Fourier Transform Infrared Spectroscopy and Chemometrics as a Tool for the Rapid Detection of Other Vegetable Fats Mixed in Cocoa Butter

Royston Goodacre[a] and Elke Anklam[b,*]

[a]Institute of Biological Sciences, University of Wales, Aberystwyth, Ceredigion, SY23 3DD, Wales, United Kingdom and [b]European Commission, DG Joint Research Centre, Institute for Health and Consumer Protection, Food Products and Consumer Goods Unit, I-21020 Ispra, Italy

ABSTRACT: The Fourier transform infrared (FTIR) technique in combination with multivariate data evaluation was used to analyze a wide variety of cocoa butters (CB), cocoa butter equivalents (CBE), and mixtures thereof. The sample set consisted of 14 CB (10 pure from various geographical origins and 4 commercial mixtures), 18 CBE (12 mixtures and 6 pure CBE from kokum, illipé, and palm midfraction), and 154 mixtures of CB with CBE at various concentrations (ranging from 5 to 20%). A total of 186 samples were analyzed in triplicate. All CB and CBE were shown to have very characteristic FTIR spectra that gave highly reproducible fingerprints. The main vibrational modes were also elucidated. FTIR can easily be employed to distinguish between pure CB and pure CBE. With prior knowledge of which cocoa butter is present in mixtures, FTIR can be applied to distinguish between CB mixed with CBE at the 10 and 20% levels (corresponding to about 2 and 5% of CBE in chocolate). However, the study revealed that a single "global" statistical model (multilayer perceptron, radial basis functions, or partial least square regression) was not able to predict the precise level of addition. The FTIR approach detailed here shows great potential as a rapid screening method for distinguishing between pure vegetable fats and, we believe, could be extended to investigate mixtures of CB and CBE by the establishment of a database.

Paper no. J9857 in *JAOCS 78,* 993–1000 (October 2001).

KEY WORDS: Chemometrics, chocolate, cocoa butter (CB), cocoa butter equivalents (CBE), Fourier transform infrared (FTIR) spectroscopy.

According to the new European Chocolate Directive 2000/36/EEC (1), the addition of up to 5% of vegetable fats other than cocoa butter (CB), the so-called cocoa butter equivalents (CBE), is allowed in chocolate products. Permitted fats are palm oil, illipé (borneo tallow or tengkawang), sal, shea, kokum gurgi, and mango kernel. The composition of cocoa butter and alternative fats, as well the analytical approaches for identification and determination, has been reviewed recently (2,3). CBE resemble the chemical composition and physical properties of CB very closely and are easily mixable with CB. The major triglycerides are the same as in CB, making the detection and quantification of such an addition difficult. The uncertainty for predicting the CBE level in chocolate is mainly caused by the large variety of CB and CBE deriving from different geographical origins the world over. In addition, the detection and quantification of some CBE in CB mixtures is extremely difficult, e.g., that of illipé (4,5).

Recently, the performance of four chromatographic methods in combination with mulivariate statistical data analysis for the major components of fat in chocolate, i.e., triglycerides and fatty acids, was reported (6). This extensive study on a large variety of CB and CBE revealed that the most suitable method for quantification should be based on the analysis of the major components, the triglycerides (4). Minor components (e.g., tocopherols, -trienols, and sterene data) have been found to be of limited use for quantitative purposes; however, they could be additional indicators for the presence of other vegetable fats in chocolate (7,8).

Nevertheless, a perceived need still exists within official control laboratories for the availability of rapid screening methods for the quantification of such vegetable fats in chocolate in order to implement the new directive and to handle a large throughput of samples (9).

Fourier transform infrared spectroscopy (FTIR) is a rapid analytical technique that measures the vibrations of bonds within functional groups. With FTIR, a particular bond absorbs electromagnetic (EM) radiation at a specific wavelength; therefore, by interrogating a biological sample with EM radiation of many wavelengths in the mid-IR range (here defined as 4000 to 600 cm^{-1}), one can construct an infrared "fingerprint" of the original food sample under investigation (10,11). Because different bonds absorb or scatter different wavelengths of EM radiation, these biological infrared fingerprints are made up of the vibrational features of all biochemical components. Therefore, FTIR gives quantitative information about the total biochemical composition of a food sample without destroying it, and produces fingerprints that are reproducible and distinct for different biological materials (12–14).

In this study we describe the potential of FTIR to distinguish between CB and CBE and to detect CBE in mixtures with CB.

EXPERIMENTAL PROCEDURES

Samples and chemicals. CB and CBE were donated by commercial suppliers. Samples consisted of 14 CB (10 pure from various geographical origins and 4 commercial mixtures), 18

*To whom correspondence should be addressed. E-mail: elke.anklam@jrc.it

CBE [12 mixtures and 6 pure CBE (kokum, illipé, and palm midfraction)] and 154 mixtures of CB with CBE at various concentrations (ranging from 5 to 20%).

In order to obey the Beer-Lambert law and limit overabsorbing bands (14), the vegetable fat samples were diluted 1:10 with analytical-grade acetone. All samples were analyzed in triplicate (in Run 1, 159 total, and in Run 2, 417 total). Therefore in total, excluding experimental optimization, 576 spectra were collected.

Diffuse reflectance-absorbance FTIR. Five microliters of the above fat samples were evenly applied onto an aluminum plate (this measured 10 × 10 cm and had 100 wells cut into the surface). Prior to analysis, the samples were oven-dried at 50°C for 30 min to evaporate the acetone. Samples were run in triplicate. The FTIR instrument used was the Bruker IFS28 FTIR spectrometer (Bruker Spectrospin Ltd., Coventry, United Kingdom) equipped with a mercury-cadmium-telluride (MCT) detector cooled with liquid N_2. The aluminum plate was then loaded onto the motorized stage of a reflectance thin-layer chromatography (TLC) accessory (15,16). The IBM-compatible personal computer used to control the IFS28 was also programmed (using OPUS version 2.1 software running under IBM O/S2 Warp provided by the manufacturers) to collect spectra over the wave number range 4000 to 600 cm^{-1}. The IR beam was focused into the sample, and data were collected automatically (because we knew the *x,y* location of the wells).

Spectra were acquired at a rate of 20 s^{-1}. The spectral resolution used was 4 cm^{-1}. To improve the signal-to-noise ratio, 256 spectra were co-added and averaged. Each sample was thus represented by a spectrum containing 882 points, and spectra were displayed in terms of absorbance as calculated from the reflectance-absorbance spectra using the OPUS software [which is based on the Kubelka-Munk theory (13)] (Fig. 1). These conditions were used for all experiments. To minimize problems arising from baseline shifts, the following procedure was implemented: (i) the spectra were first normalized so that the smallest absorbance was set to 0 and the highest to +1 for each spectrum; (ii) next, these normalized spectra were detrended by subtracting a linearly increasing baseline from 4000 to 600 cm^{-1}; (iii) finally, the smoothed first derivatives of these normalized and detrended spectra were calculated using the Savitzky-Golay algorithm (17) with five-point smoothing.

Cluster analysis. The initial stage involved the reduction of the multidimensional FTIR data by principal components analysis (PCA) (18). PCA is a well-known technique for reducing the dimensionality of multivariate data while preserving most of the variance, and Matlab was employed to perform PCA according to the NIPALS algorithm (19). Discriminant function analysis [PC-DFA, also known as canonical variates analysis (CVA)] then discriminated between groups on the basis of the retained principal components (PC) and the *a priori* knowledge of which spectra were replicates; thus, this process does not bias the analysis in any way (20). Finally, the Euclidean distance between *a priori* group centers in discriminant function (DF) space was used to construct a similarity measure, with the Gower general similarity coefficient S_G (21), and these distance

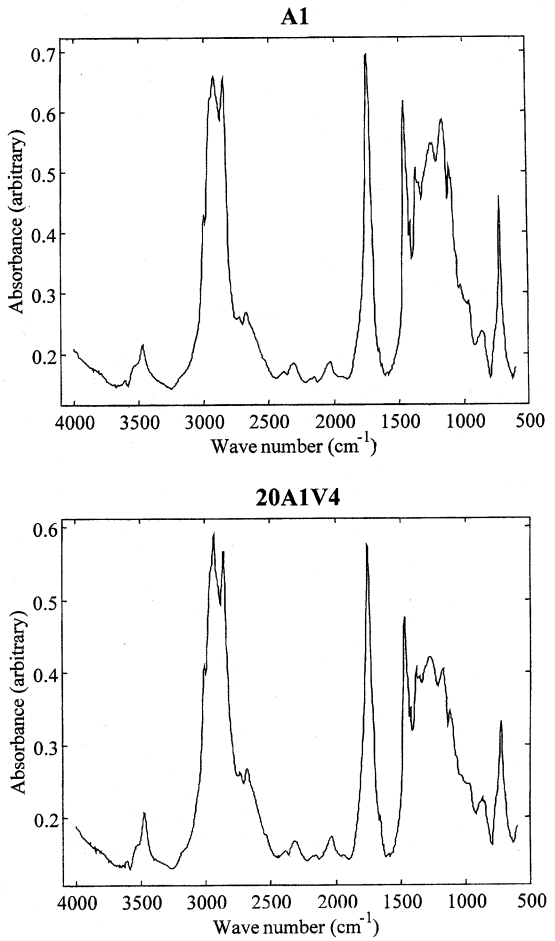

FIG. 1. Fourier transform infrared (FTIR) spectra of cocoa butter (CB) A1 (Malaysia) and of a mixture (20%) with cocoa butter equivalent (CBE) V4 (CBE mixture based on palm mid fraction with other fats in a ratio of 85:15).

measures were then processed by an agglomerative clustering algorithm to construct a dendrogram (20). These methods were implemented using Matlab version 5.0.0.4069 (The Math Works, Inc., Natick, MA), which runs under Microsoft Windows NT on an IBM-compatible personal computer.

Supervised learning methods. When the desired responses (targets) associated with each of the inputs (spectra) are known, then the system may be supervised. The goal of supervised learning is to find a model that will correctly associate the inputs with the targets; this is usually achieved by minimizing the error between the target and the model's response (output) (22). In this case we wanted to attempt to predict whether any CB was adulterated at the 10 or 20% level. Thus, a global single model was constructed.

The input data sets for all supervised learning methods contained the full FTIR spectra (882 wave number absorbances). Because we were interested only in adulteration at the 10 or 20% level, only these 369 spectra were used in the analysis, and these were partitioned into training and test sets.

It is important that the training data encompass the full range under study, because even though supervised methods are excellent at being able to interpolate, they are likely to give

poor estimates outside their "realm of knowledge," i.e., they cannot extrapolate sufficiently well (23). In order to achieve this range, the spectral data from FTIR were partitioned using the in-house program Multiplex (24), developed by Alun Jones. The Multiplex algorithm systematically placed samples into the training and test sets so that the problem domain (in terms of actual spectra) was adequately represented. This step is very important if we are attempting to achieve a global model in order to predict the level of adulteration blindly. The data were thus partitioned evenly into training and test sets; the training set comprised 186 spectra (62 in replicate), and the test set 183 spectra (61 in replicate). The output data were encoded in a single output node such that 10% adulteration was coded as 10 and 20% adulteration as 20.

Two artificial neural network (ANN)-based methods, *viz.*, standard back-propagation multilayer perceptrons (MLP) (25,26) and radial basis functions (RBF) (27,28) were used. Both ANN were carried out with a user-friendly neural network simulation program, NeuFrame version 3,0,0,0 (Neural Computer Sciences, Lulworth Business Centre, Totton, Southampton, United Kingdom), which runs under Microsoft Windows NT on an IBM-compatible personal computer (29).

The multivariate linear regression method of partial least squares (PLS) (30) was also exploited. All PLS analyses were carried out using an in-house program developed by Alun Jones (31) following the pseudocode given in Reference 30, which runs under Microsoft Windows NT on an IBM-compatible PC.

RESULTS AND DISCUSSION

The FTIR technique was used to analyze a wide variety of CB, CBE, and mixtures thereof. In total, 186 samples were analyzed in triplicate. The samples were coded in order to have no information about the composition before the analysis and statistical data evaluation were performed. The same sample sets had been used for other investigations of suitable analytical methods for the detection and determination of CBE in mixtures with CB (4,32).

Initial FTIR experiments were conducted to ascertain the optimal way to collect spectra. Several of the melted fats were applied directly to the sample carrier, but poor spectra were obtained; for example, the baseline was ~0.4 absorbance (arbitrary) units, and the highest peaks had an absorbance vastly in excess of 2.0. The conclusion was that too much sample had been applied. Therefore, to obey the Beer-Lambert law and thus limit highly absorbing bands (14), all samples were diluted 10-fold in analytical-grade acetone. These spectra typically ranged from 0 to 1.2 absorbance units (see Fig. 1). Finally, to minimize problems arising from unavoidable baseline shifts, the mathematical processing regime detailed above, and used previously (15), was adopted.

All CB and CBE had very characteristic FTIR spectra. Because it is often important to appreciate what is being measured in chemical terms, rather than to use FTIR purely as a blind fingerprinting tool, some studies were carried out using IR Mentor Pro, version 2 (Bio-Rad Laboratories, Richmond, CA), which allowed some of the major vibrational modes to be assigned. From this we could observe very strong C–H (and C=H) stretching at approximately 725, 1450, and in the region of 2800 to 3000 cm^{-1}, while the O–C=O ester bond vibration found in triglycerides was clearly evident at 1150 and 1750 cm^{-1} (Fig. 1).

In the first analytical run, 8 pure CB, 14 CBE (pure and commercial mixtures), and 32 CB/CBE mixtures (each CB was adulterated at 5, 10, 15, and 20% levels with the same CBE) were analyzed (Table 1). The adulteration range from 5 to 20% CBE at the fat level corresponds to about 1 to 5% CBE in final chocolate. These values, being below the permitted level of 5%, were prepared for analysis in order to demonstrate the analytical capability of the method.

A range of cluster analyses were undertaken on these samples, as detailed above; predominantly PC-discriminant function analysis (DFA) was employed, where the input to the DFA algorithm was the first 20 PC (which explained >99% of the total variance). The resulting PC-DFA plot (Fig. 2) shows that with the exception of V17, it was very easy to separate the pure CB from the pure CBE and pure vegetable fats.

In the PC-DFA cluster plots of five samples of only one pure CB (CB A1 from Malaysia, Tawau Crop 1992) adulterated with a CBE (CBE V4, a commercial mixture containing palm mid fraction and other exotic raw materials in a ratio of about 85:15) at levels of 0, 5, 10, 15, and 20%, a trend was observed according to the level of adulteration (Fig. 3) in the first DF. Because DF 1 was extracted to give the most variance, this trend clearly highlights that the information from the adulterant fat was present in the infrared spectra. The next stage was to ascertain if it would be possible to detect this pattern in a large variety of CB adulterated with many different CBE and with vegetable fats.

This stage was carried out with a second series of samples (Table 2) that were more extensive than the samples from the first run. Seven CB and nine CBE, both pure and commercial mixtures, were analyzed by FTIR. The aim of this work was to attempt to distinguish between 10 and 20% addition, corresponding to about 2 to 5% of CBE in chocolate, respectively.

As observed in the first run, it was very easy to distinguish the pure CB from the pure CBE. With one exception (CB K1), the dendrogram from these analyses (Fig. 4) showed that the CB grouped together and were clearly separated from the CBE. CB were observed to cluster into three groups: (i) BD (commercial mixture, deodorized) and BN (a commercial mixture identical to BD, nondeodorized) were highly similar; (ii) Z6 (a commercial mixture, deodorized) and Z7 (a commercial mixture identical to Z6, nondeodorized) cluster very closely and were also similar to K2 (pure CB from West Africa) and K3 (pure CB from Brazil); and (iii) K1 (pure CB from Malaysia) was different from the other CB. Because the nondeodorized and deodorized CB clustered together, these findings strongly suggest that the results obtained by infrared spectroscopy are based on the triglyceride structure rather than on minor components that are removed by the deodorization process.

For the CBE, four groups were seen: (i) V2 (a commercial CBE mixture based on palm midfraction and exotic raw ma-

TABLE 1
Samples of First Run of Analysis

Sample code[a]	% CBE in CB[b]	Kind of sample
A1	0	CB, Malaysia, Crop 1992
5A1V4	5	
10A1V4	10	
15A1V4	15	
20A1V4	20	
A5	0	CB, Nigeria, Crop 1993
5A5V6	5	
10A5V6	10	
15A5V6	15	
20A5V6	20	
B3	0	CB mixture, deodorized
5B3V2	5	
10B3V2	10	
15B3V2	15	
20B3V2	20	
B11	0	CB mixture, deodorized
5B11V12	5	
10B11V12	10	
15B11V12	15	
20B11V12	20	
G1	0	CB, West Africa, Ivory Coast
5G1V1	5	
10G1V1	10	
15G1V1	15	
20G1V1	20	
K1	0	CB, Malaysia
5K1V1	5	
10K1V1	10	
15K1V1	15	
20K1V1	20	
Z5	0	CB, Ghana, Crop 1994
5Z5V15	5	
10Z5V15	10	
15Z5V15	15	
20Z5V15	20	
Z4	0	CB, Ecuador, Crop 1994
5Z4V13	5	
15Z4V13	15	
20Z4V13	20	

[a]V1, CBE mixture, based on palm mid fraction; V2, CBE mixture, based on palm mid fraction; V4, CBE mixture, based on palm mid fraction; V6, CBE mixture, no information available; V12, palm mid fraction; V15, CBE mixture, no information available.
[b]CBE, cocoa butter equivalent; CB, cocoa butter.

terials in a ratio of about 70:30) and V5 (a commercial mixture, no information available) were similar; (ii) V13 (pure CBE, illipé from producer A) and V21 (pure CBE, illipé from producer B) were similar; (iii) V18 (a commercial mixture, based on shea) and V20 (a commercial mixture, based on sal) were very similar and clustered more loosely with V15 (a commercial mixture, no information available) and V19 (a commercial mixture, illipé type); and (iv) V22 (kokum) was very different from all the fats analyzed. This pattern was also observed in the DFA plot from analysis of all 139 samples (data not shown).

The next stage was to analyze each of the seven CB, both

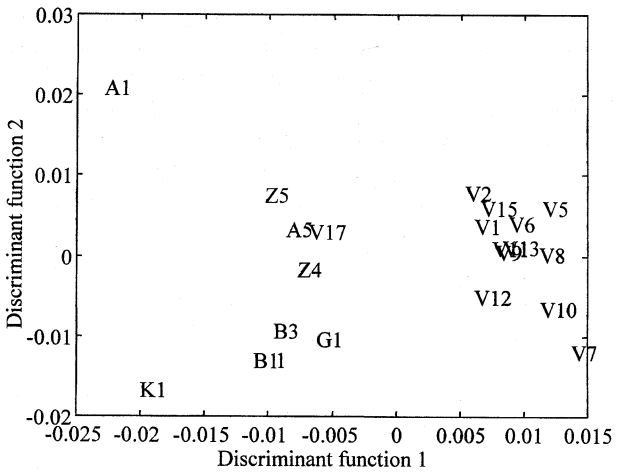

FIG. 2. Discriminant analysis (principal components-discriminant function analysis, PC-DFA) on pure CB and pure CBE. Abbreviations: A1, CB (Malaysia, Crop 1992); A5, CB (Nigeria, Crop 1995); B11, CB mixture, deodorized; B3, CB mixture, deodorized; G1, CB (West Africa, Ivory Coast); K1, CB (Malaysia); V1, CBE mixture, based on palm mid fraction; V2, CBE mixture, based on palm mid fraction; V5, CBE mixture, no information available; V6, CBE mixture, no information available; V7, CBE mixture, no information available; V8, illipé; V9, illipé; V10, palm mid fraction; V12, palm mid fraction; V13, illipé; V15, CBE mixture, no information available; V17, CBE mixture, no information available; Z4, CB (Ecuador, Crop 1994); Z5, CB (Ghana, Crop 1994). See Figure 1 for other abbreviations.

pure and adulterated, with the pure adulterants. Using CB BD as an example, the PC-DFA plots (Figs. 5A,B) showed that all the pure adulterants were very different from the pure BD and BD adulterated at 10 and 20% levels. Further analysis of BD alone and with 10 and 20% adulteration (Figs. 5C,D) showed that while each of the adulterants at the 10 and 20% levels could be separated from one another and from the pure BD, there was no distinction into two classes (one for the 10% and one for the 20% level) for all the CBE. This result is to

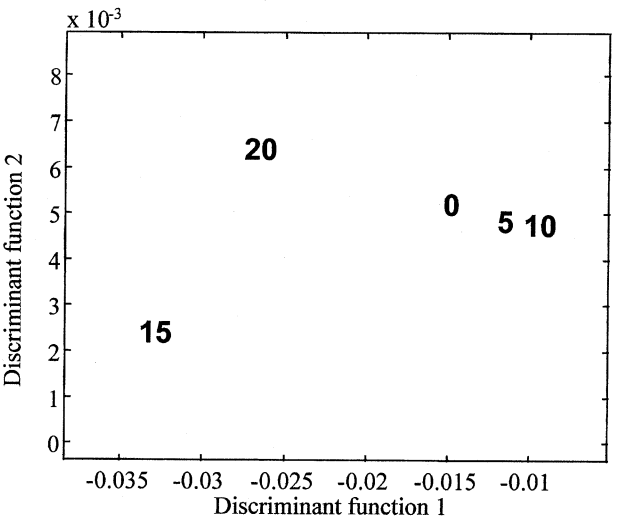

FIG. 3. Discriminant analysis plots of five samples of CB A1 (Malaysia) mixed with CBE V4 (a commercial mixture based on palm mid fraction with other fats in a ratio of 85:15) at levels of 5, 10, 15, and 20%. See Figure 1 for abbreviations.

TABLE 2
Samples of Second Run of Analysis

CB[a]	CBE[b]	%	CB	CBE	%	CB	CBE	%
BD		0	K1	V22	10	K3	V15	20
BD	V18	10	K1	V22	20	Z6		0
BD	V18	20	K1	V2	10	Z6	V18	10
BD	V19	10	K1	V2	20	Z6	V18	20
BD	V19	20	K1	V5	10	Z6	V19	10
BD	V20	10	K1	V5	20	Z6	V19	20
BD	V20	20	K1	V13	10	Z6	V20	10
BD	V21	10	K1	V13	20	Z6	V20	20
BD	V21	20	K1	V15	10	Z6	V21	10
BD	V22	10	K1	V15	20	Z6	V21	20
BD	V22	20	K2		0	Z6	V22	10
BD	V2	10	K2	V18	10	Z6	V22	20
BD	V2	20	K2	V18	20	Z6	V2	10
BD	V5	10	K2	V19	10	Z6	V2	20
BD	V5	20	K2	V19	20	Z6	V5	10
BD	V13	10	K2	V20	10	Z6	V5	20
BD	V13	20	K2	V20	20	Z6	V13	10
BD	V15	10	K2	V21	10	Z6	V13	20
BD	V15	20	K2	V21	20	Z6	V15	10
BN		0	K2	V22	10	Z6	V15	20
BN	V18	10	K2	V22	20	Z7	V13	10
BN	V18	20	K2	V2	10	Z7	V13	20
BN	V19	10	K2	V2	20	Z7	V18	10
BN	V19	20	K2	V5	10	Z7	V18	20
BN	V20	10	K2	V5	20	Z7	V19	10
BN	V20	20	K2	V13	10	Z7	V19	20
BN	V21	10	K2	V13	200	Z7	V20	10
BN	V21	20	K2	V15	10	Z7	V20	2
BN	V22	10	K2	V15	20	Z7	V21	10
BN	V22	20	K3		0	Z7	V21	20
BN	V2	10	K3	V18	10	Z7	V22	10
BN	V2	20	K3	V18	20	Z7	V22	20
BN	V5	10	K3	V19	10	Z7	V2	10
BN	V5	20	K3	V19	20	Z7	V2	20
BN	V13	10	K3	V20	10	Z7	V5	10
BN	V13	20	K3	V20	20	Z7	V5	20
BN	V15	10	K3	V21	10	Z7	V13	10
BN	V15	20	K3	V21	20		V18	100
K1		0	K3	V22	10		V19	100
K1	V18	10	K3	V22	20		V20	100
K1	V18	20	K3	V2	10		V21	100
K1	V19	10	K3	V2	20		V22	100
K1	V19	20	K3	V5	10		V2	100
K1	V20	10	K3	V5	20		V5	100
K1	V20	20	K3	V13	10		V13	100
K1	V21	10	K3	V13	20		V15	100
K1	V21	20	K3	V15	10			

[a]CB: BD, mixture deodorized; BN, mixture nondeodorized; K1, Malaysia; K2, West Africa; K3, pure CB, Brazil; Z6, commercial deodorized mixture; Z7, commercial mixture identical to Z6 but not deodorized.
[b]CBE: V2, mixture of palm mid fraction with other fat in ratio 70:30; V5, mixture with composition unknown; V13, illipé; V15, mixture with composition unknown; V18, CBE mixture based on shea; V19, CBE mixture based on illipé; V20, CBE mixture based on sal; V21, CBE illipé; V22, CBE kokum. See Table 1 for other abbreviations.

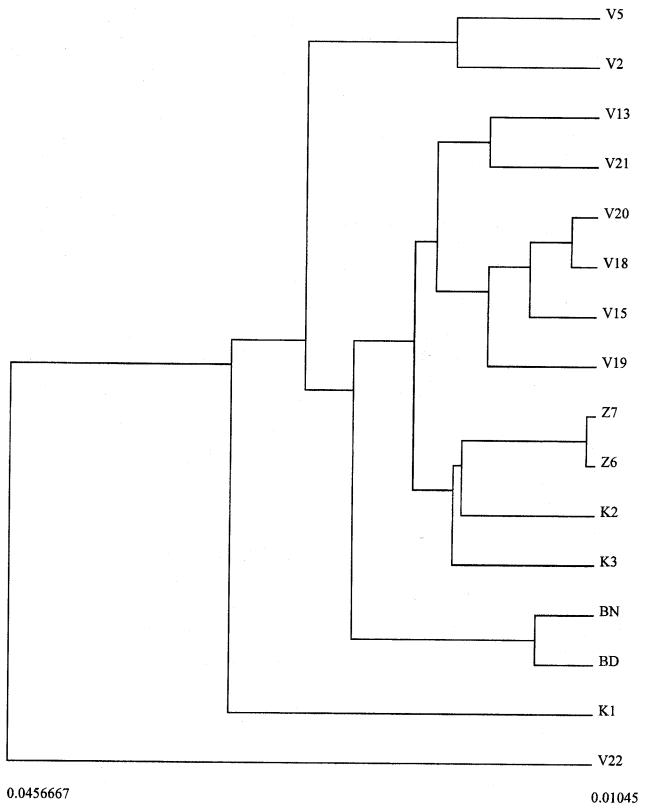

FIG. 4. Dendrogram of various pure CB and CBE. Abbreviations: V18, CBE mixture based on shea; V19, CBE mixture based on illipé; V20, CBE based on sal; V21, CBE illipé; V22, CBE kokum; Z6, CB commercial deodorized mixture; Z7, CB commercial mixture identical to Z6 but not deodorized; K2, cocoa butter (West Africa); K3, cocoa butter (Brazil); BN, cocoa butter mixture, nondeodorized; BD, cocoa butter mixture, deodorized. For other abbreviations, see Figures 1 and 2.

be expected, because the CBE are chemically different and the addition of them to a single CB will cause different spectral changes. To attempt to force the separation into two groups, DFA was performed again, but this time the group structure (*a priori* information) was based on whether the butters were adulterated at either 10 or 20%. This DFA plot showed that although there was some overlap, two groups could be seen; moreover, the 20% adulterants had more spread than the 10% samples. This result is to be expected given their increased chemical differences. Finally, to test whether DFA could be used to predict whether CB were adulterated at either the 10 or the 20% level, the first 30 spectra were used to form PC-DFA, and then the remaining 24 spectra were projected into these spaces. However, this approach proved to be unsuccessful. It is possible that the spectra used to calibrate the DFA were different from those used to test the separation, so the projection was extrapolating. All these analyses were conducted for the other six CB, and similar results were observed (data not shown).

The next stage was to analyze each of the nine CBE separately, mixed with all the cocoa butters at 10 and 20% levels. Using CBE V21 (illipé) as an example, DFA plots (data not shown) demonstrated that most of the different CB were recovered together, but again that those at the 10 and 20% levels could be separated. Further DFA analysis encoding group structure according to adulterant level (Fig. 6) indicated that, although there was some overlap, two groups could be seen, one for the 10% and one for the 20% in the first DF. Projection analyses were again conducted, and these were also un-