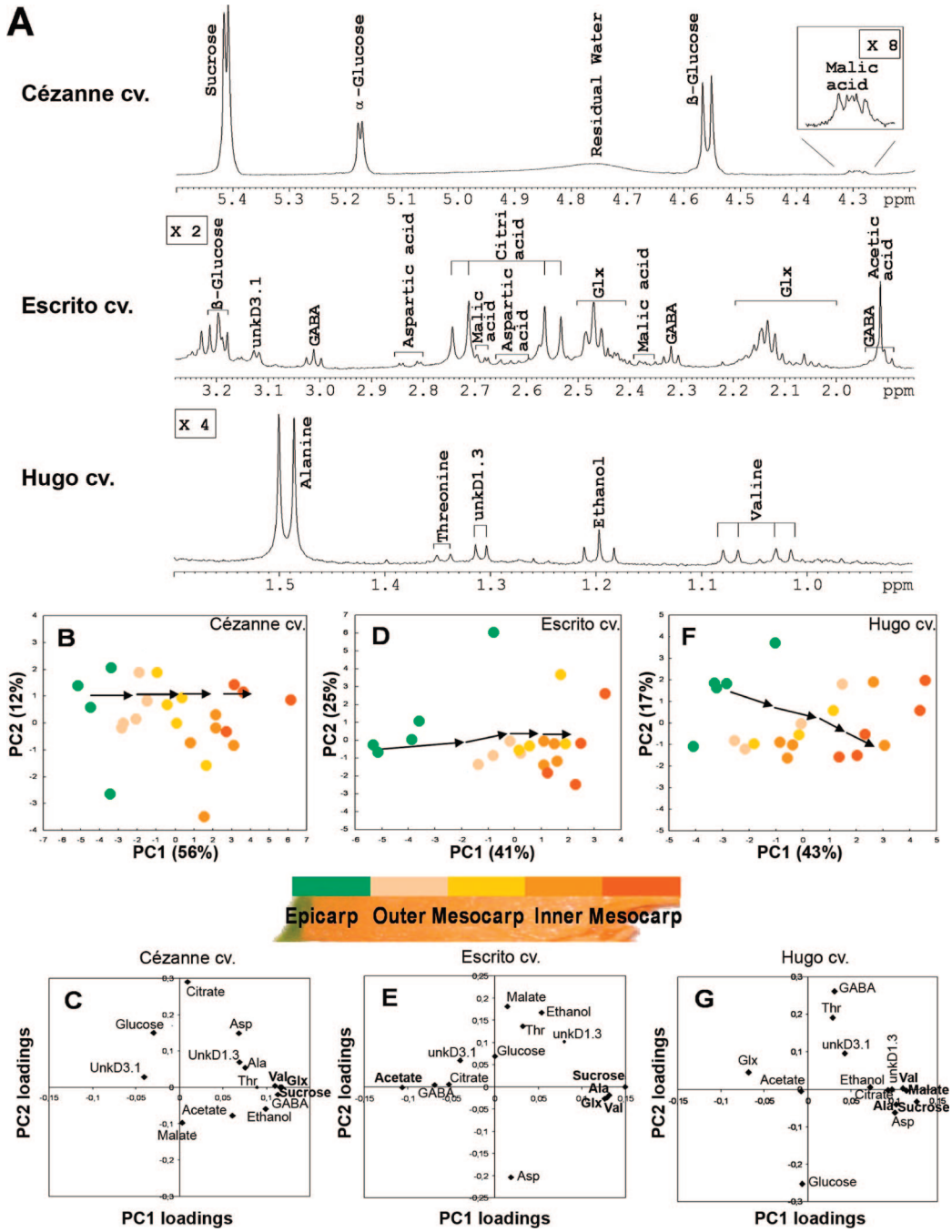

Figure 2. ^{1}H NMR spatial metabolite profiling of melon slices. (A) Representative parts of ^{1}H NMR spectra of each cultivar of melon juice in 70% methanol-d_4. Resonances are annotated according to the chemical shifts in Supporting Information Table S1. Glx corresponds to glutamine plus glutamic acid resonances. (B−G) Principal component analysis of 14 metabolites (absolute concentration) obtained from the ^{1}H NMR spectra of 22 melon sections for each of the three cultivars (Cézanne, Escrito, and Hugo). The metabolites were identified and quantified using ^{1}H NMR spectra of melon juices. Mean-centered data scaled to unit variance were used for PCA. (B, D, F) Scores plots. (C, E, G) Loading plots. Metabolites with PC1 loading values greater than 0.1 are indicated with a bold font. Acids are noted with their conjugate base name for readability purpose.

ic acids) which showed significant difference between the tissue grades and/or cultivars. Ethanol was not detected by GC−EI-TOFMS since the compound would have been lost during sample lyophilization; however, GC−EI-TOFMS did detect ethanolamine. Other unambiguously identified metabolites detected by GC−EI-TOFMS included, phenylalanine, uracil, threitol, and glutathione (Supporting Information Table S5).

^{1}H NMR Spatial Differences Highlighted Using Principal Component Analysis. In order to visualize metabolite gradients and identify discriminant metabolites, the ^{1}H NMR data were first submitted to PCA. The scores plots for Cézanne, Escrito, and Hugo cvs. are presented in Figure 2, parts B, D, and F, respectively. In each case, the first component (PC1) explained more than 40% TEV (total explained variance) for Cézanne, Escrito, and Hugo. The three scores plots showed very similar

patterns, with PC1 discriminating the samples depending upon spatial location along the melon slice from epicarp to inner mesocarp. The epicarp samples were located on the negative side of PC1, whereas the inner mesocarp samples were on the positive side. The corresponding loadings plots are presented in Figure 2, parts C, E, and G. Their observation showed the main discriminant metabolites (loading factors higher than 0.1) on PC1. Some of these discriminant metabolites were common to the three melon cvs., like sucrose, valine, and alanine, and to a lesser extent ethanol and "glx" (glutamine and glutamic acid resonances). For sugars, the gradients observed in the present experiment were in agreement with those previously showed on melon using near-infrared imaging.[37] Other metabolites appeared to be specifically discriminating for one cv., e.g., GABA for Cézanne, acetic acid for Escrito, and malic or aspartic acid for Hugo.

GC−EI-TOFMS Spatial Differences Highlighted Using Principal Component Analysis and Multiblock Hierarchical Principal Component Analysis. In order to visualize metabolite gradients and identify discriminant metabolites from the GC−EI-TOFMS data set, the data were first submitted to a standard PCA. Unfortunately, PCA did not successfully cluster the sample groups according to tissue grade (Figure 3A); this probably resulted from the high complexity of GC−EI-TOFMS data compared to ^{1}H NMR. GC−EI-TOFMS detected over 100 metabolite features compared to the 14 metabolites quantified within the ^{1}H NMR data set. The alternative sample preparations, tissue extraction, and derivatization versus juice, different locations of the tissue slices, independent sectioning of the tissue slices, and different extractions may have also contributed to this increased data complexity/variability. For future experimentation, all sample processing will be undertaken by a single technician, thus increasing the metabolic reproducibility by producing a series of identical samples for analysis upon multiple platforms.

It is also likely that the small number of bulk metabolites detected by the ^{1}H NMR and their spatial distributions were less variable than the shifts seen within many of the low-abundance metabolite species observed by GC−EI-TOFMS which were outside of the detection limits of ^{1}H NMR. It is also likely that the bulk metabolite detection is more variable and less reliable within the relatively quantified GC−EI-TOFMS data set, due to saccharide overloading and saturation of the detector, when compared to the fully quantified ^{1}H NMR data set. These issues are currently being addressed by researchers at the University of Manchester and Max Plank Institute of Molecular Plant Physiology (Golm, Germany), who are making substantial developments to the GC−EI-TOFMS methodology and data-processing strategies. Although the detection and quantification of bulk metabolite species was more reliable upon the ^{1}H NMR platform, ^{1}H NMR did not detect some of the significant metabolite gradients observed within the GC−EI-TOFMS data set including ethanolamine, β-alanine, pyruvic acid, lysine, serine, and stearic acid. It must also be stated that the GC−EI-TOFMS detected GABA amino acid gradient common to all three cultivars, whereas the ^{1}H NMR showed more variable data. However, without further biological replication and analysis of identical sample material, it is difficult to say which instrument is faithfully recording the metabolite levels, although it is likely that the ^{1}H NMR is more faithful with regard to bulk metabolites and GC−EI-TOFMS with regard to low-abundance metabolites. Care must be taken though when directly assessing relatively quantified data against absolutely quantified data since the results will differ.

Due to the lack of success with the conventional PCA approach, a multiblock HPCA model was built. The HPCA revealed the differences between epicarp and mesocarp and highlighted the discriminant metabolites for the three individual melon cvs. The super scores and block scores plots for Cézanne, Escrito, and Hugo are presented in Figure 3, parts B and C, respectively. On the super scores plot, PC1 explained 21% TEV, while PC2 explained 16% TEV. In comparison to the results of the standard PCA, PC1 in the multiblock HPCA model clearly discriminated the samples depending on the location of the section on the slice. The TEV values explained why classical PCA was not successful for the GC−EI-TOFMS data set when compared to the ^{1}H NMR data set. In the ^{1}H NMR data set, the main trend (more than 40% TEV) is due to the spatial difference of metabolites, whereas in the GC−EI-TOFMS data set, such trends are much weaker (21% TEV) and thus multiblock HPCA was required to reveal them. PC1 clustered the samples into groups (epicarp, inner and outer mesocarp) according to the spatial location. The epicarp samples were located on the positive side of PC1 while the inner mesocarp samples were on the negative side. The corresponding loadings plot is presented in Figure 3D. Observation of the loadings plots showed the main discriminant metabolites were present on PC1. Some of these discriminant metabolites were common to the three cvs., such as ethanolamine, sucrose, valine, serine, alanine, and β-alanine, and others appeared to be specifically discriminating according to cultivar, e.g., trehalose, GABA, threonine, phenylalanine, and glycerol. A univariate data analysis was undertaken using *N*-way ANOVA with a false discovery rate set at 5% (i.e., a 95% confidence limit) on the GC−EI-TOFMS data set; 47 variables were identified as significant between different melon tissue locations, whereas 82 variables were identified as significantly different between melon cvs. (Supporting Information Table S5).

GC−EI-TOFMS and ^{1}H NMR Correlation Analysis with Multiblock Hierarchical Principal Component Analysis. By putting both analytical data sets into a single HPCA model, all of the blocks scores were found to show the same trend, thus making them comparable to each other (Figure 4). More interestingly, by examining the loadings plot, it was discovered that the same compounds detected by both analytical techniques show similar positions within the plots. Four identified metabolites (common to both data sets), alanine, sucrose, valine, and GABA, were highlighted as showing similar trends. It can be clearly seen that they appear in a similar position in the loadings plots within the same cvs. of melon but analyzed by different techniques. This demonstrates the great potential of using multiblock PCA to combine different data sets together and correlate the knowledge discovered by the different metabolomic platforms.

Spatial Gradients for Individual Metabolite. The concentrations of individual discriminating metabolites were visualized with heat maps (as explained in Figure 1B) to highlight metabolite gradients. Figure 5 presents the graphics obtained from ^{1}H NMR for sucrose (Figure 5A), alanine (Figure 5B), valine (Figure 5D),

(37) Sugiyama, J. *J. Agric. Food Chem.* **1999**, *47*, 2715–2718.

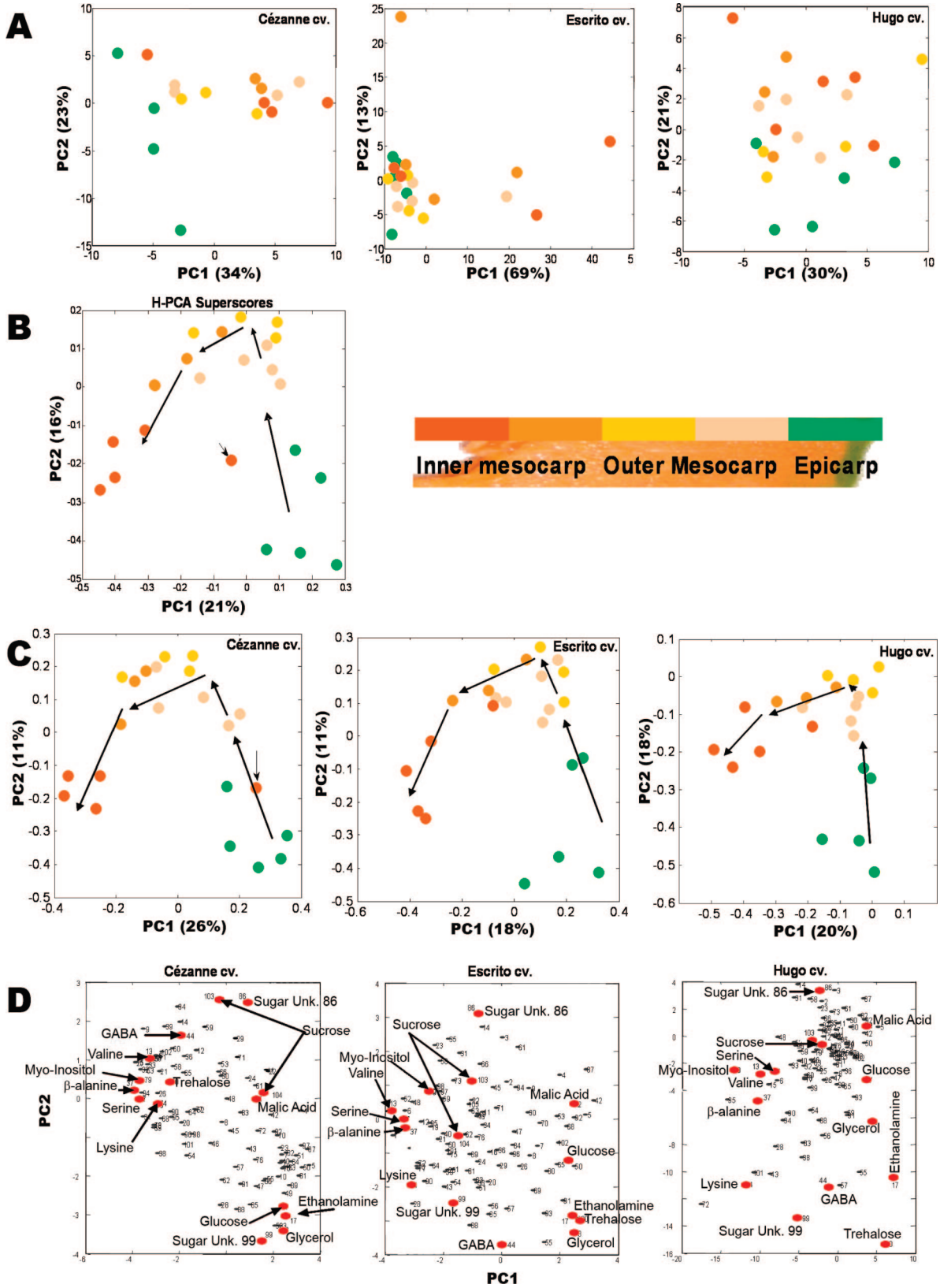

Figure 3. Principal component analysis of all detected metabolite features obtained from the GC−EI-TOFMS analysis of 22 melon sections for each of the three cultivars (Cézanne, Escrito, and Hugo). Mean-centered data scaled to unit variance were used for standard PCA and HPCA. (A) Standard PCA. (B) Multiblock HPCA super scores plot. (C) HPCA block scores plot. (D) HPCA block loadings plot (the most influential metabolites, selected by *N*-way ANOVA ($P < 0.022$), are labeled).

and ethanol (Figure 5C). Parts E and F of Figure 5 represent sucrose and alanine, respectively, as detected by GC−EI-TOFMS and are included to illustrate metabolite gradients detected by both platforms. To represent metabolite gradients detected by

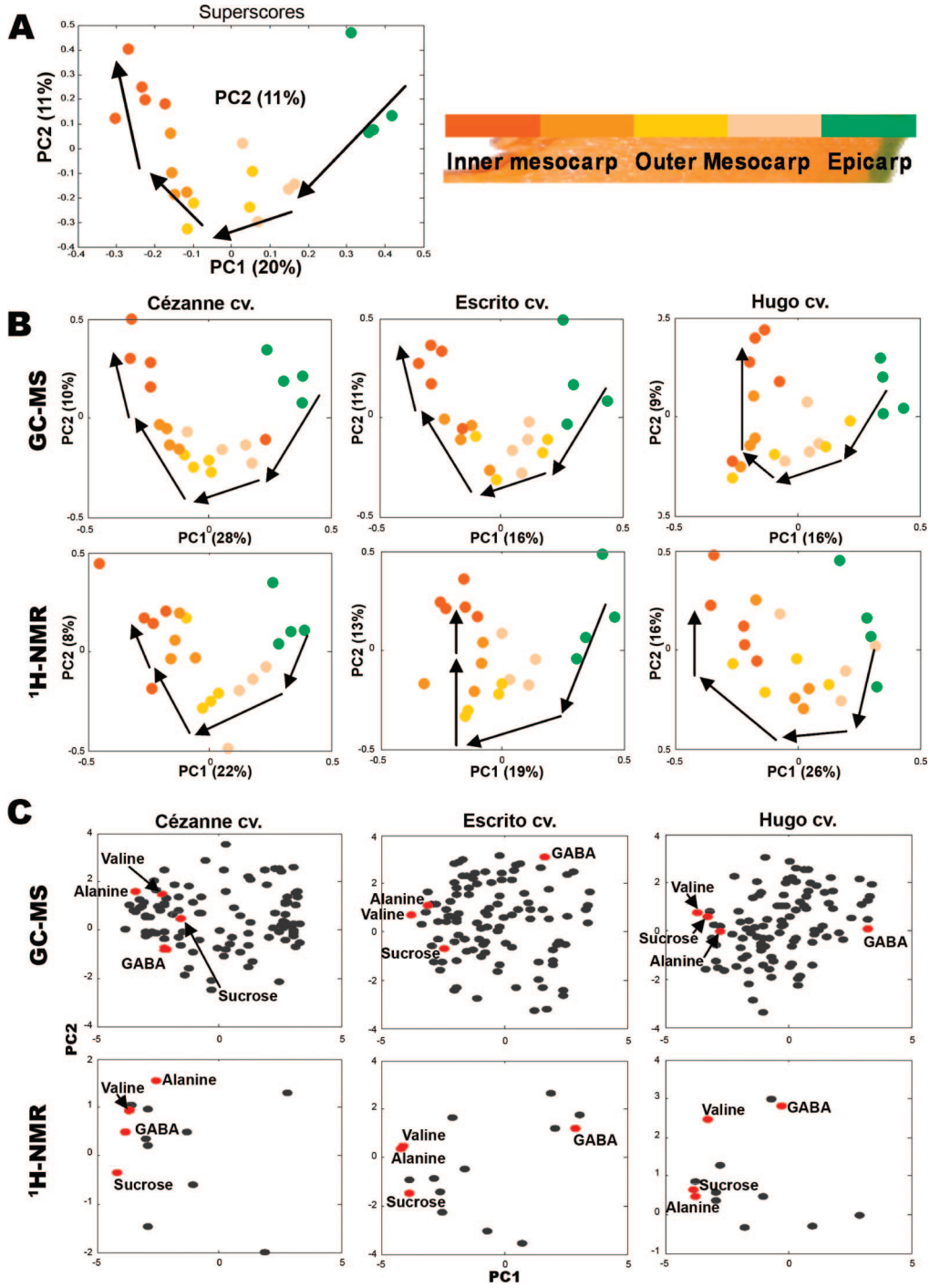

Figure 4. Correlation analysis of all detected GC−EI-TOFMS and quantified ^{1}H NMR features by multiblock HPCA for an equalized number of melon sections for each of the three cultivars (Cézanne, Escrito, and Hugo). Mean-centered data scaled to unit variance were used. (A) HPCA super scores plot. (B) HPCA block scores plot. (C) HPCA loadings plot.

GC−EI-TOFMS parts G and H of Figure 5 are included which represent serine and ethanolamine, respectively. The three cvs. all showed strong sucrose gradients in both analytical data sets, although the ^{1}H NMR trends were more defined. This may result from the differences in sampling and extraction (i.e., different melon slices, different extraction buffers, or analysis of juice compared to flesh). From these analyses it has become clear that ^{1}H NMR is more appropriate for the study of melon extracts containing high sugar levels; however, GC−EI-TOFMS offers improved sensitivity detecting amino and organic acids beyond the ^{1}H NMR detection limit. Sucrose concentration increased regularly from the epicarp and the green flesh to the inner mesocarp for all three cvs.; a gradient of two amino acids was also confirmed (alanine and valine). ^{1}H NMR highlighted the valine gradient more clearly than GC−EI-TOFMS; however, the alanine gradients were very similar between both instruments; further, GC−EI-TOFMS also discriminated β-alanine, the gradient of which was highly complementary to alanine. GC−EI-TOFMS confirmed two further amino acid gradients, serine and lysine, which again increased in concentration from the green epicarp flesh to the inner mesocarp. Finally, ^{1}H NMR revealed that ethanol showed a

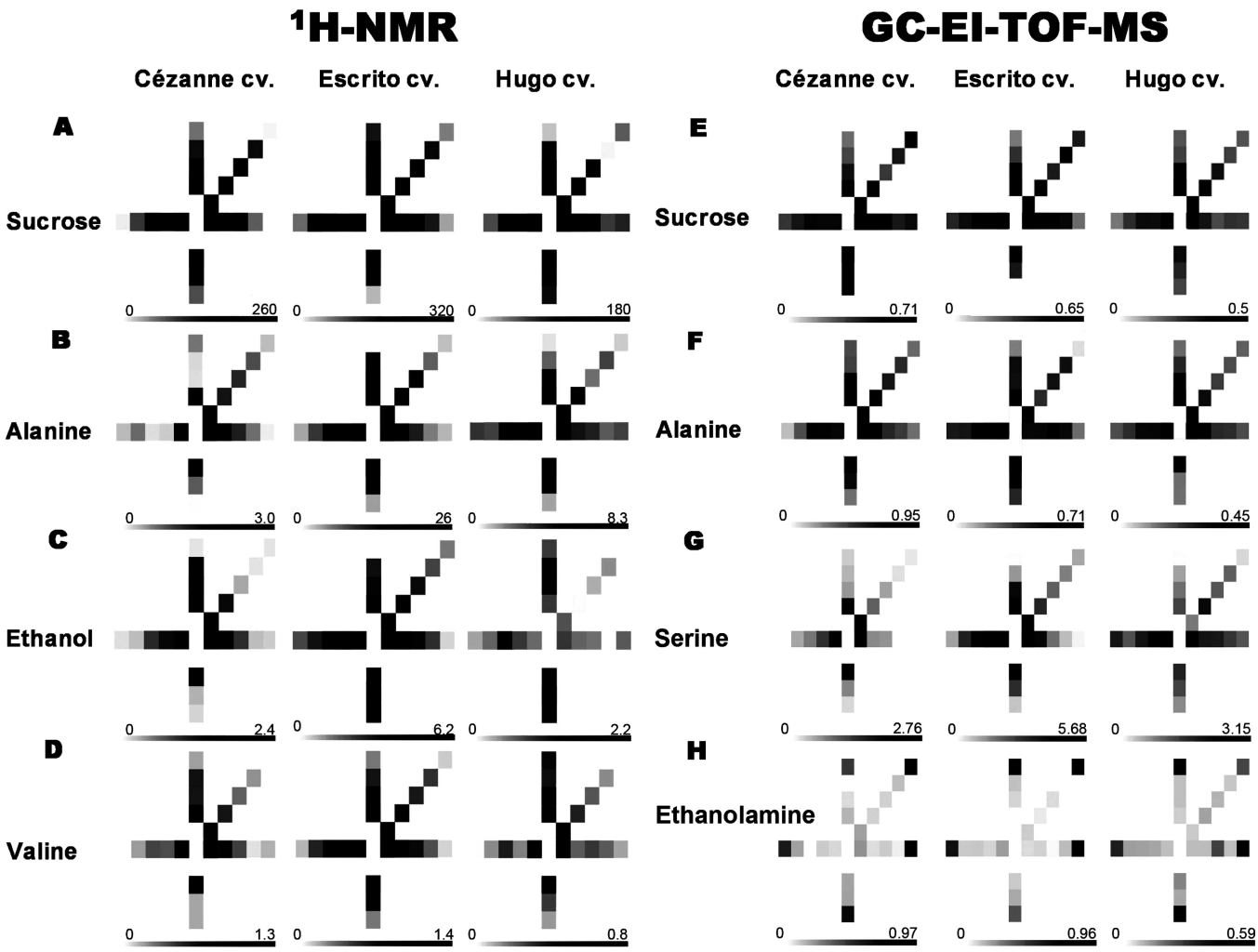

Figure 5. Schematic representation of metabolites concentration gradients in melon slices as detected by ^{1}H NMR for (A) sucrose, (B) alanine, (C) ethanol, (D) valine, and GC−EI-TOFMS for (E) sucrose, (F) alanine, (G) serine, (H) ethanolamine.

clear gradient being more concentrated in the center of the fruit, except for Hugo cv. where the concentration was more heterogeneous. A reverse trend (the metabolite concentration being greatest in the epicarp) was observed for ethanolamine, glycerol, and malic acid, in the GC−EI-TOFMS data.

To confirm the tendencies observed on the heat maps, univariate analyses were performed. The concentrations of each metabolite detected by ^{1}H NMR were submitted to *N*-way ANOVA and Tukey's test, and abundances for metabolites detected by GC−EI-TOFMS to *N*-way ANOVA, to verify that the observed spatial differences were significant. The sections were grouped as explained in Materials and Methods, and the results are presented for ^{1}H NMR in Supporting Information Table S4 and GC−EI-TOFMS in Supporting Information Table S5. The concentration of some metabolites appeared to be changing depending on the location in the slice. The sucrose concentration detected by ^{1}H NMR increased greatly from the epicarp to the inner mesocarp. A similar pattern, common to the three cvs., was shown for alanine, valine, and ethanol. Aspartic acid also showed a significant difference with higher concentration close to the center of the fruit. Some other metabolites appeared to be organized in gradients but were not common to the three cvs., or concentration differences between locations were not statistically significant (e.g., glx, threonine, and citric acid for ^{1}H NMR). The ^{1}H NMR detected that the threonine concentration increased from the epicarp to the inner mesocarp, but only at a significant level in one cv. However, GC−EI-TOFMS with its lower detection limit better illustrated the threonine gradient and thus need to be confirmed with further biological replication (Supporting Information Table S5). The metabolite gradients common to the three cvs. have fueled hypotheses generation in terms of fruit physiology regarding for instance changes in activities of key enzymes of sugar metabolism in cucurbits[30,38] or shifts in metabolism in relation with oxygen availability[39,40] in the fruit tissues.

CONCLUSIONS

In this study, a metabolomics approach with complementary analytical methods was used to biochemically characterize three cvs. of melons. It highlighted a number of metabolite concentration differences between the cvs. Our study also showed that metabolic ^{1}H NMR profiling in methanol-d_4 can be used directly on fresh juice for a fast screening of metabolite spatial

(38) Gao, Z.; Schaffer, A. A. *Plant Physiol.* **1999**, *119*, 979–987.
(39) Ismond, K. P.; Dolferus, R.; De Pauw, M.; Dennis, E. S.; Good, A. G. *Plant Physiol.* **2003**, *132*, 1292–1302.
(40) Schotsmans, W.; Verlinden, B. E.; Lammertyn, J.; Nicolaï, B. M. *Postharvest Biol. Technol.* **2003**, *29*, 155–166.

localization in fleshy fruit such as melon. The GC−EI-TOFMS spatial method provided useful information which correlated with the results of the ^{1}H NMR study, via an indirect comparison of independently performed PCA and multiblock HPCA, and via a correlation-based superblock HPCA for direct comparison of both analytical data sets. GC−EI-TOFMS also indicated a number of gradients for metabolites beyond the detection limits of ^{1}H NMR. The study has shown that GC−EI-TOFMS may not be as apt for the analysis of bulk metabolic components as ^{1}H NMR, but still provides data of much value in this combined multiplatform approach. The multiblock HPCA method has great potential for correlation of data between metabolomic platforms other than just between ^{1}H NMR and GC−EI-TOFMS, although due to both techniques covering primary metabolism they are particularly comparable. Useful insights toward more appropriate metabolite extraction and analytical methods for melon analysis by GC−EI-TOFMS were also gained. A number of metabolite gradients from the epicarp to the inner mesocarp were found in all cvs. that can be related to metabolism changes. The remaining question is to know how these gradients become established during the fruit development; further investigation on several replicate fruits at different stages of growth could answer this point and is currently being undertaken.

ACKNOWLEDGMENT

This study was partially funded by the EU within the plant metabolomics project META-PHOR (FOOD-CT-2006-036220). We gratefully thank Sylvie Bochu and Françoise Leix-Henry from CEFEL (France) for providing the melons, and Drs. Arthur A. Schaffer (Volcani Centre, Israel) and Robert Hall (Plant Research International, Holland) for critical reading of the manuscript. The first two authors contributed equally to this work.

SUPPORTING INFORMATION AVAILABLE

Supplementary methods including (1) melon sample growth, collection, and handling, (2) extraction of polar metabolites from ground flesh samples for ^{1}H NMR analysis, (3) spatial analysis of pH, refractive index, and osmolarity of melon juice, (4) ^{1}H NMR analysis of melon juice, and supplementary data including (Figure S-1) baseline ^{1}H NMR metabolite profiling of ground samples of melon fruit, (Table S-1) ^{1}H chemical shifts used for identification and/or quantification of metabolites in polar extracts (in D_2O) and in juices (in CD_3OD/H_2O) of melon fruits, (Table S-2) concentration of metabolites identified in flesh polar extracts of three melon cultivars (Cézanne, Escrito, and Hugo), (Table S-3) pH, osmolarity, and refractive index measurements of juice collected in different locations of three melon cultivars (Cézanne, Escrito, and Hugo), (Table S-4) concentration changes, depending on the location in the slice of one melon, for metabolites identified in three melon cultivars (Cézanne, Escrito, and Hugo), and (Table S-5) a selection of unambiguously identified metabolites detected in polar extracts of three melon cultivars (Cézanne, Escrito, and Hugo) by GC−EI-TOFMS. This material is available free of charge via the Internet at http://pubs.acs.org.

Received for review December 2, 2008. Accepted March 2, 2009.

AC9001996