Metabolomic Differentiation of *Cannabis sativa* Cultivars Using ^{1}H NMR Spectroscopy and Principal Component Analysis

Young Hae Choi,[†] Hye Kyong Kim,[†] Arno Hazekamp,[†] Cornelis Erkelens,[‡] Alfons W. M. Lefeber,[‡] and Robert Verpoorte*[,†]

Division of Pharmacognosy, Section Metabolomics, Institute of Biology, Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands, and Division of NMR, Leiden Institute of Chemistry, Gorlaeus Laboratories, Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands

Received February 18, 2004

The metabolomic analysis of 12 *Cannabis sativa* cultivars was carried out by ^{1}H NMR spectroscopy and multivariate analysis techniques. Principal component analysis (PCA) of the ^{1}H NMR spectra showed a clear discrimination between those samples by principal component 1 (PC1) and principal component 3 (PC3) in cannabinoid fraction. The loading plot of PC value obtained from all ^{1}H NMR signals shows that Δ^9-tetrahydrocannabinolic acid (THCA) and cannabidiolic acid (CBDA) are important metabolites to differentiate the cultivars from each other. The discrimination of the cultivars could also be obtained from a water extract containing carbohydrates and amino acids. The level of sucrose, glucose, asparagine, and glutamic acid are found to be major discriminating metabolites of these cultivars. This method allows an efficient differentiation between cannabis cultivars without any prepurification steps.

Cannabis (marihuana) preparations, derived from *Cannabis sativa* L. (Cannabinaceae), are considered as one of the most dangerous illicit drugs because of their narcotic and addictive properties. Nevertheless, their promising therapeutic potential has driven researchers to pay attention to various possible clinical uses,[1] such as for menstrual cramps and convulsions, inflamed tonsils, migraine and headaches,[2] glaucoma,[3] asthma,[4] and pain relief.[5] The active constituents belong to a group of compounds classified as cannabinoids inclusive of Δ^9-tetrahydrocannabinol (THC), Δ^8-tetrahydrocannabinol, cannabigerol (CBG), and cannabinol (CBD).[6,7] In addition to these major cannabinoids, approximately 60 further cannabinoids including metabolites have been isolated so far.[6,7]

Cannabis is widely distributed throughout the world, both in cultivated forms and as wild plants.[8] Whether a cannabis plant predominantly produces fiber (hemp) or narcotic resin is governed by both genetic and climatic factors.[9] Studies on a large number of cannabis plants originating from different parts of the world have led to the acceptance that a number of chemical races of *C. sativa* exist. Several classification systems have been proposed to distinguish psychoactive and fiber strains of cannabis based on their cannabinoid composition.[7,9] For example, Grlic proposed a classification system including the use of a selection of chemical, spectroscopic, microbiological, and pharmacological tests whose results were dependent on the levels of cannabidiolic acid (CBDA), cannabidiol (CBD), Δ^9-tetrahydrocannabinol (Δ^9-THC), and cannabinol (CBN) in the sample.[10] These markers were regarded as indicative of successive stages of "ripening" or subsequent decomposition of the resin. In addition, some methods based on quantitative analysis of specific cannabinoids such as Δ^9-THC, cannabigerol (CBG), CBD, CBN, Δ^9-tetrahydrocannabivarin (THCV), cannabidivarin (CBDV), cannabichromene (CBC), and cannabigerol monomethyl ether (CBGM) have been suggested.[7,11−13] Despite these extensive studies on the chemical composition of cannabis, previous results are limited to the differentiation of the drug and fiber types of cannabis. However, large-scale commercial production takes place in some areas (e.g., The Netherlands) as a

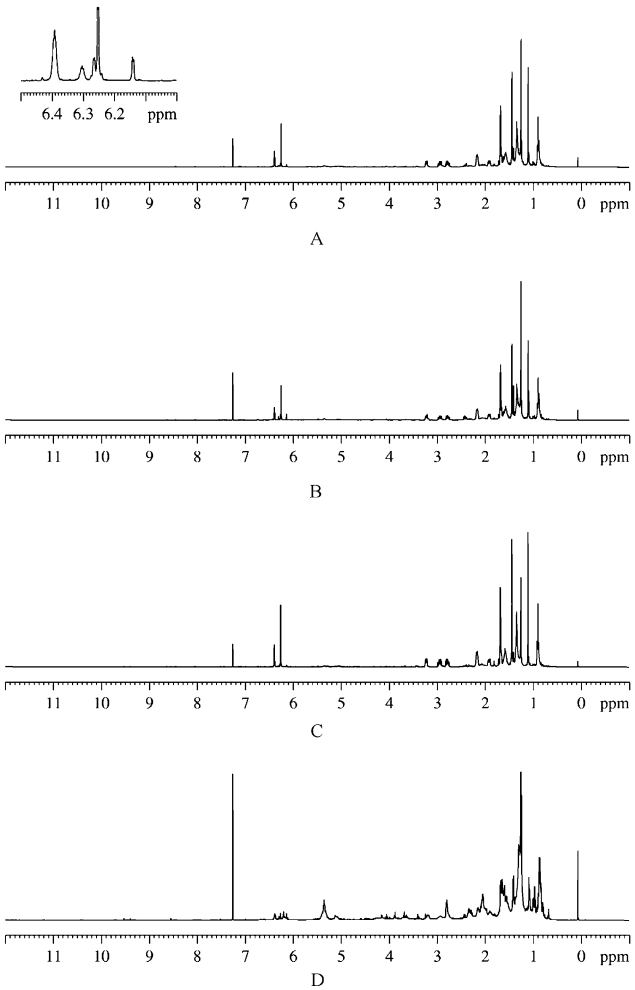


Figure 1. ^{1}H NMR spectra of $CHCl_3$ extract of *Cannabis sativa* Simm 4 flowers (A), Simm 12 flowers (B), Simm 18 flowers (C), and Simm 4 leaves (D).

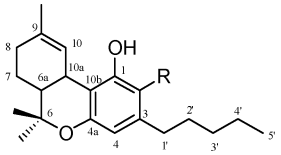
* To whom correspondence should be addressed. Tel: +31 71 527 4510. Fax: +31 71 527 4511. E-mail: E-mail: verpoort@chem.leidenuniv.nl.
[†] Institute of Biology.
[‡] Institute of Chemistry.

10.1021/np049919c CCC: $27.50
Published on Web 05/04/2004

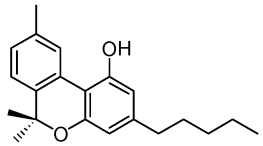
source of medicinal use. There are numerous varieties cultivated as drug-type cannabis species. The classification or discrimination of each cultivar is still unclear when the classical methods using the content of cannabinoids are applied, because they all produce similar amounts of cannabinoids. For the reliable differentiation of cannabis cultivars, a systemic method including a variety of metabolites (metabolomic profiling) is desirable.

The term "metabolome" has been used to describe the observable chemical profile or fingerprint of the metabolites in whole tissues.[14,15] In metabolite profiling, it would be preferable to use a wide-spectrum chemical analysis technique, which is rapid, reproducible, and stable in time while needing only a very basic sample preparation. NMR spectroscopy is one of the techniques that could meet such demands. The last few decades, a number of techniques have been devised to develop NMR spectroscopy as a fingerprinting tool for the interpretation and quality assessment of industrial and natural products and multivariate or pattern recognition techniques such as the well-described principal component analysis (PCA). Recently, the NMR and PCA method has been applied to the metabolomic profiling of several kinds of wine,[16] coffee,[17] juice,[18] beer,[19] and some plants.[20,21]

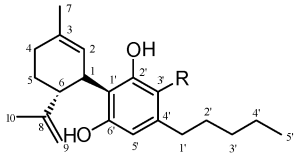
In this study, we report a ^{1}H NMR spectroscopic method coupled with PCA for the metabolomic analysis of 12 cultivated forms of *C. sativa*. This may lead to the differentiation of cannabis cultivars based on a variety of metabolites.



R = H, Δ9-tetrahydrocannabinol (THC)
R = COOH, Δ9-tetrahydrocannabinolic acid (THCA)



cannabinol (CBN)



R = H, cannabidiol (CBD)
R = COOH, cannabidiolic acid (CBDA)

Results and Discussion

For the identification of nonpolar metabolites such as cannabinoids, $CHCl_3$ extracts were investigated. Similar metabolomic patterns were observed by visual inspection of ^{1}H NMR spectra of the $CHCl_3$ extracts of the various *C. sativa* flowers (Figure 1). Δ^9-Tetrahydrocannabinolic acid (THCA) was the predominant compound in $CHCl_3$ extracts of *C. sativa* flowers. The characteristic signals were well distinguishable in the ^{1}H NMR spectrum of $CHCl_3$ extracts. H-10 at δ 6.39 (brs), H-4 at δ 6.25 (s), H-10a at δ 3.23 (dm, J = 7.0 Hz), H-1′ at δ 2.94 (m) and 2.78 (1H, m), C-9 methyl at δ 1.68 (s), C-6 β-methyl at δ 1.44 (s), C-6 α-methyl at δ 1.11 (s), and H-5′ at δ 0.90 (t, J = 6.9 Hz) were observed as major signals in the $CHCl_3$ extract (Figure 1). A hydroxyl proton of the compound was also detected at δ 12.19 (s). In addition to these signals of Δ^9-THCA, those of Δ^9-THC and CBN were detected as minor signals. The signals of H-10, H-4, H-2, and H-1″ of Δ^9-THC were detected at δ 6.30 (brs), 6.26 (d, J = 1.6 Hz), 6.14 (d, J = 1.6 Hz), and 2.42 (m), respectively. For CBN, H-10 at δ 8.16 (s), H-7 at δ 7.14 (d, J = 7.9 Hz), H-8 at δ 7.07 (d, J = 7.9 Hz), and H-4 at δ 6.43 (s) were clearly distinguishable in the ^{1}H NMR spectra of each $CHCl_3$ extract.

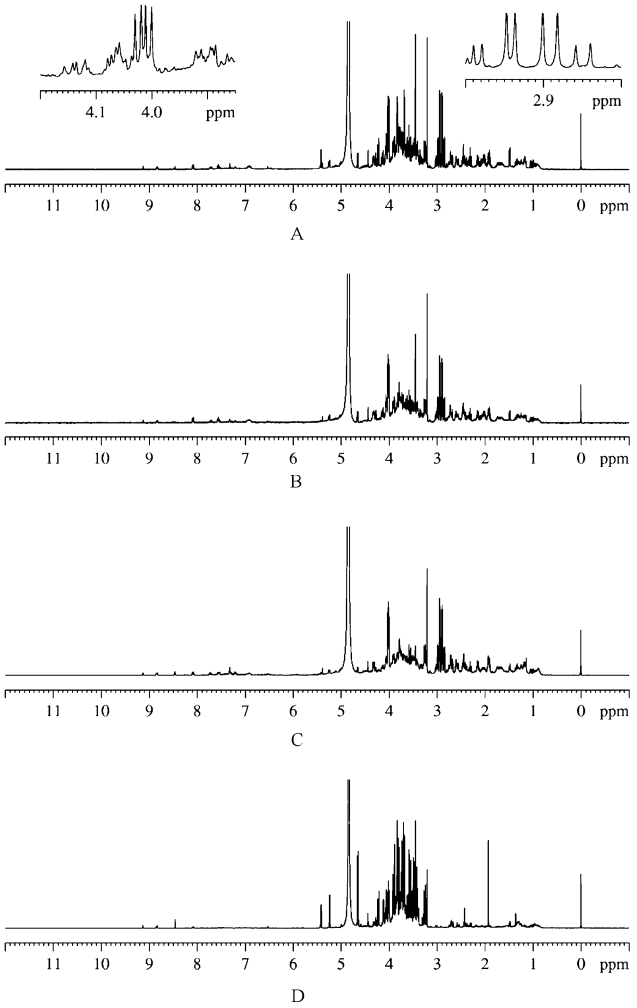


Figure 2. ^{1}H NMR spectra of water extract of *Cannabis sativa* Simm 4 flowers (A), Simm 12 flowers (B), Simm 18 flowers (C), and Simm 4 leaves (D).

The extract of cannabis leaves (Simm 4) shows a quite different ^{1}H NMR spectrum from that of flowers (Figure 1D). The content of cannabinoids is somewhat lower than that of flowers. Moreover, the major cannabinoid was not Δ^9-THCA but Δ^9-THC. The signal at δ 6.38 (brs) was thought to be that of H-10a of Δ^9-THCA, but it showed different ^{1}H−^{1}H COSY and HMBC correlations from the reference compound of Δ^9-THCA, and other ^{1}H NMR signals of the compound such as H-4 at δ 6.25 (s), H-10a at δ 3.23 (dm, J = 7.0 Hz), H-1′ at δ 2.94 (m), and δ 2.78 (1H, m), if observed at all, had a very low intensity. The methyl signals of fatty components at δ 1.2−1.4 and olefinic signals of fatty components and sterols at δ 5.0−5.5 were detected as major signals in the ^{1}H NMR spectra of each $CHCl_3$ extract of the leaves.

In the water extract (Figure 2), the major signals are shown in the anomeric signals of carbohydrates such as δ 5.42 (d, J = 3.8 Hz), 5.24 (d, J = 3.7 Hz), and 4.64 (d, J =

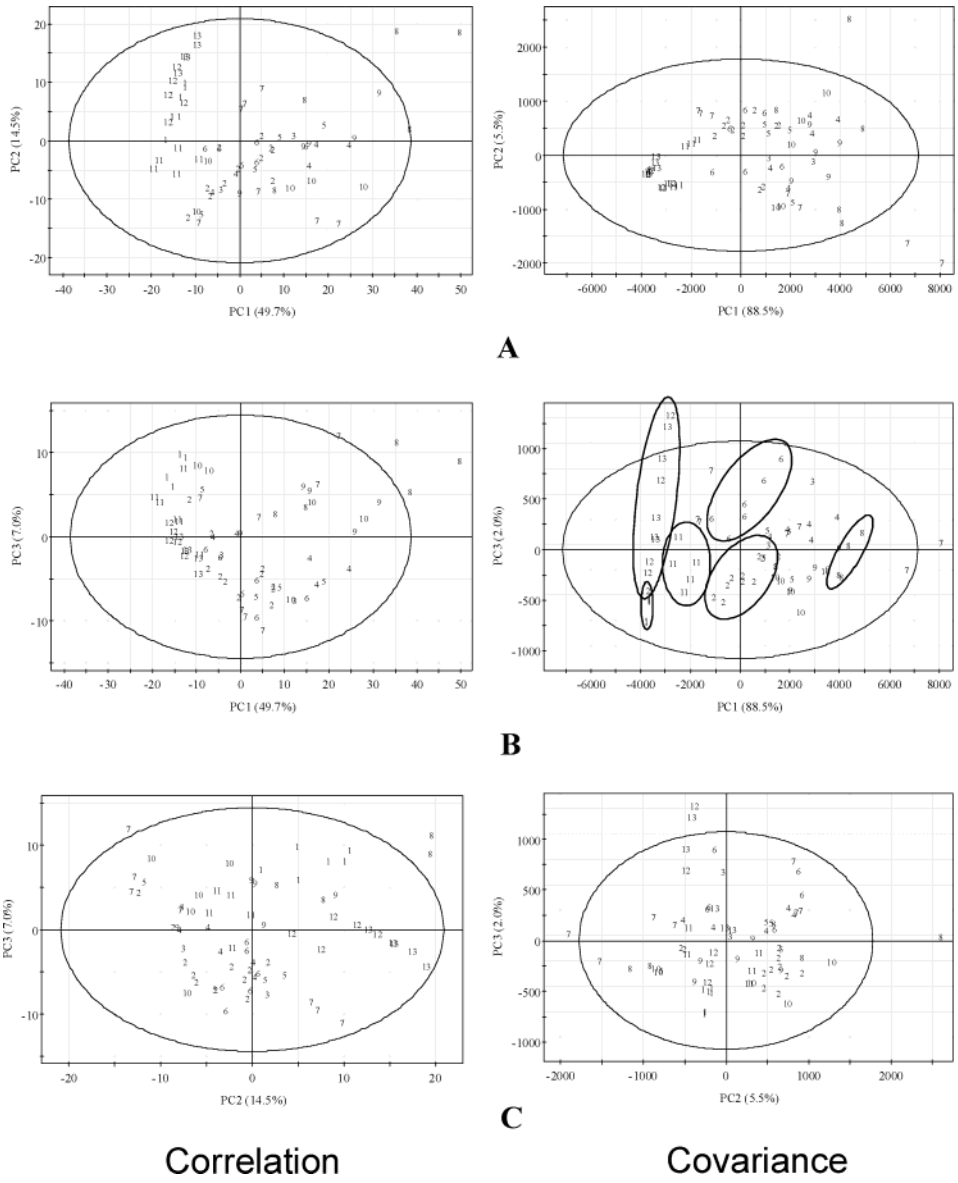


Figure 3. Score plot of principal component analysis of the $CHCl_3$ extracts of *Cannabis sativa* cultivars obtained by correlation and covariance method: (A) PC1 vs PC2, (B) PC1 vs PC3, (C) PC2 vs PC3. The ellipse represents the Hotelling T2 with 95% confidence level in score plots: (1) Simm 4 leaves, (2) Simm 2 flowers, (3) Simm 4 flowers, (4) Simm 5 flowers, (5) Simm 6 flowers, (6) Simm 12 flowers, (7) Simm 18 flowers, (8) Simm 27 flowers, (9) Simm 40 flowers, (10) Simm 49 flowers, (11) Simm 750A flowers, (12) Simm 750B flowers, (13) Simm 750C flowers.

8.0 Hz). They were assigned to the anomeric protons of sucrose, α-glucose, and β-glucose, respectively.[22] Another anomeric signal obtained from the fructose moiety of sucrose is also well distinguishable at δ 4.22 (d, J = 8.8 Hz). The residual proton signals of the sugars shown in a crowded region (δ 3.0−4.0) were assigned by the comparison of ^{1}H NMR spectra of the reference compounds and from ^{1}H−^{1}HCOSY and TOCSY spectra. For amino acids, HMBC spectra gave informative evidence about the assignments, together with their chemical shifts. H-2 or H-3 can correlate with the carbonyl group of the amino acid. As a result, the most abundant signals at δ 4.01 (dd, J = 7.6 Hz, 4.3 Hz), 2.96 (dd, J = 16.9 Hz, 4.3 Hz), and 2.87 (dd, J = 16.9 Hz, 7.6 Hz) were assigned to be H-2, H-3a (close to NH_2), and H-3b (close to H-2) of asparagine. In addition, H-3 of alanine at δ 1.48 (d, J = 7.3 Hz) and H-4 of valine at δ 1.05 (d, 6.8 Hz) and 0.99 (d, 6.8 Hz) were assigned. Other possible ^{1}H NMR signals of amino acids such as glutamic acid at δ 2.14 (H-4, m) and 2.38 (H-3, m) were detected in the ^{1}H NMR spectra of the water extract of flowers. The leaves show a higher level of carbohydrates such as sucrose and glucose but lower amount of amino acids than the flowers.

Principal component analysis is an unsupervised clustering method requiring no knowledge of the data set and acts to reduce the dimensionality of multivariate data while preserving most of the variance within.[23] The principal components can be displayed in a graphical fashion as a "scores" plot. This plot is useful for observing any groupings in the data set. PCA models are constructed using all the samples in the study. Coefficients by which the original variables must be multiplied to obtain the PC are called loadings. The numerical value of a loading of a given variable on a PC shows how much the variable has in common with that component.[24] Thus for NMR data, loading plots can be used to detect the spectral areas responsible for the separation in the data. Generally, this separation takes place in the first two or three principal components (PC1, PC2, and PC3). If the data are mean-centered with no scaling, then a covariance matrix is produced, but if the data mean-centered and the columns of the data matrix scaled to unit variance, a correlation

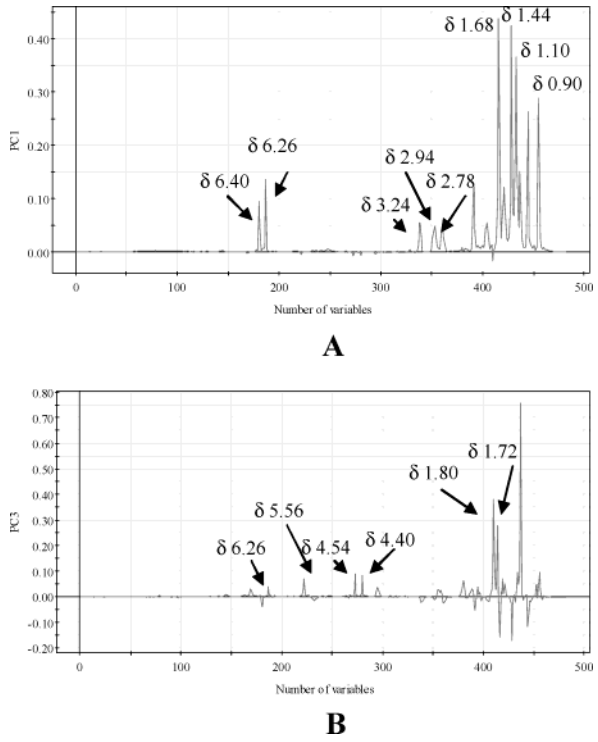


Figure 4. Loading plot of principal component analysis of the $CHCl_3$ extracts of *Cannabis sativa* cultivars obtained by covariance method: (A) PC1, (B) PC3.

matrix is produced. An advantage of the covariance matrix is that the loadings retain the scale of the original data. For the correlation method, however, a weaker signal to possess a discriminatory power can be considered as the same level as stronger signals.[25,26] In this study, both methods were evaluated, but the covariance method showed better separation results (Figure 3). For the data set obtained from the analysis of $CHCl_3$ extract, a six-component model could explain 99% of the variance, with the first three components explaining 96%. Among the PCs, the combination of PC1 and PC3 can give well-separated clusters for all cannabis cultivars (Figure 3). Examination of the scores and loading plots for PC1 versus PC3 showed that *C. sativa* Simm 4 leaves and the flowers of Simm 2, 5, 12, 27, 750A, 750B, and 750 C were clearly separated from each other (Figure 3). However, Simm 18 flowers showed quite a broad area in the score plot because the content of metabolites including cannabinoids was not reproducible due to collection time or variation between individual plants. Investigation of the loading plot of PC1 indicated that the first component explained the variance in Δ^9-THCA due to the signals at δ 6.40 (H-10), 6.26 (H-4), 3.24 (H-10a), 2.94 (H-1′), 2.78 (H-1′), 1.68 (C-9 methyl), 1.44 (C-6 β-methyl), 1.11 (C-6 α-methyl), and 0.90 (H-5′) (Figure 4). Δ^9-THCA had an effect on the positive value of PC1. Therefore, cannabis Simm 27 flowers showed the highest content of Δ^9-THCA. For PC3, CBDA is an important factor affecting the separation. In the ^{1}H NMR spectra of the plants of the $CHCl_3$ extracts, the signals at δ 6.26, 5.56, 4.54, 4.40, 1.80, and 1.72 greatly affected the PC3 value. The signals were assigned as H-5′, H-2, H-9 (*trans*), H-9 (*cis*), H-7, and H-10 of CBDA based on comparison of compound CBDA.[27] In the ^{1}H NMR spectra, the signals of CBDA showed quite low intensity, which made it difficult to identify the compound. However, the loading plot, processed by PC values, showed very clear signals of CBDA. On the basis of these data, Simm 12 flowers were concluded to contain relatively high amounts of CBDA.

In the ^{1}H NMR spectra of the water extracts of cannabis

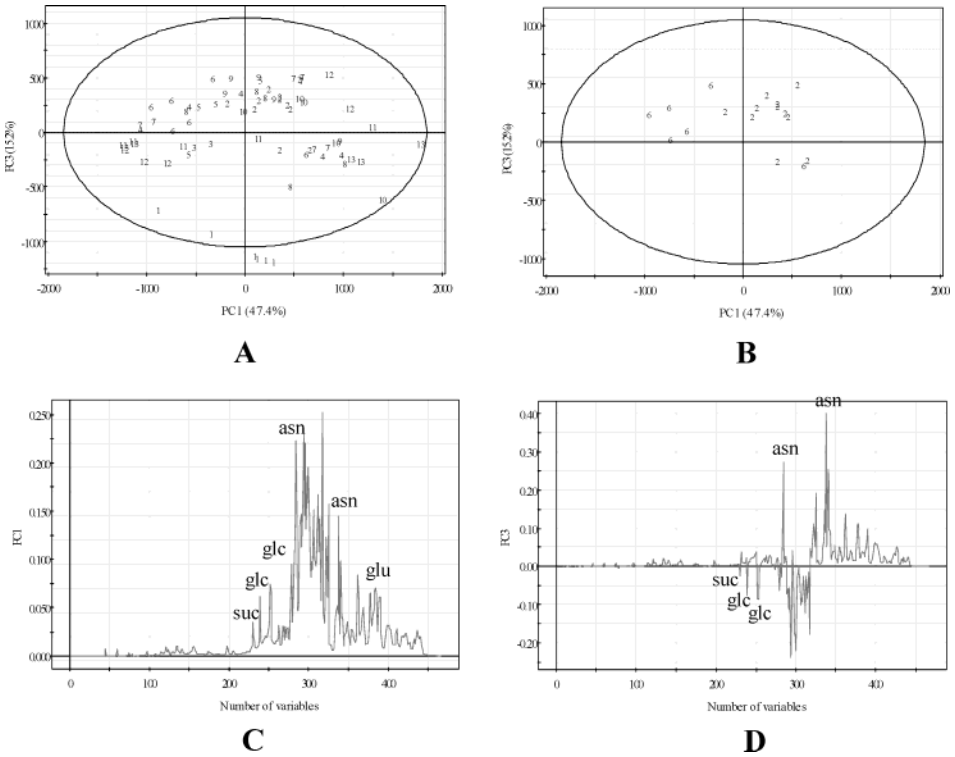


Figure 5. Score and loading plot of principal component analysis of the water extracts of *Cannabis sativa* cultivars obtained by principal component analysis (covariance method). The ellipse represents the Hotelling T2 with 95% confidence in score plots: (A) score plot of the whole cultivars, (B) score plot of Simm 2 and Simm 12 flowers, (C) loading plot of PC1, (D) loading plot of PC 3. (1) Simm 4 leaves, (2) Simm 2 flowers, (3) Simm 4 flowers, (4) Simm 5 flowers, (5) Simm 6 flowers, (6) Simm 12 flowers, (7) Simm 18 flowers, (8) Simm 27 flowers, (9) Simm 40 flowers, (10) Simm 49 flowers, (11) Simm 750A flowers, (12) Simm 750B flowers, (13) Simm 750C flowers (suc, sucrose; glc, glucose; asn, asparagine; glu, glutamic acid).

flowers, as mentioned earlier, primary metabolites such as carbohydrates and amino acids can be detected. The covariance method with PC1 and PC3 also shows clear separations. For the data set obtained from the analysis of each water extract, a six-component model explained 96% of the variance, with the first three components explaining 83%. The separation of each water extract is less than that of each $CHCl_3$ extract. However, Simm 2 and Simm 12, which are commonly used for medicinal purposes in The Netherlands, were clearly separated from one another (Figure 5B). The main factors to differentiate the flowers were carbohydrates and amino acids. Simm 2 flowers were found to contain more glucose, sucrose, asparagine, and glutamic acid from a loading plot (Figure 5).

In this study, multivariate statistical methods have been used to analyze the ^{1}H NMR spectra of the flowers of *C. sativa* cultivars. This method was found to be an ideal method to compare *C. sativa* cultivars with each other based on a variety of metabolites. ^{1}H NMR spectroscopy shows the quantity and quality of the diverse metabolites present in the plants at the same time without any chromatographic purification, and each procedure could be performed within 12 min. For reliable and easy comparison of the metabolomic profiling, the large ^{1}H NMR data set obtained from various metabolites can be reduced to PC1, PC2, or PC3 using principal component analysis. Samples of the same variety showed clustering in PC space and gave a high success rate in assigning samples to the groups. The PC loadings were examined in order to obtain information about the chemical basis for the clustering behavior. Although the major metabolites were important, it was apparent that minor constituents could also contribute significantly to the discrimination. Considering all the results, ^{1}H NMR/PCA seems to be a very promising tool for the authentication and quality control of cultivars of *C. sativa*.

Experimental Section

Plant Material. Flowers of *Cannabis sativa* L. (cultivars SIMM 2, 4, 5, 6, 12, 18, 27, 40, 49, 750A, 750B, and 750C) and leaves (cultivar SIMM 04) harvested in October 2002 were obtained from Stichting Institute for Medicinal Marijuana (SIMM) in Rotterdam, The Netherlands. The plant materials were air-dried in the dark at ambient temperature for 2 weeks.

Solvents and Chemicals. First-grade chloroform and methanol were purchased from Merck Biosolve Ltd. (Valkenswaard, The Netherlands). $CDCl_3$ (99.96%) and D_2O (99.00%) were obtained from Cambridge Isotope Laboratories Inc. (Miami, FL), and NaOD was purchased from Cortec (Paris, France). Potassium dihydrogen phosphate was from Merck (Darmstadt, Germany). Hexamethyldisilane (HMDS) and trimethylsilane propionic acid sodium salt (TSP) were purchased from Merck (Darmstadt, Germany).

Extraction. Three hundred milligrams of ground material was transferred into a centrifuge tube. Five milliliters of 50% aqueous methanol and 5 mL chloroform were added to the tube followed by vortexing for 30 s and sonication for 1 min. The material was then centrifuged at 3000 rpm for 20 min. The extraction was performed twice. The water and chloroform fractions were collected separately to a 10 mL round-bottomed flask and dried in a rotary vacuum evaporator.

NMR Measurements. Potassium dihydrogen phosphate was added to D_2O as a buffering agent. The pH of the D_2O for NMR measurements was adjusted to 6.0 using a 1 N NaOD solution. All spectra were recorded on a Bruker AV-400 NMR instrument operating at a proton NMR frequency of 400.13 MHz. For each sample, 128 scans were recorded with the following parameters: 0.126 Hz/point, pulse width (PW) = 4.0 μs (90°), and relaxation delay (RD) = 1.0 s. FIDs were Fourier transformed with LB = 1.0 Hz, GB = 0, and PC = 1.0. For quantitative analysis, peak height was used. The spectra were referenced to the residual solvent signal of $CDCl_3$ (7.26 ppm) for the $CHCl_3$ extract and TSP at 0.00 ppm for the water extract. Hexamethyldisilane (HMDS, 0.01%, v/v) for $CDCl_3$ and trimethylsilane propionic acid sodium salt (TSP, 0.01%, w/v) were used as internal standards.

Data Analysis. The ^{1}H NMR spectra were automatically reduced to ASCII files using AMIX (v. 3.7, Bruker Biospin). Spectral intensities were scaled to HMDS for the $CHCl_3$ extract and TSP for the water extract and reduced to integrated regions of equal width (0.02 ppm) corresponding to the region δ 0.30−10.00. The region δ 4.6−5.8 was excluded from the analysis because of the residual signal of water. Principal component analysis was performed with the SIMCA-P software (v. 10.0, Umetrics, Umeå, Sweden).

Acknowledgment. The financial support of the van Leersumfonds (KNAW) is gratefully acknowledged.

References and Notes

(1) Formukong, E. A.; Evans, A. T.; Evans, F. J. *Phytother. Res.* **1989**, *3*, 219−235.
(2) Russo, E. *Pain* **1988**, *76*, 3−8.
(3) Helper, R. S.; Frank, I. M. *JAMA, J. Am. Med. Assoc.* **1971**, *217*, 1392.
(4) Abdoud, R. T.; Sanders, H. D. *Chest* **1976**, *70*, 480−485.
(5) Iversen, L.; Chapman, V. *Curr. Opin. Pharmacol.* **2002**, *2*, 50−55.
(6) Mechoulam, R.; McCallum, K.; Burstein, S. *Chem. Rev.* **1976**, *76*, 75−112.
(7) Turner, C. E.; Elsohly, M. A.; Boeren, G. *J. Nat. Prod.* **1980**, *43*, 169−234.
(8) Raman, A. In *Cannabis*; Brown, D. T., Ed.; Harwood Academic Publishers: Amsterdam, 1998; Chapter 2, pp 29−54.
(9) Raman, A.; Joshi, A. In *Cannabis*; Brown, D. T., Ed.; Harwood Academic Publishers: Amsterdam, 1998; Chapter 3, pp 55−70.
(10) Grlic, L. *Bull. Narc.* **1968**, *20*, 25−29.
(11) Waller, C. W.; Scigliano, J. A. *Natl. Acad. Sci. NRC* **1970**, *4*, 28−32.
(12) Fetterman, P. S.; Keith, E. S.; Waller, C. W.; Guerrero, O.; Doorenbos, N. J.; Quimby, M. W. *J. Pharm. Sci.* **1971**, *60*, 1246−1249.
(13) Small, E.; Beckstead, H. D. *J. Nat. Prod.* **1973**, *36*, 144−165.
(14) Ott, K.-H.; Aranibar, N.; Singh B.; Stockton, G. W. *Phytochemistry* **2003**, *62*, 971−985.
(15) Sumner, L. W.; Mandes, P.; Dixon, R. A. *Phytochemistry* **2003**, *62*, 817−836.
(16) Brescia, M. A.; Caldarola, V.; De Giglio, A.; Benedetti, D.; Fanizzi, F. P.; Sacco, A. *Anal. Chim. Acta* **2002**, *458*, 177−186.
(17) Charlton, A. J.; Farrington, W. H. H.; Brereton, P. *J. Agric. Food Chem.* **2002**, *50*, 3098−3103.
(18) Vogels, J. T. W. E.; Terwel, L.; Tas, A. C.; Van den Berg, F.; Dukel, F.; Van der Greef, J. *J. Agric. Food Chem.* **1996**, *44*, 175−180.
(19) Duarte, I.; Barros, A.; Belton, P. S.; Righelato, R.; Spraul, M.; Humpfer, E.; Gil, A. M. *J. Agric. Food Chem.* **2002**, *50*, 2475−2481.
(20) Nord, L. I.; Kenne, L.; Jacobsson, S. P. *Anal. Chim. Acta* **2001**, *446*, 199−209.
(21) Forveille, L.; Vercauteren, J.; Rutledge, D. N. *Food Chem.* **1996**, *57*, 441−450.
(22) Agrawal, P. K. *Phytochemistry* **1992**, *31*, 3307−3330.
(23) Goodacre, R.; Shann, B.; Gilbert, R. J.; Timmins, E. M.; McGovern, A. C.; Kell, D. B.; Logan N. A. *Anal. Chem.* **2000**, *72*, 119−127.
(24) Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wold, S. *Multi- and Megavariate Data Analysis*; Umetrics Academy: Umeå, Sweden, 2001; pp 43−69.
(25) Belton, P. S.; Colquhoun, I. J.; Kemsley, E. K.; Delgadilo, I.; Roma P.; Dennis, M. J.; Sharman, M.; Holmes, E.; Nicholson, J. K.; Spraul, M. *Food Chem.* **1998**, *61*, 207−213.
(26) Ward, J. L.; Harris, C.; Lewis, J.; Beale, M. H. *Phytochemistry* **2003**, *62*, 949−957.
(27) Choi, Y. H.; Hazekamp, A.; Peltenburg-Looman, A. M. G.; Frédérich, M.; Erkelens, C.; Lefeber, A. W. M.; Verpoorte, R. *Phytochem. Anal.* **2004**, in press.

NP049919C