Study of the Self-Association of Alcohols by Near-Infrared Spectroscopy and Multivariate 2D Techniques

Laila Stordrange,[†] Alfred A. Christy,[‡] Olav M. Kvalheim,*[,†] Hailin Shen,[§] and Yi-zeng Liang[||]

Department of Chemistry, University of Bergen, N-5007 Bergen, Norway, Department of Chemistry, Faculty of Mathematics and Natural Sciences, Agder University College, Kristiansand, Norway, School of Chemistry, University of Bristol, Bristol, U.K., and College of Chemistry and Chemical Engineering, Central South University, Changsha 410083, People's Republic of China

Received: October 1, 2001; In Final Form: June 17, 2002

Self-association of 1-propanol, 1-butanol, 1-pentanol, 1-hexanol, 1-heptanol, 2-propanol, 2-methyl-1-propanol, and 2-methyl-2-propanol in carbon tetrachloride (concentration range 0.01−1.00 M) has for the first time been investigated by means of multivariate curve resolution in the near-infrared region of 1900−2200 nm. Rank analysis was carried out prior to resolution to determine the number of distinct species. In addition to principal component analysis (PCA) and evolving factor analysis (EFA), smooth PCA was used with good results. For all of the investigated alcohols, the rank was unambiguously assessed as three over the entire concentration range. At low concentrations, EFA detected a concentration region with only two species. Eigenstructure tracking analysis (ETA) revealed one-component regions in the spectral direction. The first loadings of the selective regions gave the concentration profiles of the two species. The spectral profiles of these two species were then resolved by using the least squares of these concentration profiles. The third profile was resolved by iterative target transformation factor analysis (ITTFA). The quality of the resolved profiles was checked by using the method of orthogonal projections (OP). The resolved profiles were subsequently used to determine the association numbers and equilibrium constants for the alcohols. The resolved spectral and concentration profiles of the linear primary alcohols and the branched alcohols separated into two distinct groups. The average sizes of the associated species increase with increasing concentration. At higher concentrations of alcohol, the cyclic/nonpolar species exist in higher populations than do the linear species. A model that interprets the systems in terms of equilibria between a monomer, linear aggregates, and cyclic aggregates shows good agreement with the resolved profiles.

Introduction

Hydrogen bonding[1] arises from the polarization of the covalent bond between a hydrogen atom and a more polar atom. The positive charge on the hydrogen atoms induce dipole−dipole interaction between molecules. Hydrogen bonding in water leads to some anomalous properties. For example, water has high melting and boiling points compared to those of chemical substances with similar molecular weight. The molecules of water tend to associate in many different forms, giving rise to different aggregates of molecules of different molecular weight. Similar hydrogen bonding also arises in alcohols, where one hydrogen atom in the water molecule is replaced by an alkyl group. Although the extent and probability of hydrogen bonding is reduced by half because of the presence of an alkyl group, the hydrogen bonding in alcohols gives rise to high boiling points for alcohols compared to those of hydrocarbons of similar molecular weight.[2]

The self-association of alcohols that has been discussed in the literature for more than a half century is based on techniques such as dielectric[3,4] and infrared spectroscopy both in the mid-IR[5−7] and near-IR[4,8−12] regions.

Dielectric studies of alcohols in nonpolar solvents[3,4] have proven that high polar aggregates exist at low concentrations near 0.2 M whereas some low-polar aggregates exist at higher concentrations. This indicates the presence of cyclic aggregates in alcohol solutions at higher concentrations.

The strength of a covalent O−H bond is assumed to decrease in the proximity of a hydrogen bond. We know from Hooke's law that the frequency is proportional to the force constant. OH stretches of aggregates with high association numbers are therefore expected to be observed at frequencies that are lower than those of small aggregates. Generic structures are given in Figure 1. The frequencies due to the free-OH group of monomer (a), the terminal, free-OH group of linear aggregates (b), and the hydrogen-bonded OH group of linear aggregates (c and d) and cyclic aggregates (e) are expected to follow the order

$$\nu_{\text{monomer,free}} > \nu_{\text{linear,free}} > \nu_{\text{linear,H-bonded}} > \nu_{\text{cyclic,H-bonded}}$$

Using infrared spectroscopy, Van Ness et al.[5] indicated a three-component system of monomer, cyclic dimer, and linear polymer for solutions of ethanol in *n*-heptane and in toluene. Fletcher and Heller,[8] using the near-infrared absorption of the first overtone of the OH absorptions, predicted a monomer, linear tetramer, and cyclic tetramer model for solutions of 1-octanol and 1-butanol in *n*-decane. These workers claimed first that the dimers of alcohol molecules did not exist. Later, they concluded that the dimers exist in low concentrations but

* Corresponding author. E-mail: Olav.Kvalheim@kj.uib.no. Tel: + 47 55 58 33 66. Fax: + 47 55 58 94 90.
[†] University of Bergen.
[‡] Agder University College.
[§] University of Bristol.
[||] Central South University.

10.1021/jp013670f CCC: $22.00
Published on Web 08/16/2002

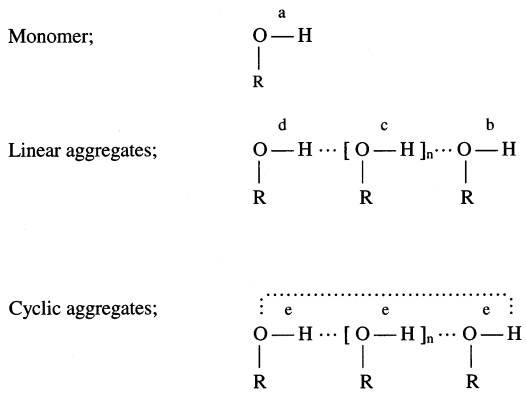

Figure 1. Generic structures of alcohols in inert solutions; monomer, linear aggregates, and cyclic aggregates.

in open-chain form.[9] Brink and Glasser[3] used dielectric spectroscopic techniques in the study of ethanol in different nonpolar solvents and concluded that the ethanol molecules exist in monomer, open dimer, and cyclic tetramer species. Shinomiya and Shinomiya[4] used near-IR and dielectric spectroscopic techniques in the study of 1- and 3-pentanol in *n*-heptane at high concentrations and proposed a model of a monomer, open-chain dimer, cyclic tetramer, and open-chain octamer for the association of alcohols. Recent studies by Czarnecki[11] using 2D FT-NIR correlation spectroscopy of 1-octanol at different temperatures and concentrations in CCl_4 show that in solutions at low or moderate concentrations the system comprises mainly cyclic aggregates with relatively low populations of open species.

All of these studies suggest that the alcohols exist in a "free" unbound form at very low concentrations and that they tend to associate and form larger aggregates of different sizes and shapes at high concentrations. However, controversy about the sizes and shapes of the aggregates exists.

In 1936, Errera and Mollet[13] used infrared spectroscopy to study the self-association of alcohols. Two peaks were observed, one narrow band at 3650 cm^{-1} and a broad band in the region 3590−3100 cm^{-1}. The narrow band was assigned to free alcohol molecules. Concerning the broad band, Errera and Mollet wrote, "It seems possible that the large band at 3350 cm^{-1} can be decomposed into two bands corresponding to different molecular aggregations."

Curve-resolution techniques make it possible to resolve the 2D data into spectra and concentration profiles of the significant components in the system. Førland et al.[7] and Nodland[6] used infrared spectroscopy in combination with curve resolution to study the self-association of alcohols in the low-concentration region. They found three different species that they assigned to monomers, linear aggregates, and cyclic aggregates (see Figure 1). The resolved profiles were used for qualitative and quantitative interpretations (i.e., band assignments and estimation of aggregate sizes).

The advantage of NIR spectroscopy is that the molar absorptivities of overtone and combination bands are much lower than the absorptivities of fundamentals in the mid-infrared region. Thus, the spectra of high-concentration or neat samples can be obtained. NIR spectroscopy is a great tool to use when studying the self-association of alcohols in high-concentration samples.[11,12,14] FT-NIR in combination with 2D correlation techniques has been used to study the self-association of alcohols.[11,14] The technique is very useful for band assignments and qualitative studies but provides no quantitative information. Iwahashi et al.[12] proposed a method to quantify the association degree of alcohols in the near-infrared region. They used one frequency under the monomer peak, which is partially superimposed by bands arising from hydrogen-bonded species, to determine the association degree of the polymer. Using this method, they were able to quantify only two species: the monomer and the polymer. In earlier studies, there is a general opinion that at least three species exist.[3,4,6,7] It is of great interest to be able to quantify the number of species and the size of these aggregates in order to understand the association of alcohol molecules. Being able to resolve data into spectral and concentration profiles will increase the understanding of the self-association phenomena.

One goal of this study is to find a way to resolve the NIR spectra of alcohols into pure spectral profiles of different associated species and thereby increase our insight into the self-association phenomenon. The study comprises several straight-chain and branched-chain alcohols in varying concentrations in carbon tetrachloride; it is also our intention to examine the difference in the behavior of these alcohols with increasing concentration using the concentration profiles obtained from the curve resolution of the NIR spectra.

Theory

Dynamic Process. Alcohol molecules associate and form aggregates with increasing alcohol concentration. This represents a dynamic process that is reflected in the NIR spectra if the change in total alcohol concentration is sufficient.

If we assume that the dynamic process evolves because of a change in the concentrations of the components in the mixture and that the spectra of the different components obey additivity, then each absorbance measured at a particular wavelength, λ, can be expressed as the sum of the components, A, of the concentration, c_a, and molar absorptivity, s_a, of the different components in the solution, respectively:

$$x(k, \lambda) = \sum_{a=1}^{A} c_a(k)\, s_a(\lambda) \tag{1}$$

In eq 1, $x(k, \lambda)$ is the absorbance of light by a solution with concentration k measured at a particular wavelength λ. If the absorbance of many solutions is measured at several wavelengths, then the spectra can be organized as follows:

$$\mathbf{X} = \mathbf{C}\mathbf{S}^{\mathrm{T}} \tag{2}$$

In eq 2, $\mathbf{X}$ is a matrix containing N rows (absorbance spectra at N different concentrations) and M columns (absorbances at M wavelengths). $\mathbf{C}$ is the concentration matrix of dimension $N \times A$ containing concentration profiles of A different aggregates in solution. $\mathbf{S}$ is the molar absorptivity (pure spectra of the A species in the solution) matrix of order $M \times A$ containing A columns of molar absorption coefficients at M wavelengths.

The decomposition of $\mathbf{X}$ into $\mathbf{C}$ and $\mathbf{S}$ can be carried out using curve-resolution techniques such as heuristic evolving latent projections (HELP),[15] iterative target factor analysis (ITTFA),[16] and orthogonal projection (OP).[17]

Equilibrium Constants. When the concentration of an aggregate is known, the average size of the aggregated species can be determined. If the aggregate contains n monomers (i.e., n is the association number), then the equilibrium of a monomer−aggregate model can be written as shown in eq 3.

$$n\mathrm{R{-}OH} \leftrightarrow (\mathrm{R{-}OH})_n \tag{3}$$

Earlier studies reveal that alcohol solutions can be expected to

be a three-component system consisting of monomer, linear (polar) aggregates, and cyclic (nonpolar) aggregates.[3,4,6,7,11] A relationship similar to that given in eq 4 can be expected.

$$\begin{array}{ccc} & \overset{K_1}{} & \\ \text{nR-OH} & \leftrightarrow & (\text{R-OH})_{n^*\_\text{linear}} \\ {\scriptstyle K_3}\updownarrow & & \updownarrow{\scriptstyle K_2} \\ & (\text{R-OH})_{n'\_\text{cyclic}} & \end{array} \tag{4}$$

n monomers form linear and cyclic aggregates of average sizes equal to n^* and n', respectively. Linear aggregates can be expected to form cyclic aggregate and vice versa. In addition, the aggregates can break up and form monomers. Equation 4 gives three equilibrium equations: K_1, K_2, and K_3. This makes it difficult to estimate the aggregate sizes and the equilibrium constant of the alcohol systems. Simplification of a three-component system to a two-component system by the assumption of a monomer−aggregate model (eq 3) makes it possible to calculate the association number, *n*, and the equilibrium constant, *K*.

$$K = [(\text{R}-\text{OH})_n]/[\text{R}-\text{OH}]^n \tag{5}$$

Equation 5 is useful because it permits the evaluation of a system where a monomer is in equilibrium with other species. The association numbers calculated here are a result of this rough estimate, and *K* is the relationship of the concentration of aggregated and monomer species.

To be able to find estimates of the association numbers and the equilibrium constants, the absolute concentration profiles have to be calculated. For a system where the total concentration is known, absolute concentration profiles can be calculated using the assumption that the sum of the concentrations of different species is equal to the total concentration.

$$\mathbf{c}_{\text{total}} = \sum_{a=1}^{A} \mathbf{c}_a \text{b}_a = \mathbf{C}\mathbf{b} \tag{6}$$

The vector **b** used for scaling the resolved concentration profiles is calculated by least squares:

$$\mathbf{b} = (\mathbf{C}^{\text{T}}\mathbf{C})^{-1}\mathbf{C}^{\text{T}}\mathbf{c}_{\text{total}} \tag{7}$$

The association number is determined by the same method as that used by Nodland[6] and Førland et al.[7] The scaled concentration profiles of associated species is related to free species by eq 8:

$$(\mathbf{c}_{\text{associated}})^{1/n} = k \times \mathbf{c}_{\text{free}} \tag{8}$$

A plot of $(\mathbf{c}_{\text{associated}})^{1/n}$ versus $\mathbf{c}_{\text{free}}$ with the correct association number, *n*, will give a straight line passing through the origin. When the absolute concentration profiles and the association numbers are known, the equilibrium constants are calculated by inserting these values into eq 5.

Curve Resolution. For every alcohol, a matrix **X** containing *N* spectra (objects) of different concentrations and spectra of *M* wavelengths (variables) was created.

A NIR spectrophotometer produces spectra with many wavelengths where the same information can be found in several of the wavelengths. It is possible to find a small number of latent variables that explain the maximum amount of information in the data set.

$$\mathbf{X} = \sum_{a=1}^{A} \mathbf{t}_a \mathbf{p}_a^{\text{T}} + \mathbf{E} = \mathbf{T}\mathbf{P}^{\text{T}} + \mathbf{E} \tag{9}$$

The latent variables, **T** and **P**, are in this case called principal components (PCs), and the method is called principal component analysis (PCA).[18] Decomposition onto latent variables to explain as much of the variance in the data as possible is also called singular value decomposition (SVD).[19] The chemical rank *A* is equal to the number of components that gives rise to significant variation in the data. **T** is the latent variable in the concentration direction, and $\mathbf{P}^{\text{T}}$ is the latent variable in the wavelength direction. They are usually referred to as scores and loadings of the matrix **X**. The dimensions of **T** and $\mathbf{P}^{\text{T}}$ are $N \times A$ and $A \times M$, respectively. The matrix **E** in eq 9 is noise. There exist many rank analysis methods with which to decide the number of significant components in a system. Familiar methods are PCA[18] and the evolving factor analysis (EFA).[20,21] A newer method is smooth PCA.[22] Smooth PCA is an improvement of PCA for rank determination. In PCA, the eigenvalues are found by solving eq 10.

$$\mathbf{X}^{\text{T}}\mathbf{X}\boldsymbol{\nu} = \lambda\boldsymbol{\nu} \tag{10}$$

X is the matrix, $\boldsymbol{\nu}$ is the eigenvector, and λ is the eigenvalue. In smooth PCA, this equation is expanded with a smoothing factor $(\mathbf{I} + k\,\mathbf{G}^{\text{T}}\mathbf{G})^{+}$.

$$(\mathbf{I} + k\,\mathbf{G}^{\text{T}}\mathbf{G})^{+}\,\mathbf{X}^{\text{T}}\mathbf{X}\boldsymbol{\nu} = \lambda\boldsymbol{\nu} \tag{11}$$

I is the identity matrix, **G** is a second-order differentiation matrix, and *k* is the degree of smoothing. To determine the number of significant eigenvalues in a principal component analysis of a multicomponent system, eqs 10 and 11 can be resolved as general eigenvalue problems. The logarithm of the ratio of the eigenvalues from eqs 10 and 11 shows a significant increase at the $A + 1$ eigenvalue when the true rank is equal to *A*. Although the eigenvalues of the significant components remain the same after smoothing, the noise residuals will decrease, and it is this change that is reflected in the plot of the ratio of the eigenvalues.

Multivariate Curve Resolution of 2D Data. A system where no a priori information about the chemical components is available is classified as a black system.[23] The aim of performing curve resolution on black systems is to estimate the spectra (qualitative analysis) and concentration profiles (quantitative analysis) of all the chemical components present in the system. This can be written as a decomposition of a data matrix **X** to a concentration matrix **C** and a spectral matrix **S** (eq 2). Selective regions are used for the unique resolution of multicomponent systems into pure spectra and concentration profiles.[15] Selective regions can be detected visually by latent projection graph (LPG)[15] and eigenstructure tracking analysis (ETA)[24] methods. When the overlap between spectral bands is severe, LPG may not be sensitive enough. A possible solution is to use eigenstructure tracking analysis (ETA).[24] If selective concentration regions can be found for all of the significant components, then it is possible to calculate the concentration profiles directly by using eq 2.

$$\mathbf{C} = \mathbf{X}\mathbf{S}_{\text{s}}(\mathbf{S}^{\text{T}}_{\text{s}}\mathbf{S}_{\text{s}})^{-1} \tag{12}$$

The symbol *s* stands for spectra from selective regions. Here, the estimated spectra for every chemical component is set equal to the first loading vector calculated for every independent selective concentration profile. For a system of *A* components

TABLE 1: Manufacturer and Quality of Alcohols Used in This Work

alcohol	manufacturer	quality
1-propanol	Riedel de Haen	minimum 99.5%, p.a.
1-butanol	Riedel de Haen	minimum 99.5%, p.a.
1-pentanol	Fluka Chemika	minimum 99%, p.a.
1-hexanol	Riedel de Haen	minimum 98%, p.a.
2-propanol	Vinmonopolet (Norway)	technical quality
2-methyl-1-propanol	Merck A/S	minimum 99%, p.a
2-methyl-2-propanol	Merck A/S	minimum 99.5%, p.a
carbon tetrachloride	Merck A/S	minimum 99.8%, p.a

where selective regions are decided for all of the components, PCA performed on each selective region provides the selective spectrum as the first loading vector (**p**). The spectra are then $\mathbf{S}_s = [\mathbf{p}_1\mathbf{p}_2...\mathbf{p}_A]$. On the contrary, selective regions in the spectral direction identify the column vectors that are proportional to the concentration profiles of the components. The best estimate of a concentration profile for a component is found as the score vector of the first principal component. The concentration profiles are then $\mathbf{C}_s = [\mathbf{t}_1\mathbf{t}_2...\mathbf{t}_A]$. The pure component spectra are subsequently obtained by eq 13.[15]

$$\mathbf{S}^T = (\mathbf{C}_s{}^T\mathbf{C}_s)^{-1}\mathbf{C}_s{}^T\mathbf{X} \tag{13}$$

Frequently, not all the components have selective regions. Iterative target transformation analysis (ITTFA)[16] is a method where selective regions are not essential. Orthogonal projection (OP) was introduced in quantitative analysis by Lorber.[17] He showed that the net analytical signal for a component is equal to the part of the spectrum that is orthogonal to the spectra of the other components. A profile can be calculated using ITTFA and thereafter checked by means of OP.

Experimental Section

Eight alcohols were studied in this work (Table 1). The near-infrared measurements were made using a Perstorp Analytical 6500 NIR instrument equipped with a fiber optic probe. A known amount of carbon tetrachloride was poured into a glass cell assembly. A reference spectrum of carbon tetrachloride was measured in the range 1100−2500 nm. Thirty-two scans at a resolution of 2 nm were averaged in each measurement. The alcohol to be investigated was injected using a 100-μL syringe. The solution was then stirred to homogenize the mixture. Measurements (30−42) were made in the concentration range 0.01−1.00 M. Carbon tetrachloride is cancer-causing and should be handled with care in good ventilation. Data analysis was performed using Matlab (The Math Works Corp.) and Xtricator (Pattern Recognition Systems AS).

Results and Discussion

Wavelength Region for Investigation of the Self-Association of Alcohols. Figure 2 shows NIR spectra of 1-pentanol (0.1 and 0.5 M) and *n*-pentane in CCl_4. The spectra of 1-pentanol show overlap at the first and second overtones of the CH stretch and in the combination region of the CC and CH stretches. To be able to resolve the profiles of a multicomponent system, there must be a significant change in the spectra when the concentration of the alcohol changes. A change is observed in the first overtone of the OH stretch (1300−1650 nm) and in the combination region of the OH stretch (1900−2200 nm). These two regions are of interest in the study of the self-association of alcohols.

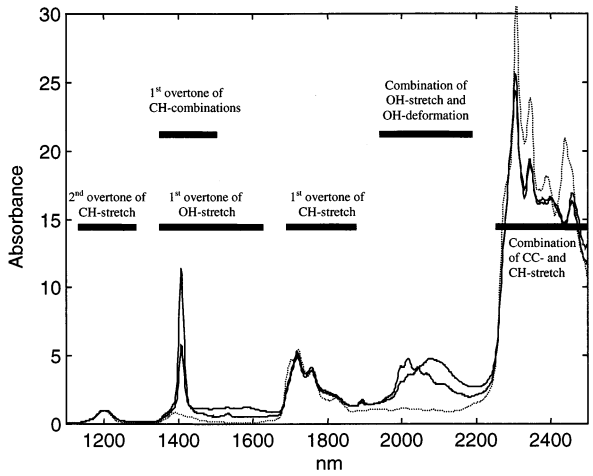

Figure 2. Spectra of 0.1 and 0.5 M 1-pentanol (—) and *n*-pentane (· · ·). All of the spectra are normalized after the band at the second overtone of the CH stretch.

One problem in NIR spectroscopy is that CH stretch falls within first overtone of the OH stretch region. In the first overtone of the OH stretch, a broad absorption band from about 1350−1550 nm is observed (Figure 2). Most of the curve resolution methods are dependent on selective regions, and the signal from the CH stretch must be removed before continuing. One possibility is to use subtraction methods such as direct subtraction[25] and generalized background subtraction (GBS)[26] by using the *n*-pentane spectrum. These methods were tested with no success. Instead, curve resolution of the combination region (1900−2200 nm), which is not as strongly influenced by the background for CH vibrations, was performed. A detailed description of the resolution of the OH combination region follows.

All of the alcohols tested here behave in the same manner. Because of this, the alcohols were investigated using the same procedure. Curve resolution is therefore shown only for one alcohol, 1-heptanol. The resolved spectra and equilibrium constants for the other alcohols will be presented.

Baseline Correction. The spectrum of *n*-pentane in the combination region of the OH stretch from about 1900−2200 nm is not equal to zero (Figure 2). A baseline is assumed to be present. To be able to resolve the spectra, this baseline has to be removed. A simplification of the baseline correction proposed by Liang et al.[27] was performed. One "zero-concentration" region before and one after the OH combination band are selected. Thereafter, a straight line is fitted through the points in the zero-concentration regions. This gives an estimate of the baseline in the region between the two zero-concentration regions. The baseline is subtracted from the raw spectrum. The baseline correction was performed individually for every spectrum in a series. In Figure 3, spectra before (a) and after (b) baseline correction and their respective LPG plots are shown. The LPG plot (Figure 3a) shows an offset and drift in the raw spectrum because the line does not start and stop in the origin. The spectra were corrected, and after baseline correction, the LPG plot (Figure 3b) indicates that the baseline has been removed.

Rank Analysis. The baseline-corrected spectra (Figure 4a) were normalized by dividing the spectra by their respective concentrations for the exploration of data (Figure 4b). With increasing concentration, a relative decrease in the intensity in the range 1912−2050 nm is revealed after normalization. Two peaks at 1996 and 2014 nm are observed in this wavelength region. In the region from 2050 to 2180 nm, a relative increase

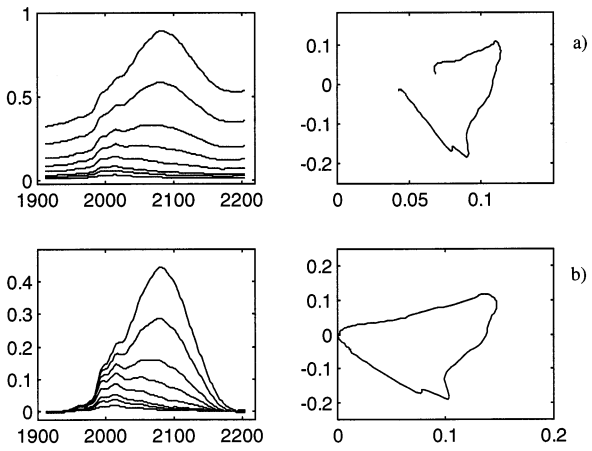

Figure 3. Spectra and LPG plot of (a) raw data and (b) baseline-corrected data of 1-heptanol.

in the intensity with increasing alcohol concentration is revealed. This region contains a broad band with maximum absorbance at 2082 nm.

Baseline-corrected raw data (Figure 4a) has been used in the following analysis of the alcohols.

Rank determination of a system is very important before curve resolution is performed because the solution is being approached iteratively and by induction. First, a principal component decomposition was performed. In Table 2, the eigenvalues and explained variance is listed for the first 10 principal components. One PC explains more than 97% of all the variance in the data, and this indicates a one-component system. Figure 4b shows that the system consists of at least two components. Two components explain 99.99% of variance in data. Components four and five explain almost the same amount of variance. Between the fifth and sixth principal components, a new jump that is observed before the remaining components explains the same amount of variance. Principal decomposition of raw data indicates a system consisting of two, three, or five components. Investigating scores and loadings did not give unique rank determination. Scores gave two−three components whereas loadings gave five components. Correlated noise is often present in spectra. Correlated noise can be removed by differentiation in the spectral direction.[28] In Figure 5, the first six loadings are shown after performing a first-order differentiation in the spectral direction. The loading vectors in the spectral direction show structure in the three first loadings. The remaining loadings have a noisy pattern. This strongly indicates a three-component system.

An EFA plot in the concentration direction before and after baseline correction is given in Figure 6. Before baseline correction, the rank is equal to four. After baseline removal, the rank is three or five. The fourth and fifth components are noisy and close to each other. EFA suggests a rank of three after baseline correction.

In smooth PCA, the ratio between eigenvalues before and after smoothing is calculated. This gives an informative plot for rank determination. By using smooth PCA, the rank was finally determined (Figure 7). The rank is equal to the first values that are close to zero (log 1 = 0). The line (−∗−) indicates different degrees of smoothing that are forced onto the system by choosing different k values (see eq 11). In Figure 7, a clear increase is observed from the third to the fourth eigenvalue in both the concentration and in the spectral direction, revealing a system of chemical rank equal to three.

Investigating loadings in the spectral direction, EFA in the concentration direction, and smooth PCA in both the spectral and concentration directions for all the alcohol solutions (0.01−1.00 M) revealed three significant components.

Shen et al.[29] tested a rank-analysis method called morphological scores on four of the alcohol systems inspected here. He showed that the rank was three for 1-propanol, 1-butanol, 1-heptanol, and 2-butanol.

Curve Resolution

Resolution of a Two-Component Region. Upon inspecting the EFA plot (Figure 6b), a two-component region seemed to be present at low concentrations. A two-component region from 0.006 to 0.113 M was established (see Figure 8). A system with two components is usually easy to resolve. If selective regions exist for both components, the least-squares method can be used to resolve the profiles. If no selective regions exist and the two components are completely overlapped, it may still be possible to resolve the profiles by using sequential rank analysis.[30]

To obtain a good resolution of spectra, one has to find selective regions. The ETA plot is a good method for revealing selectivity. By means of the ETA plot, six selective regions were found in the spectral direction. When cross-correlating the first scores of the selective regions, two groups were revealed. These were determined to belong to components 1 and 2, respectively. The first score of a selective region is equal to the concentration profile of that species. The spectra were calculated using the least-squares method, $\mathbf{S}^T = (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{X}$. The profiles are shown in Figure 9. Monomers dominate at low concentrations. When the concentration increases, monomers associate and form aggregates. In Figure 9, it is assumed that the dotted line belongs to the monomer. At higher concentrations, the alcohols are expected to associate into linear aggregates. The spectral profile of the second component is asymmetric. This may be due to the fact that at higher frequencies (2050 nm) many smaller linear aggregates exist, whereas lower intensity at lower frequencies (2100 nm) is caused by the presence of fewer but longer linear aggregates. This is the same observation that was made by Førland et al.[7] Component 2 is expected to correspond to linear aggregates of different lengths (see Figure 1).

The monomer band has a negative part. This may be the result of the selective regions not being completely pure (i.e., contaminated by absorbance from other components). In such cases, negative regions might be expected when using least squares to resolve the profiles. Another reason might be that the baseline correction is too rough.

Resolution of the Third Component. Cyclic aggregates are expected to exist at a lower frequency than linear aggregates. This information can be used for an independent resolution to decide the profiles of the third component in the system.

The matrix was expanded to include the whole concentration range from about 0.01−1.0 M. The spectral region is the same as it was for the resolution of the two-component system.

The ITTFA method was used to find the profile of the third component. The third component is expected to show up at a lower frequency than the second component if the second and the third components are linear and cyclic aggregates, respectively. Different start estimates were chosen at a frequency slightly lower than the band maximum of the second component. The iteration was performed 15 times. For every iteration, regions of negative intensity were set to zero. The different start estimates gave resolved spectra of the same shape and almost the same maximum. For 1-heptanol, the start estimates were chosen as 2084, 2100, and 2110 nm. After iteration, peaks with maxima equal to 2086, 2090, and 2088 nm were obtained, respectively. The final start estimate was chosen as 2088 nm.

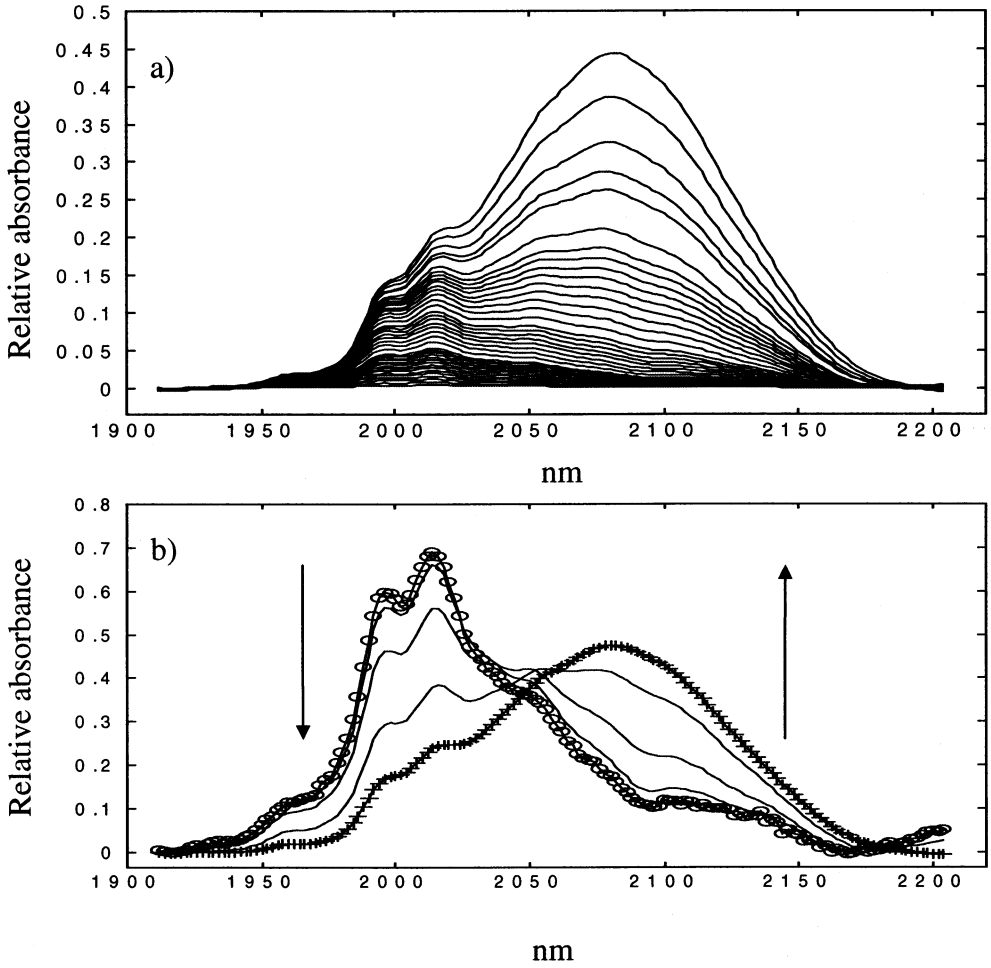

Figure 4. (a) Baseline-corrected raw data of 1-heptanol (1912−2204 nm) and (b) normalized spectra of lowest concentration 0.01 M (−○−○−) 0.06 M, 0.15 M, 0.38 M, and highest concentration 0.92M (−+−+−).

TABLE 2: PCA of 1-Heptanol (1912−2204 nm) Baseline-Corrected Data

PC no.	eigenvalue	explained variance (%)
1	37.8129092	97.2437565
2	1.0688723	2.7488274
3	0.0025931	0.0066688
4	0.0001992	0.0005122
5	0.0000825	0.0002121
6	0.0000031	0.0000079
7	0.0000011	0.0000027
8	0.0000008	0.0000020
9	0.0000007	0.0000019
10	0.0000004	0.0000011

The calculated spectral profiles were normalized (Figure 9), and then the concentration profiles were estimated using least squares (Figure 10).

The third concentration profile was checked by projecting the whole matrix onto a plane perpendicular to the two-component region in the concentration direction. The bands in the spectra in the projected matrix that have nothing in common with the spectra in the two-component region will be left in every row of the resulting matrix. The sum is taken for every row, and if the profile calculated by using OP coincides with ITTFA, the third profile is accepted (see Figure 10).

Interpretation of the Resolved Profiles. Figure 11 shows the resolved profiles of the spectra of all of the alcohols. The spectra of the monomers show several sharp bands, whereas the associated forms show one large band. This confirms that the first component corresponds to a free-OH stretch. On the contrary, the associated forms are expected to give a broad band. Table 3 provides an overview of the bands present in the resolved spectral profiles of the investigated alcohols. The resolved monomer profiles of the linear alcohols are similar. The branched alcohols differ greatly in band frequencies. The position of the OH group in the molecule is reflected in the spectra of the monomer as the number of bands, their shapes, and their intensities in the combination region of the OH stretch.

A molecule may have a large number of vibrational modes. Some of them can be associated with the vibrations of individual bonds or functional groups (as the fundamental OH stretch of the free-OH stretch), whereas others must be considered to be vibrations of the whole molecule. The OH combination region investigated here is assigned to the combination of the OH stretch at about 3500 cm^{-1} and OH deformations at about 1300 cm^{-1} in the MIR region.[31] More precisely, a sharp band at 3650−3590 cm^{-1} is assigned to the free-OH stretch, a broad band at 3600−3200 cm^{-1}, to the hydrogen-bonded OH stretch, and a series of absorption bands at low energy at about 1410−1260 cm^{-1}, to OH deformation.[32] The last region belongs to the fingerprint region and exists because of weak vibrations of the whole molecule. The SDBS web site[33] lists infrared spectra of different alcohols in CCl_4. The spectrum of 1-heptanol shows bands at 3638 and 3344 cm^{-1} that are assigned to the fundamental OH stretch and sharp band at 1468, 1460, 1379, and 1342 cm^{-1} that are assigned to OH deformations. This is expected to result in many bands in the OH combination region

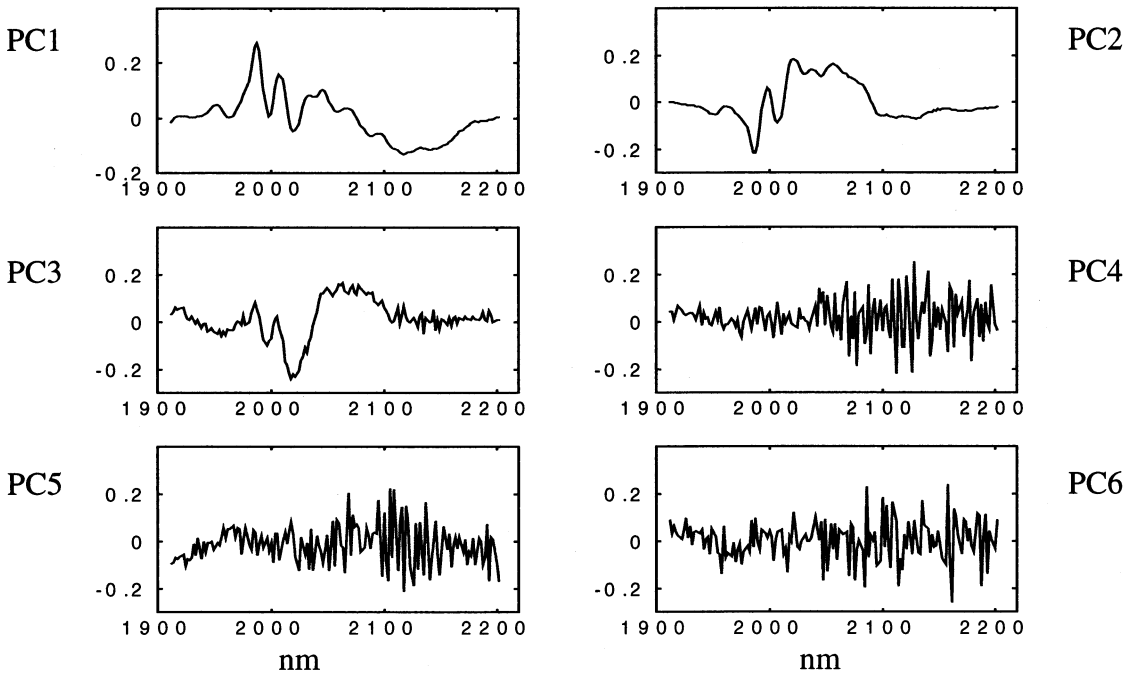

Figure 5. Six first loadings plotted vs wavelength. Data has been differentiated in the spectral direction.

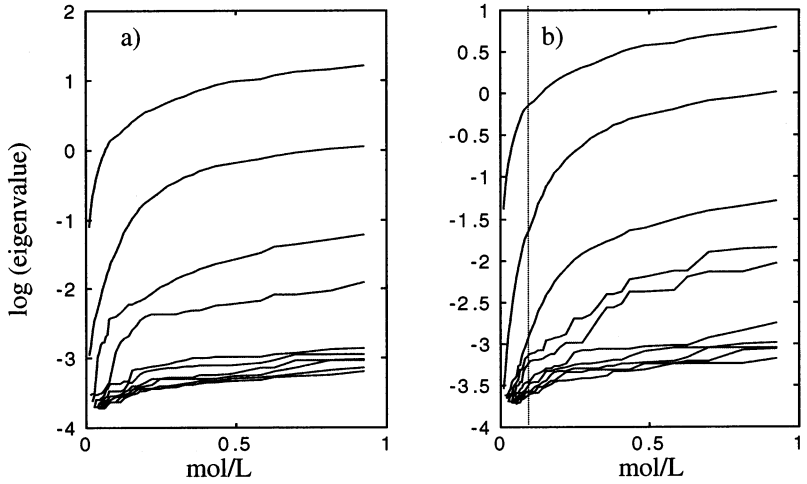

Figure 6. EFA plot (a) before and (b) after baseline correction in the concentration direction of 1-heptanol.

in the NIR frequency region. To be able to give exact band assignments to each sharp band that is observed in the resolved monomer spectra in the combination region of each alcohol, it is necessary to scan the spectra of alcohols solutions in the infrared region and to perform curve resolution or a correlation study with the peaks obtained in the combination region. The structures of the resolved profiles of the monomer bands are expected to result from the combination of the fundamental OH bands and the OH deformation bands and are characteristic for each alcohol. However, the bands due to associated species reveal only one broad band per associated species. Because no narrow peaks are observed in these profiles, is seems that these bands are masked by the broad bands from the fundamental frequencies.

The spectra of associated components give on the other side a maximum intensity at almost the same wavelength, independent of the type of alcohol. For linear components, the maximum intensity is around 2078 nm, whereas for cyclic components, it lies about 10 nm higher. The broad bands assigned to the associated species are expected to originate from a combination of a band in the fingerprint region and the broad band at about 3600−3200 cm^{-1}. Nodland[6] and Førland et al.[7] obtained a small, sharp peak under the monomer band of the resolved spectrum of the linear aggregates, which they have assigned to the free-OH stretch at the end of the linear aggregate (see Figure 1). By inspecting the resolved bands in Figure 9, one can observe that the band is missing. The gradient between the monomer and the linear aggregates is not so different because they both increase with increasing concentration in the two-component region. The zero-concentration regions chosen for baseline correction increase with increasing concentration and with different gradients (see Figure 3a). The baseline correction used here is expected to influence the gradient of the absorption bands. Small bands such as the free-OH stretch of the linear aggregates are expected to be the most vulnerable because of this baseline correction. The free-OH of linear aggregates may obtain a gradient that is similar to that of the free monomer band, and the result will be that this band is resolved together with the other monomer bands.

The separation between bands assigned to the two different associated species in the fundamental region is larger than 100 cm^{-1}.[6,7] Here, the average separation between the resolved

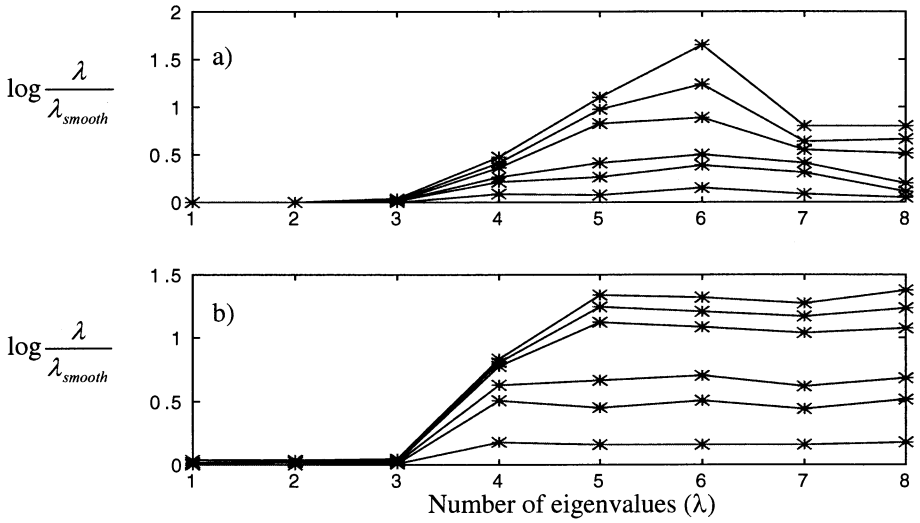

Figure 7. Smooth PCA in (a) the concentration direction and in (b) the spectral direction with second-order differentiation. (*) indicates the degree of smoothing, *K*. *K* values used here are 0.1, 0.5, 1, 5, 10, and 20.

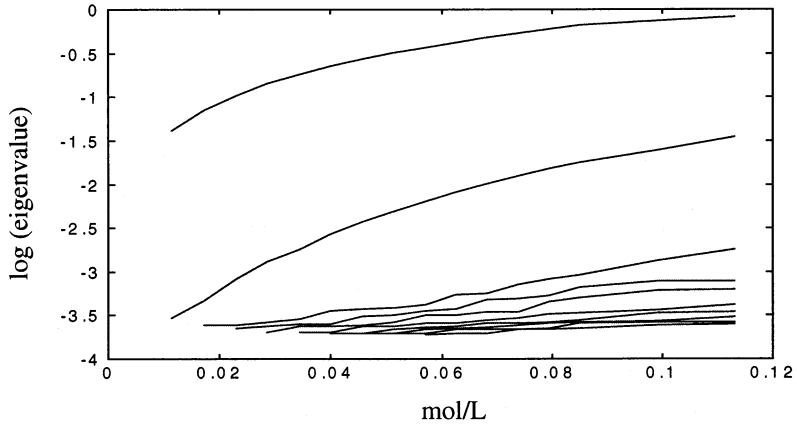

Figure 8. EFA plot of 1-heptanol in the concentration region 0.006−0.113 M (baseline-corrected data).

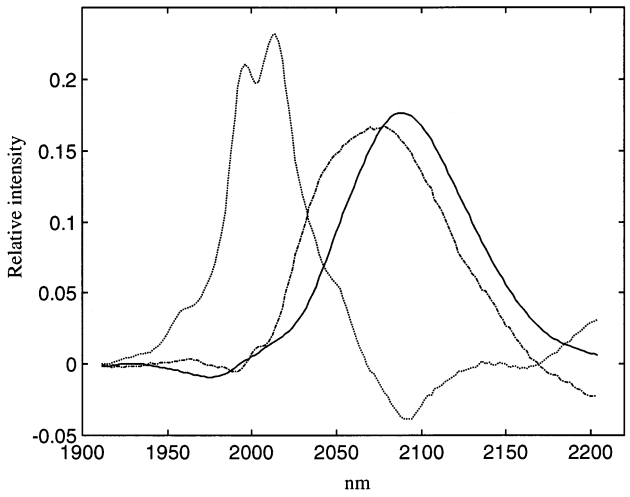

Figure 9. Resolved spectral profiles of 1-heptanol from 1912−2204 nm. First (− − −), second (− · − · −), and third components (—) found by ITTFA.

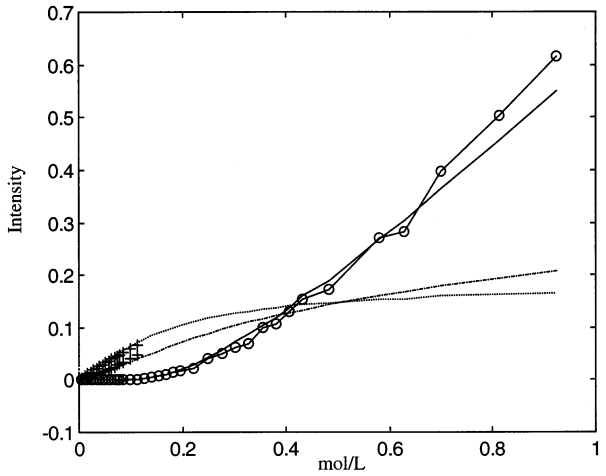

Figure 10. Resolved concentration profiles of 1-heptanol in the near-infrared region 1912−2204 nm. The first (− − −) and second components (− · − · −) were found with ETA and the least-squares method, respectively, and the third component was found by ITTFA (—) and OP (−○−○−). In addition, the concentration profiles of the first and second components (−+−+−) found in the two-component system are shown. All of the profiles are scaled to the right intensity.

spectral profiles for the linear and cyclic aggregates is found to be 23 cm^{-1}. An explanation for the decrease in separation is that bands in the NIR region are broader than the bands in the MIR region. In addition, the contribution from several bands from the fingerprint region is expected to cause an additional broadening of the combination bands.

The concentration profiles were scaled to the right intensity before further analysis (Figure 12). The concentration profiles split into a linear and a branched group. The concentration of the monomer band stagnates at about 0.3 M. The gradient of

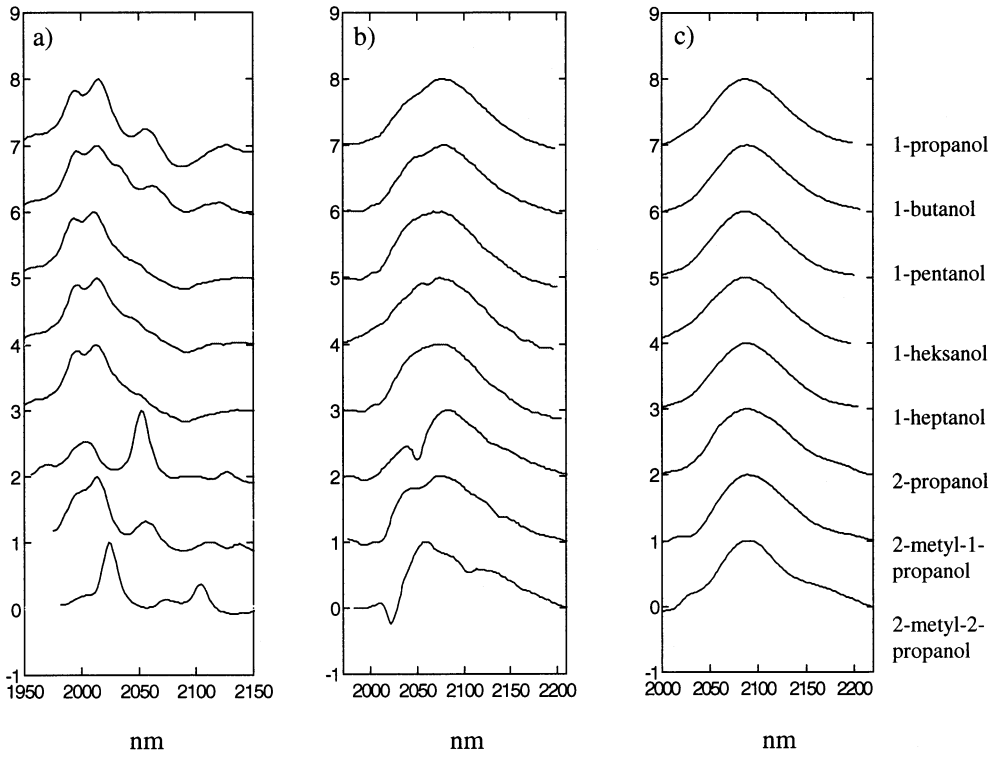

Figure 11. Resolved spectra of the different alcohols investigated: (a) monomers, (b) linear aggregates, and (c) cyclic aggregates.

TABLE 3: Band Observed in the Resolved Spectra

alcohol	spectral region (nm)	monomer (nm)	linear aggregates (nm)	cyclic aggregates (nm)
1-propanol	1946−2198	1994, 2016, 2056	2078	2088
1-butanol	1936−2206	1996, 2014, 2062	2078	2088
1-pentanol	1910−2200	1996, 2014	2078	2086
1-hexanol	1936−2206	1996, 2014	2074	2086
1-heptanol	1912−2204	1996, 2014	2078	2088
2-propanol	1956−2220	1970, 2004, 2052	2038, 2084	2088
2-methyl-1-propanol	1976−2224	2014, 2056	2074	2090
2-methyl-2-propanol	1981−2222	2024, 2074, 2104	2058	2088

the cyclic aggregates is higher than that for the linear aggregates. The ratios between the concentration profiles of the aggregates and the monomer are smaller for the branched alcohols than for the linear alcohols. This result is expected to reflect the fact that it is more difficult for branched molecules to form aggregates because of steric hindrance.

Association Degrees and Equilibrium Constants. The formula for calculating the equilibrium constant for the self-association of alcohols is given in eq 5. Table 4 provides the association numbers and the equilibrium constants for the alcohols investigated in this paper. The linear aggregates have, in general, smaller average sizes and equilibrium constants than do the cyclic aggregates. In addition, it is observed that the alcohols divide into two groups. The branched alcohols have smaller equilibrium constants than the linear alcohols. This is expected because branched alcohols have greater difficulty in aggregating than do linear alcohols. These are the same observations that Nodland[6] and Czarnecki et al.[11] made. Branched alcohols give an average size of the aggregates equal to 2−3 for linear and 5−6 for cyclic aggregates.

Linear alcohols show, in general, an average size equal to three for linear aggregates and 6−11 for cyclic aggregates (i.e., the average size is generally greater than that observed for the steric hindered alcohols). The linear alcohols show a relatively large range of association numbers that cannot be explained by structural differences. The average sizes of the linear aggregates of 1-hexanol and of the cyclic aggregates of 1-butanol are small, and the sizes of the cyclic aggregates of 1-pentanol and 1-hexanol are large compared to the average size obtained for linear alcohols. The small structural differences in the linear alcohols do not agree with these observations. The method for the determination of n is expected to contribute to the differences observed. It is the relationship between the concentration profiles of the monomer and the aggregates that is being studied. When calculating the equilibrium constants, the aggregate concentrations are being divided by the monomer concentration raised to the power of n (see eq 5). The result is that the equilibrium constants change to a relative large extent if n changes. If the monomer concentration decreases, this method provides unreliable results. This is especially the case for the linear alcohols.

Another contribution to the differences in the quantitative results given in Table 4 is the problem of deconvoluting spectra. The NIR region consists of many broad bands, and the combination region investigated here has a baseline (see Figure 2). To be able to resolve the spectra into pure spectral and concentration profiles, a slightly rough baseline correction was performed. In addition, some of the selective regions may not be completely pure. This affects the concentration profiles and result in less precise quantitative results. Spectra with no baselines having contributions only from the alcohol species

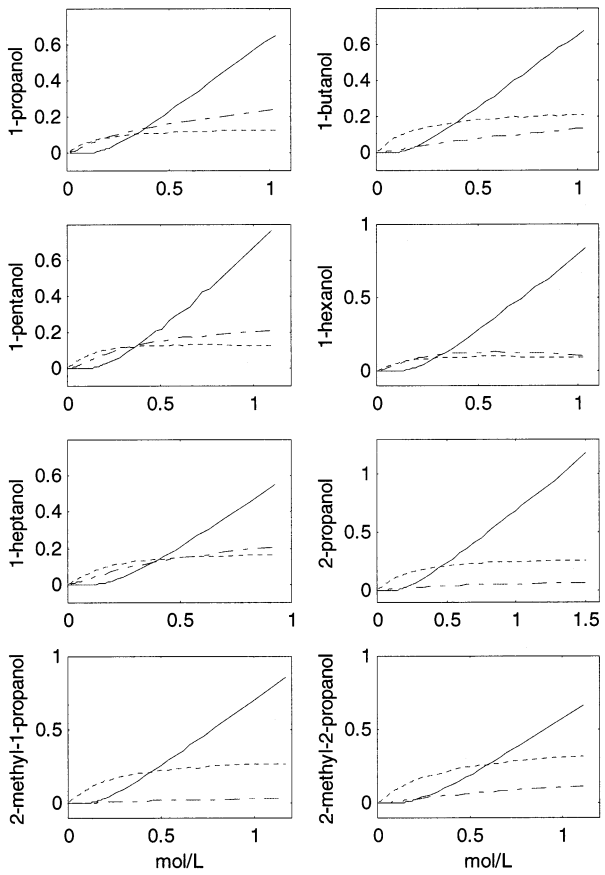

Figure 12. Scaled concentration profiles of the resolved alcohol systems: monomers (− − −), linear aggregates (EnDash− - − -), and cyclic aggregates (—).

TABLE 4: Association Numbers and Equilibrium Constants Calculated for the Given Concentration Regions

	linear aggregates		cyclic aggregates		
alcohol	size (*n*)	equilibrium constant (*K*)	size (*n*)	equilibrium constant (*K*)	concentration (M)
1-propanol	3	109.9	8	8.1×10^6	0.30−0.70
1-butanol	3	13.5	6	7.7×10^3	0.40−1.03
1-pentanol	3	9.6	10	2.5×10^8	0.20−1.10
1-hexanol	2	13.4	11	3.4×10^{10}	0.23−0.65
1-heptanol	3	44.9	8	9.0×10^5	0.41−0.92
2-propanol	2	0.9	6	2.6×10^3	0.38−1.04
2-methyl-1-propanol	3	1.7	6	2.1×10^3	0.42−1.17
2-methyl-2-propanol	2	1.1	5	211	0.40−1.11

are expected to provide more precise results. The problem is to find a way to remove baseline contributions from the spectra without affecting the "true" profiles. Here, linear removal of the baseline was used. Upon inspecting Figure 2, one observes that the *n*-pentane spectrum rises in intensity around 2150 nm. This information could be used, and instead of drawing a linear function from the two zero-concentration regions, a nonlinear or an exponential function could provide an acceptable baseline correction.

The combination region was used because of problems in removing the CH stretch in the first overtone of the OH stretch region. Because little knowledge of the OH combination region exists, the first overtone of the OH stretch might be a better region to use in studying the self-association of high-concentration samples of alcohols. Curve resolution in this region is complicated because of the contribution from the CH stretch. GBS[26] using *n*-pentane was tested on 1-pentanol. After the subtraction, negative spectral regions were obtained, and the rank in the first overtone of the OH stretch region was the same before and after removal of the CH stretch. Iwahashi et al.[12] used compounds in which the OH hydrogen was replaced by deuterium. The oxygen−deuterium bond is not expected to give absorptions in the first overtone of the OH stretch region, whereas the CH stretches in the deuterated alcohol are expected to have force constants closely related to the CH stretches in an alcohol having the same carbon skeleton. If this is true, GBS could be performed using the deuterated alcohol spectra; thereafter, resolution of the spectra in the first overtone of the OH stretch region can be performed. This method could probably also be used for removal of the baseline in the combination region.

Conclusions

A procedure for resolving the NIR spectra of the OH combination region (1900−2200 nm) has been presented. For all of the alcohols studied, the rank was equal to three. The resolved profiles indicate a system consisting of monomers, linear aggregates, and cyclic aggregates where the size of the aggregates increases with increasing concentration. For the branched alcohols, the average size of the aggregates is smaller than that for linear alcohols because of steric hindrance. The cyclic aggregates had higher equilibrium constants than the linear aggregates, and the alcohol systems at low to moderate concentrations are expected to comprise a majority of the cyclic aggregates.

Acknowledgment. This project has been has been supported through a Ph.D. grant to Laila Stordrange from the Norwegian Research Council, project no. 138494/432. Dr. Egil Nodland at the University of Bergen is acknowledged for valuable comments.

References and Notes

(1) Holtzclaw, H. F.; Robinson, W. R.; Odom, J. D. *General Chemistry with Qualitative Analysis*, 9th ed.; D. C. Heath: Lexington, MA, 1991.

(2) Hart, H. *Organic Chemistry*, 8th ed.; Houghton Mifflin Company: Boston, MA, 1991.

(3) Brink, G.; Glasser, L. *J. Phys. Chem.* **1978**, *82*, 1000.

(4) Shinomiya, K.; Shinomiya, T. *Bull. Chem. Soc. Jpn.* **1990**, *63*, 1093.

(5) Van Ness, H. C.; Van Winkle, J.; Richtol, H. H.; Hollinger, H. B. *J. Phys. Chem.* **1967**, *71*, 1483.

(6) Nodland, E. *Appl. Spectrosc.* **2000**, *54*, 1339.

(7) Førland, L. Y.-z.; Kvalheim, O. M.; Høiland, H.; Chazy, A. *J. Phys. Chem. B* **1997**, *101*, 6960.

(8) Fletcher, A. N.; Heller, C. A. *J. Phys. Chem.* **1967**, *71*, 3742.

(9) Fletcher, A. N.; Heller, C. A. *J. Phys. Chem.* **1968**, *72*, 1839.

(10) Fletcher, A. N. *J. Phys. Chem.* **1972**, *76*, 2562.

(11) Czarnecki, M. A. *J. Phys. Chem. A* **2000**, *104*, 6356−6361.

(12) Iwahashi, M.; Suzuki, M.; Katayama, N.; Matsuzawa, H.; Czarnecki, M. A.; Ozaki, Y.; Wakisaka, A. *Appl. Spectrosc.* **2000**, *54*, 268−276.

(13) Errera, J.; Mollet, P. *Nature (London)* **1936**, 882.

(14) Czarnecki, M. A.; Maeda, H.; Ozaki, Y.; Suzuki, M.; Iwahashi, M. *Appl. Spectrosc.* **1998**, *52*, 994−1000.

(15) Kvalheim, O. M.; Liang, Y.-z. *Anal. Chem.* **1992**, *64*, 936.

(16) Vandeginste, B. M.; Derks, W.; Kateman, G. *Anal. Chim. Acta* **1985**, *173*, 253.

(17) Lorber, A. *Anal. Chem.* **1986**, *58*, 1167.

(18) Kvalheim, O. M.; Liang, Y.-z. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 283.

(19) Shrager, R. I. *Chemom. Intell. Lab. Syst.* **1986**, *1*, 59.

(20) Maeder, M.; Zuberbuehler, A. D. *Anal. Chim. Acta* **1986**, *181*, 287.

(21) Maeder, M. *Anal. Chem.* **1987**, *59*, 527.

(22) Silverman, B. W. *Ann. Stat.* **1996**, *24*, 1.

(23) Liang, Y.-z.; Kvalheim, O. M.; Manne, R. *Chemom. Intell. Lab. Syst.* **1993**, *18*, 235.

(24) Liang, Y.-z.; Kvalheim, O. M.; Rahmani, A.; Breereton, R. G. *J. Chemom.* **1993**, *7*, 15.

(25) Iwahashi, M.; Hayashi, Y.; Hachiya, N.; Matzusawa, H.; Kobayashi, H. *J. Chem. Soc., Faraday Trans. 1* **1993**, *89*, 707.
(26) Lorber, A.; Goldbart, Z.; Halon, A. *Anal. Chem.* **1985**, *57*, 2537.
(27) Liang, Y.-z.; Kvalheim, O. M.; Rahmani, A.; Breereton, R. G. *Chemom. Intell. Lab. Syst.* **1993**, *18*, 265.
(28) Chatterjee, S.; Price, B. *Regression Analysis by Example*; Wiley & Sons: New York, 1997.
(29) Shen, H.; Stordrange, L.; Manne, R.; Kvalheim, O. M.; Liang, Y.-z. *Chemom. Intell. Lab. Syst.* **2000**, *51*, 37.
(30) Liang, Y.-z.; Manne, R.; Kvalheim, O. M. *Chemom. Intell. Lab. Syst.* **1994**, *22*, 229.
(31) Osborne, B. G.; Fearn, T.; Hindle, P. H. *Practical NIR Spectroscopy with Applications in Food and Beverage Analysis*, 2nd ed.; Longman Scientific & Technical: Essex, England, 1993.
(32) Williams, D. H.; Flemming, I. *Spectroscopic Methods in Organic Chemistry*, 5th ed.; McGraw-Hill: New York, 1995.
(33) http://www.aist.go.jp/RIODB/SDBS/.