

2020

Quantification of Low-content Encapsulated Active Cosmetic Ingredients in Complex Semi-solid Formulations By Means of Attenuated Total Reflectance-infrared Spectroscopy

Franck Bonnier

Université Francois-Rabelais de Tours, franck.bonner@univ-tours.fr

Lynda Miloudi

Université de Tours

Sandra Henry

Université de Tours

See next page for additional authors

Follow this and additional works at: <https://arrow.tudublin.ie/nanolart>

 Part of the [Chemistry Commons](#)

Recommended Citation

Bonnier, F. et al (2020). Quantification of low-content encapsulated active cosmetic ingredients in complex semi-solid formulations by means of attenuated total reflectance-infrared spectroscopy. *Analytical and Bioanalytical Chemistry*, vol.412, pp.159-160. doi:10.1007/s00216-019-02221-2

This Article is brought to you for free and open access by the NanoLab at ARROW@TU Dublin. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie.



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](#)

Authors

Franck Bonnier, Lynda Miloudi, Sandra Henry, Dominique Bertrand, Clovis Tauber, Xavier Perse, Florent Yvergnaux, Hugh Byrne, Igor Chourpa, and Emilie Munnier

Quantification of low content encapsulated active cosmetic ingredients in complex semi-solid formulations by means of Attenuated Total Reflectance-Infrared spectroscopy

Franck Bonnier^{*a}, Lynda Miloudi^a, Sandra Henry^a, Dominique Bertrand^b, Clovis Tauber^c, Xavier Perse^a, Florent Yvergnaux^d, Hugh J. Byrne^e, Igor Chourpa^a, Emilie Munnier^a

^aUniversité de Tours, EA 6295 Nanomédicaments et Nanosondes, 31 avenue Monge, 37200 Tours, France

^bData_Frame, 25 rue Stendhal, 44300 Nantes, France

^cUniversité de Tours, UMR U1253 iBrain, Inserm, 37032 Tours, France

^dBioeurope - Solabia Group, Anet, France

^eFOCAS Research Institute, Technological University Dublin, Kevin Street, Dublin 8, Ireland

***Corresponding author**

Franck Bonnier

EA 6295 NMNS "Nanomédicaments et Nanosondes"

UFR Sciences Pharmaceutiques

31, avenue Monge, 37200 Tours

E-Mail: franck.bonnier@univ-tours.fr

Tel: +33 2 47 3 67307

Key words: Infrared spectroscopy, Attenuated Total Reflectance, Omegalight®, Alginate Nano-Carriers, multivariate analysis, Hydrogels, label free quantification

Abstract

Attenuated Total Reflectance-Infrared (ATR-IR) spectroscopy is a robust tool for molecular characterisation of matter. Applied to semi-solid formulations, it enables rapid and reliable data collection without pre-analytical requirements. Based on nanoencapsulated Omegalight®, a skin lightening active cosmetic ingredient (ACI), incorporated in a hydrogel, it is first demonstrated that, despite the high water content and the chemical complexity of the samples (i.e. number of ingredients), the spectral features of the ACI can be detected and monitored. Secondly, with a total of 105 samples divided into a training set (n = 60) and a unknown set (n = 45) covering a 0.5% w/w – 5% w/w concentration range, the study further investigates the quantitative performance of ATR-IR coupled with Partial Least Squares Regression (PLSR). Through a step by step approach to testing different cross validation protocols, accuracy (Root Mean Square Error of Cross Validation – RMSECV) and linearity between the experimental and predicted concentrations (R^2) of ATR-IR are consistently evaluated to be respectively 0.097% (w/w) and 0.995 with a lower LOD = 0.067% (w/w). Subsequently, further evaluation of the accuracy (relative error of the predicted concentration compared to the true value, expressed as %) of the analysis was undertaken with the 45 unknown samples were defined as and analysed by PLSR. The outcome of the analysis demonstrates the ruggedness and the consistency of the determination performed using the ATR-IR data. With an average relative error of 2.5% w/w and only 5 samples out of 45 blind samples exhibiting relative error above the 5% threshold, high accuracy-quantification of the nano-encapsulated ACI can be unambiguously achieved by means of the label-free and non-destructive technique of ATR-IR spectroscopy. Ultimately, the study demonstrates that the analytical capabilities of ATR-IR hold significant potential for applications in the cosmetics industry and although the path remains long; the present study is one step further to support validation of the technique, albeit for the specific case of Omegalight®.

1. Introduction

Fourier Transform Infrared (FTIR) spectroscopy measures the absorption of incident IR radiation following interaction with a sample (1) and is a well-established, non-destructive and label free characterisation technique enabling collection of specific molecular finger prints from samples (2). It is a particularly popular analytical technique in chemistry laboratories, due to the low instrumental costs and ease of collection of IR signatures, providing detailed information about the chemical composition, useful for monitoring chemical reactions or analysing synthesis end products. (3). Bench top hyper spectral imaging FTIR apparatus delivers high performance for analytical applications (2), and, has gained interest in cosmetic and pharmaceutical analysis, since they allow to map larger sample area like that of products packaging (4), polymer films (5), tablets (6) or biological materials such as tissue sections or cells (7).

The advent of Attenuated Total Reflectance (ATR) accessories has greatly contributed to increasing the speed of acquisition and reliability of IR analysis through direct deposition of samples on top of the ATR crystal without any particular preparation (8). Despite the low penetration depth of the evanescent wave (typically 0.5 – 5 μm) (9) rapid and specific analysis of solid raw material such as powders or liquid samples can easily be achieved. (8, 10). Ultimately, ATR-IR spectroscopy is an attractive low cost alternative to other gold standards such as High Performance Liquid Chromatography (HPLC) (11), the main advantage remaining the minimal sample preparation and lack of requirement for separation steps prior to analysis (12).

Despite the wide range of applications of ATR-FTIR reported in scientific literature, their translation to cosmetic research and industry remains limited. Although it is widely accepted that Active Cosmetic Ingredients (ACI) or Active Pharmaceutical Ingredients (API) can be detected and characterised using their vibrational bands (13), the ATR-IR technique seems to suffer from the lack of recognition as a quantitative tool in these particular fields and probably from a widespread assumption that it does not allow analysis of liquid samples. Therefore, examples of investigation of liquid samples rather focus on the quality control and quantitative analysis performed on particular systems such as essential oils (14) and other plants extracts (15, 16), generally lipid based, thus with reduced contribution from water. However, it has been recently demonstrated that both sensitivity and specificity of the ATR-IR spectroscopy make it suitable for monitoring of ACIs when in semi-solid forms such as hydrogels (17). To date, while reported studies of formulation stability show an increasing popularity of the technique in the cosmetics field, there has been no in depth investigation of the analytical power of the ATR-IR spectroscopy to monitor quality of final cosmetic products (18).

Commercialised cosmetic products are complex systems with various composition and texture, depending on the targeted applications. Additionally, new concepts are emerging to achieve improved efficacy (19-21), enhanced stability (22, 23) or protection of ACIs against exogenous factors such as light (24) by formulating them in nano-carriers (NC). Although not yet implemented in cosmetic

products, development of coreshell nano-systems is a hot topic, and recent developments in preparation protocols are encouraging for rapid transfer to final products in the near future (25), especially for dermatological applications (20, 26) However, routine chemical screening of semi-solid formulations with encapsulated active ingredients can be challenging. While quantification can be performed with a combination of chemical extraction protocols followed by separative analytical techniques such as HPLC coupled to mass spectrometry (23, 27), extraction of ACI from nanovectors remains fastidious and expensive. In addition, this technique uses large volumes of solvent and consequently generates considerable amount of chemical waste which is in conflict with current concerns aiming to develop green chemistry alternatives (28).

The present study has been conducted on Omegalight® (Bioeurope, France) ACI, a commercialised lightening agent targeting melanocytes (25), encapsulated in an Alginate Nano-Carrier (ANC) and integrated into a typical commercialised hydrogel product (as specified by BioEurope). Semi-solid formulations (cream or gel) are largely found in commercialised cosmetic products, offering a convenient and pleasing texture for local application on skin. However, considering 1%-2% (w/w) Omegalight® is the concentration typically recommended by the providers, the ACI itself remains a small fraction of the overall composition. It implies other ingredients, present at higher concentrations, dominate the IR spectral signatures collected. The aim of the present investigation was to evaluate the analytical power of the ATR-IR technique to quantify the ACI®, but also to overcome some of the most common limitations encountered in recent studies (17, 29). The present study therefore includes a large number of samples in the training set (n=60) and employs unknown samples for validation of observations (n = 45) for the data mining.

A combination of optimal data pre-processing (30, 31) and multivariate statistical analysis such as Partial Least Squares Regression (PLSR) (32) (33) is necessary to provide optimal data mining. In this manner, the reliability and ruggedness of the ATR-IR methods was evaluated to highlight and explore possible validation of the technique for online screening and monitoring of encapsulated active cosmetic during the production process in an industrial environment. Multivariate calibration methods such as PLSR applied to spectroscopic data require the method to be trained according to the specific formulation analysed, prior to application of the technique to unknown samples. The present study has therefore investigated the reliability of the analysis when different batches of unknown samples (3 independent sets of 15 samples each) were prepared and over a 3 month period, providing additional insight about the reproducibility of the approach and the possibility to tailor the method for specific ACI/formulations.

2. Materials and methods

2.1 Preparation and characterisation of Omegalight®-loaded alginate-based nanocarriers (ANC_OL)

2.1.1. Reagents

Omegalight®, the skin-lightening ACI evaluated in this study, was provided by Bioeurope (Solabia group) under collaborative agreement. A number of other ingredients are present, either as part of nanocarriers or as additives commonly found in cosmetic products, conferring a high chemical complexity to the samples. The ingredients found in the nanocarrier composition are: polysorbate 80 (Seppic, France), sorbitan monooleate (Seppic, France), sodium alginates (Setalg, France), calcium chloride (Fisher Bioblock, France). To prepare cosmetic hydrogels corresponding closely to those commercially available, the following cosmetic excipients were mixed: sodium carboxymethyl cellulose as gelling agent (Acros organics, France), a mixture of preservatives, dehydroacetic acid and benzyl alcohol (Cosgard®, Aroma zone, France) and glycerol as a humectant (Cooper, France). The latter are categorised by the term additives, as they are not integrated in the nanocarrier shell.

2.1.2 Preparation of ANC_OL

Alginate nanocarriers (ANC) are core-shell nano-systems composed of a hydrophobic oily core surrounded by a hydrophilic alginate based shell, enabling incorporation of lipophilic ACI in aqueous samples such as hydrogels. Protocols used for preparation of ANC_OL are based on previous worked published by Miloudi et al. (29). Briefly, ANC_OL were prepared following the method of oil-in-water emulsification and ionic gelation, described in detail by Nguyen et al. (23), optimised to provide ANC_OL with the oily core containing only the hydrophobic ACI itself. Briefly, sodium alginate is placed to soak in ultrapure water until well swollen, followed by stirring to dissolution before filtration through a 0.45 µm nylon filter. The 0.6 g/L alginate solution prepared is then supplemented with polysorbate 80 (0.04 g/L) to form the aqueous phase of the emulsion. Secondly, the oil phase was prepared by mixing Omeagalight® and sorbitan monooleate (0.1g/L). Finally, the nano-emulsification is achieved by mixing the two phases under sonication (Vibra-cell ultrasonic processor, Sonics, 20 kHz) for 3 minutes. The gelation of the surface of the nanocarriers is achieved by addition of a solution of calcium ions (1 g/L).

2.1.3. Physico-chemical characterisation of ANC_OL suspensions

Physico-chemical characterisation was performed based on procedures developed for ANC_OL suspensions and described in reference (29). The average hydrodynamic diameter and polydispersity index of ANC_OL were assessed by Dynamic Light Scattering (DLS) using a NanoZS instrument (Malvern Instruments, UK). A 1/100 dilution in ultrapure water has been applied to each sample prior to measurement. The analysis was performed with a 633 nm laser source and a detection angle set at 173°. ANC_OL zeta potential (ζ) was measured with the same instrument with a detection angle set at 13°. All measurements have been performed in triplicate and at 25°C.

2.1.4 Preparation of Cosmetic product-like models

The study aims to evaluate ATR-IR spectroscopy as a quality control technique for cosmetic products of complex formulation, in this instance, the preparation of samples has been done using the formulation of commercialised hydrogel cosmetic products as a base in which the ANC_OL suspension is added. For instance, CarboxyMethyl Cellulose (CMC) is one of most commonly used gelling agent in hydrogel providing the viscous texture to the product. This ingredient is usually found at concentrations of about 1.5% (w/w). Additives such as humectant and preservatives also are common ingredients present in formulations. Cosgard® and glycerol were added to samples at final concentrations of respectively 1% (w/w) and 20% (w/w), to match the composition of end products. Ultimately, the samples analysed throughout this study have been prepared to accurately mirror industrial conditions with formulations corresponding to hydrogel cosmetic products currently commercialised.

To evaluate ATR-IR spectroscopy as quality control tool for complex cosmetic products, sets of samples over a range of concentrations for Omegalight® comprised between 0.5% and 5% (w/w) were included. Firstly, a training set of 60 samples (Set_01) was prepared by adding different amounts of ANC-OL (prepared as described above) into the formulation to reach target concentrations. For clarity, the list of samples from Set_01 and corresponding concentrations are summarised in table S1. Due to the viscosity of few ingredients, the concentrations provided in the table reflect the exact weight introduced during preparation, aiming to be as close as possible to targeted values. Secondly, 3 unknown sets of 15 samples each have been prepared at month+1, month+2 and month+3 (respectively Unknown_Set_01, Unknown_Set_02 and Unknown_Set_03) (Table S2). For each set of samples, freshly prepared ANC-OL suspensions were used, possibly exhibiting different properties. Unknown_Set_02 and Unknown_Set_03 were included in the study in an attempt to demonstrate the stability of the analysis over time and investigate how the ATR-IR performed when ANC have different size and concentration in ACI. The study aims to investigate the final step of production when the ACI (ANC_OL) is added to the matrix (gel). Under industrial conditions, these two fractions of the samples are prepared independently and it is assumed they are subjected to independent quality analysis prior to mixing to obtain the final cosmetic products. Therefore, it has been considered they met quality requirements for use and that their compositions (concentrations) have been confirmed. In that context, modifications in the matrix composition have not been treated in the present study and PLSR analysis has been performed on the basis of a binary ACI-matrix systems.

2.2 ATR-FTIR data collection analysis data handling

2.2.1. ATR-FTIR data collection

IR spectra were acquired using a Frontier spectrometer (Perkin Elmer, France) equipped with a Quest single reflection diamond attenuated total reflectance (ATR) accessory (Specac, UK). The spectral range was set between 4000-900 cm^{-1} and the spectral resolution at 4 cm^{-1} . The samples

prepared are semi-solid hydrogels which are comparable to liquids, but with higher viscosity. Drops of 200 μL were deposited directly onto the diamond surface and spectroscopic measurements were performed without delays. Prior to sample measurement, a background spectrum was recorded in air (4 scans) and automatically ratioed with the sample spectrum (4 averaged scans) by the software. Between each measurement, the ATR crystal has been cleaned with dionised water and tissue paper. For each sample, 5 deposits have been measured and 3 spectra per drop have been collected. Hydrogel samples behave similarly to liquid, and therefore no pressure was applied for recording of data. Ultimately, 15 spectra were recorded from each ANC_OL concentration, capturing the inter- and intra-variability during measurements. Spectra from pure compounds have also been collected using similar parameters.

2.2.2 Data Handling

The data pre-processing and analyses were performed using MATLAB (Mathworks, USA). Analysis can be strongly influenced by any interferences (instrumental response, artefact, sample heterogeneity) possibly affecting the overall quality of the data collected. Consequently, data pre-processing methods are critical in ensuring the best outcome, i.e. precision and accuracy (34). Preliminary studies demonstrated that Extended Multiplicative Signal Correction (EMSC) (35) was an appropriate pre-processing method for this type of sample. It has been applied using the EMSC toolbox for MATLAB, freely available from Nofima Data Modelling (<https://nofimamodeling.org/software-downloads-list/>). The toolbox enables data uploading and processing in a user-friendly interface. Full details about the correction and the use of the interface can be found in the tutorial published by Asfeth and Kohler (35). A basic EMSC model has been applied to the range 2200-900 cm^{-1} and found to be optimal, with the mean spectrum used as reference. The preprocessed data were analysed first using Principal Component Analysis then Partial Least Squares Regression (PLSR) (36). Principal Component Analysis (PCA) is a multivariate analysis technique that is widely used to simplify a complex data set of multiple dimensions (37). It allows the reduction of the number of variables in a multidimensional data set, although it retains most of the variation within the data set. The other advantage of this method is the derivation of PC loadings which represent the weight of each variable (wavenumber) in the data distribution across the scatter plot and for a given PC. Analysis of the loading of a PC can give information about the source of the variability inside a data set, derived from variations in the chemical components contributing to the spectra. Presently, PCA has been used for data exploration and therefore applied to raw spectra. PLSR delivers however an outcome evaluated in terms of ruggedness of the regression model and accuracy in predicted concentrations. The 3 parameters used to evaluate the quality of the analysis are the linearity between the experimental and predicted concentrations (R^2) and accuracy (Root Mean Square Error of Cross Validation – RMSECV, relative error of the predictive

concentration compared to the true value, expressed as %) The RMSECV is determined according to (38):

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{i=N} (\hat{c}_i - c_i)^2}{N}}$$

where \hat{C}_i is the predicted concentration, C_i the observed concentration and N the number of samples. Additionally, the lower Limit Of Detection (lower LOD) and the limit of quantification (LOQ) have been estimated based on the work published by Allegrini et al (39).

For the purpose of the study, 4 sets of samples are used; Set_01 containing 60 samples as the training set, and Unknown_set_01, Unknown_set_02 and Unknown_set_03, each containing 15 samples each. In order to avoid any bias during the analysis possibly leading to overoptimistic results, it is common to implement a Cross Validation (CV) procedure within the PLSR analysis. It implies using a calibration set to construct the predictive model, an independent test set to determine the optimal number of latent variable and independent sets of unknown samples. The experimental design of this study didn't include separate samples for calibration and test sets, but rather they were created by randomly splitting the 60 samples of Set_01. There are a number of different methods to randomly split the data into subgroups. In the current study, the Leave K-Out Cross Validation (LKOCV) method was used to ensure the 15 spectra collected from each sample are either all in calibration or in the test set. Ultimately 40 samples (2/3) of set_01 were randomly selected as calibration set, while the remaining 20 samples (1/3) were used as the test set. Once the number of latent variables is defined, blind samples, contained in Unknown_set_01, Unknown_set_02 and Unknown_set_03, were added to the model to estimate their concentrations. One can see that, following this approach, unknown sets are kept independent and don't have any influence on the construction of the PLSR model. Randomly dividing the training Set_01 into calibration and test sets requires verification that the outcome of the PLSR is consistent when different samples are selected as calibration. In order to include this variability into the results, a 100 fold iteration process was applied to the analysis, which is reflected in subsequent figures of the manuscript with error bars.

3. Results and discussions

3.1 Physico-chemical characterisation of ANC_OL suspensions

Average hydrodynamic diameter, polydispersity and ζ value of ANC_OL suspensions are provided in table 1. Notably, a polydispersity index below 0.2, indicates a narrow distribution. The physico-

chemical properties are in accordance with those observed in previous studies (23) and comply with use of nano-systems in skin cosmetic products.

Table 1. Summary of ANC_OL suspensions properties

[OL] % (w/w)	16.7
hydrodynamic diameter (nm)	198 ± 3
polydispersity index	0.12 ± 0.01
Z (mV)	-16 ± 1

3.2 – Characterisation of spectral variability in data set

The current study has been performed on hydrogel products which contain CMC as gelling agent (see 2. Materials and Methods). However, in order to produce a semisolid form, only 1.5% CMC is required, meaning the samples are predominantly composed of water. In addition, additives such as glycerol and Cosgard®, commonly found in commercial cosmetic products, have been added, respectively in proportions of 20% (w/w) and 1% (w/w), for consistency with final concentrations used in the cosmetics industry. Although the 1% (w/w) Cosgard® is expected to have little or no contribution to the IR spectra, Glycerol is a polyol with a number of strong features in its IR spectrum which could possibly interfere with the quantitative analysis. The ACI studied is a strongly hydrophobic molecule, difficult to include in hydrophilic formulations such as hydrogels, except via encapsulation in core-shell NC. Alginate is a natural biodegradable polysaccharide extracted from brown algae, and is particularly suited to formation of NC shells (40), in order to improve stability and compatibility with homogenous suspensions. Although the final encapsulating system is named Alginate Nano-Carriers (ANC), the shell contains other ingredients such as sorbitan monooleate and polysorbate 80, which act as surfactants in the NC shell.

Figure 1 presents the 15 raw spectra collected either from the lowest concentration, sample 01 with 0.5% (w/w), or from the highest concentration, sample 60 with 4.99% (w/w), tested in this study. Only the finger print region, 900-1800 cm^{-1} , is presented, which contains the most relevant spectral changes (highlighted by black arrows). The underlying signature originates from the water contribution, which is constituted by a main broad band centered on 1637 cm^{-1} (δ OH mode) (11). While the 950-1150 cm^{-1} region reflects the presence of glycerol at relatively high concentration, the C=O stretching band at 1739 cm^{-1} , the C-O stretching at 1111 cm^{-1} and 1163 cm^{-1} , the C=C stretching of the aromatic ring at 1446 cm^{-1} and 1504 cm^{-1} , the C-C stretching at 1493 cm^{-1} , are bands that can be specifically assigned the ACI. Other bands at 1039 cm^{-1} (ν (C-O), ν (C-C)), 1377 cm^{-1} (δ C-C, δ C-C-H) and 1464 cm^{-1} (δ (CH₃),

δ (CH₂) δ (C-C)) are shared with the sorbitan monooleate, but only subtle contributions from this ANC shell ingredient are suggested. Another observation that can be made is the high degree of reproducibility of the measurements performed. Figure 1 displays the 15 spectra collected from sample 01 and sample 60 exhibit quite low spectral variability. 5 drops have been deposited on the ATR crystal for each sample to ensure the samples containing the ANC_OL have good homogeneity in order to avoid interferences in the data analysis. However, due to overlapping spectral signatures originating from the different ingredients, and the weak variation of the active cosmetic ingredient, a multivariate analysis is required to deliver accurate quantitative analysis.

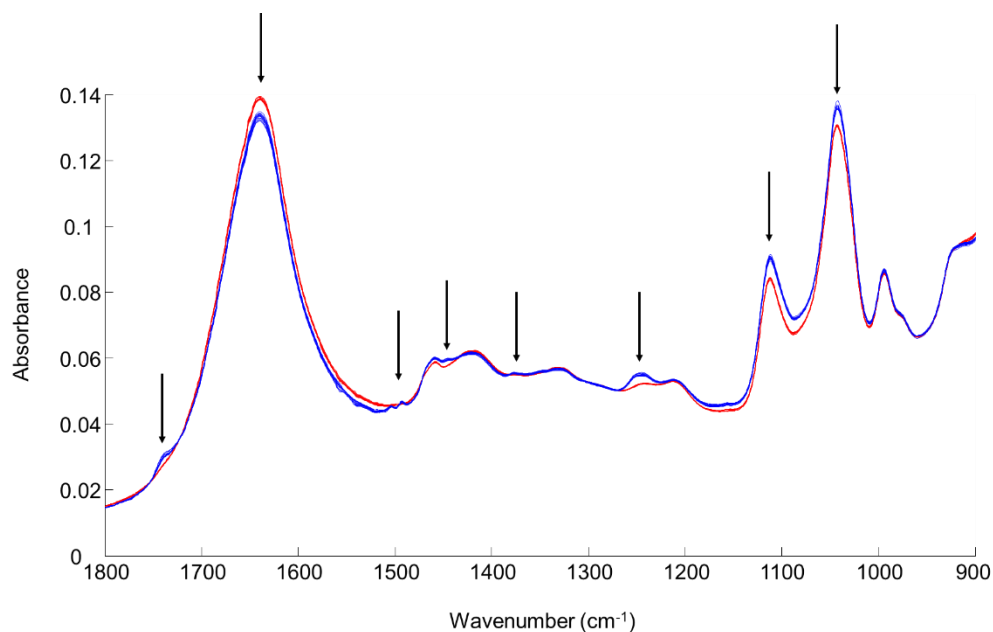


Figure 1. Raw spectra collected from sample 01 (red) and sample 60 (blue) respectively containing 0.5% and 4.99% (w/w). Arrows indicate the main features of interest that can be assigned to the presence of the ACI.

PCA is a valuable method for data visualisation, enabling identification of the spectral features exhibiting the highest variability within the data sets recorded. Figure 2 displays the PCA scatterplot of the two extreme concentrations tested in the present study, respectively 0.5% w/w (figure 2 red dots) and 4.99% w/w (figure 2 green dots). The two groups of spectra are well separated according to PC1, which accounts for 97.4% of the explained variance, while PC2 only accounts for 1.78%. The loading of PC1 (figure 3A) has unambiguous similarities with the pure spectrum of Omegalight®, including matching peak positions for previously identified features at 1739 cm⁻¹, 1504 cm⁻¹, 1493 cm⁻¹, 1464 cm⁻¹, 1447 cm⁻¹, 1378 cm⁻¹, 1111 cm⁻¹ and 1039 cm⁻¹. The negative band at 1700-1540 cm⁻¹ indicates that changes in water content also affect the IR spectra collected.

Notably, the use of nano-carriers could also affect the optical properties of samples, influencing the overall signal intensity. PCA has been performed on the raw spectra first, while some preprocessing could have a significant effect on the scatter plot. Nevertheless, it suggests that an increase in ACI would naturally induce noticeable spectral variability and that the loading of PC1 obtained from the PCA analysis can be considered as a reference for subsequent quantitative analysis, as it is expected that similar features would be identified as most relevant for the construction of PLSR predicted models.

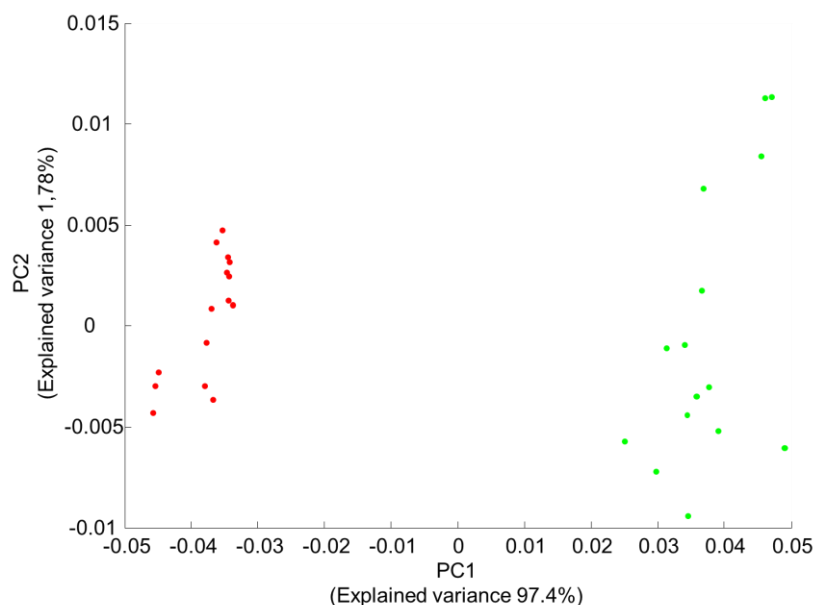


Figure 2. PC1 versus PC2 Scatter plot from PCA analysis performed on raw spectra collected from ANC_OL samples at 0.5% w/w (Red) and 4.99% w/w (green)

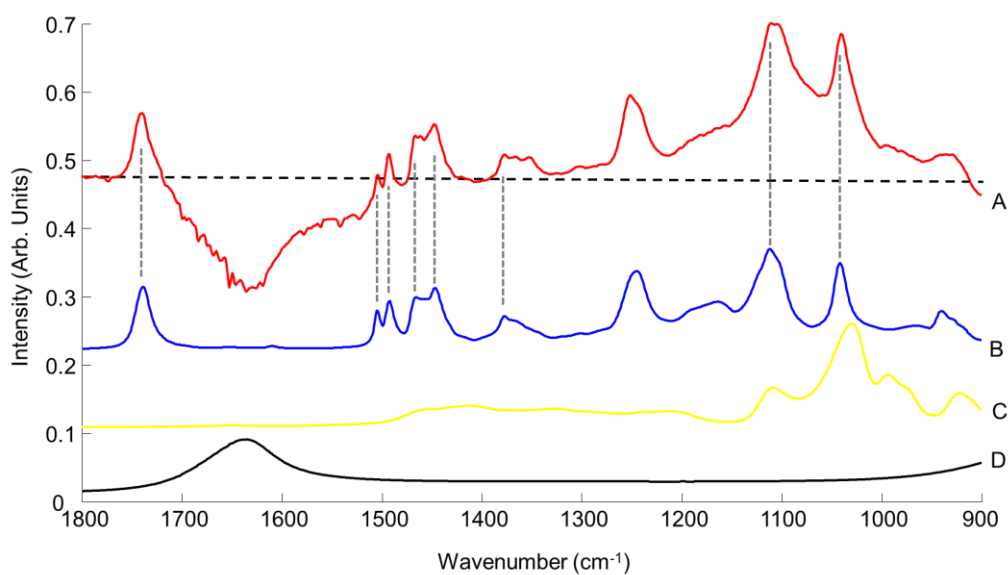


Figure 3. Loading 1 (A) obtained from the PCA analysis compared to pure spectra of ACI (B), glycerol (C) and dionised water (D). Spectra are offset for clarity.

3.3 Leave K-Out Cross Validation (LKOCV)

a) Construction of regression plot

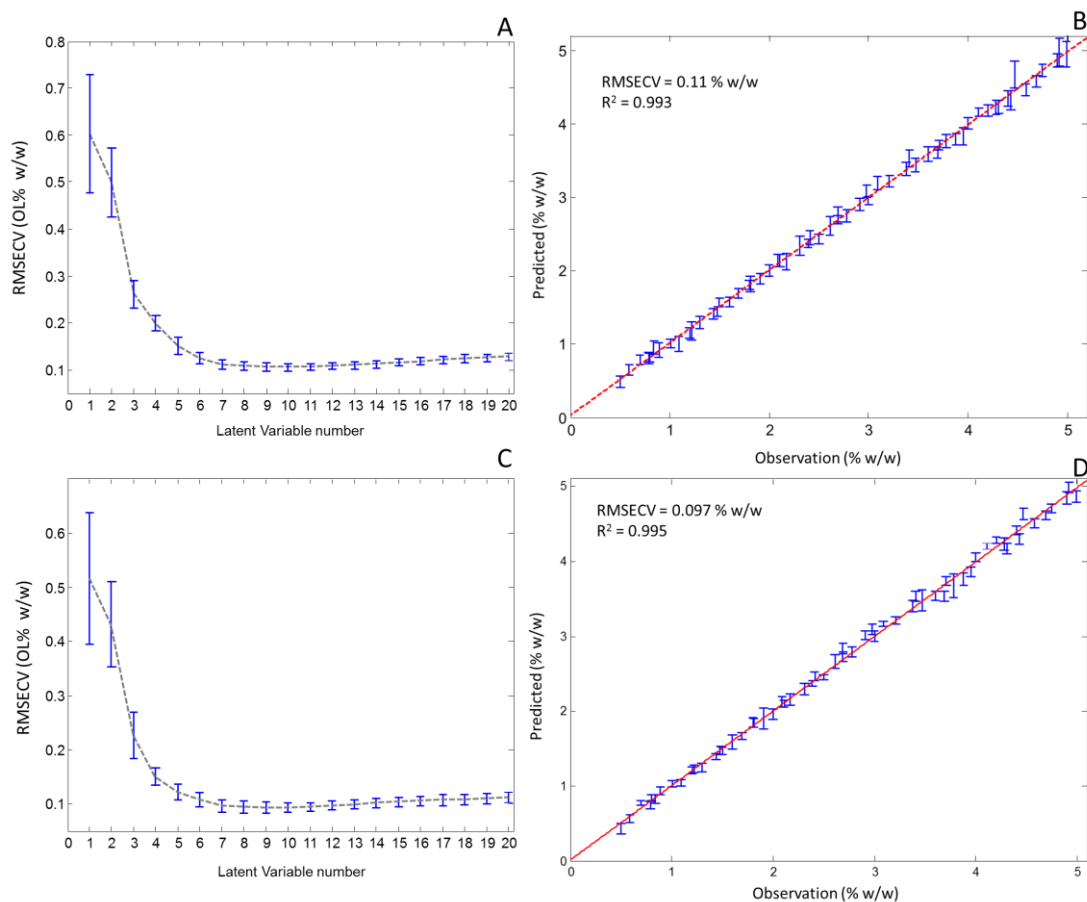


Figure 4 Construction of the predictive model with PLSR. **A and C:** Evolution of the RMSECV according to the latent variable number for respectively raw and EMSC corrected data and **B and D:** PLSR with projection of true concentrations (observation) against experimental concentrations (predicted) for respectively raw and EMSC corrected data.

Leave K-out Cross Validation (LKOCV) is widely used for quantitative analysis, to avoid overfitting in the results. In order to avoid any bias in the analysis due to the presence of samples represented by repetition in both the calibration and test set, the “K-out” is defined as K-groups-out or, more specifically for this study, K-samples-out. It means that the 15 spectra collected from a sample are either in the calibration or the test set, but cannot be split and found in both. The number of samples K is freely defined by the analyst, and presently 40 samples were used as calibration and the 20 remaining samples as test. Finally, the objectivity of the analysis is ensured by random selection of the K-samples with an iterative routine (n=100) implemented. Raw data deliver a RMSECV of 0.11% (w/w) with a R² value of 0.993 with 7 latent variables used for the model (Figure 4A and 4B). The corresponding lower

LOD and LOQ were respectively calculated to be 0.2% (w/w) and 0.06% (w/w). The outcome of the analysis is rather satisfying considering no preprocessing has been applied to data. The number of latent variables is defined by the analyst, based on the curve presented in figure 4A. Strictly speaking, the minimum RMSECV is reached with 10 latent variables, but it also appears that variations in RMSECV are minor above 7. In order to limit the interferences from noise contained in the spectra, it is rather advised to limit as much as possible the number of latent variables. Thus it has been preferred to not include more than 7 for this study. However, using less than 6 latent variables considerably decreases the quality of the predictive model. Ostensibly, the ACI is the only ingredient in the samples whose concentration varies, although it is necessary to bear in mind that it is achieved by changing the amount of ANC loaded in the formulation. Therefore, modifying the ACI concentration leads to modification of the chemical composition including all nano-carriers ingredients. As a result, expression of other vibrational modes can also be affected. Decreasing the API concentration would lead to a stronger contribution of other constituents in the IR spectra and vice versa (i.e glycerol and water). In addition to the modification in spectral features directly linked to the chemical composition, the optical properties can also be affected, generating underlying variations in the signatures not necessarily perceived by eye. Therefore, it is not surprising that the PLSR model requires an increased number of latent variables to deliver its best outcome. After EMSC, PLSR results are slightly improved, yielding RMSECV = 0.0936% (w/w) and $R^2 = 0.995$, although the difference compared to raw data remains insignificant based on those criteria. However, the lower LOD is reduced to 0.067% (w/w) and the LOQ is found to be 0.201% (w/w) suggesting the preprocessing of the data increases the reliability of the analysis. The LKOCV protocol highlights the ruggedness of the data sets in terms of consistency and reliability of the information contained in the spectra. Despite removing 20 concentrations from the training set (1/3 of the entire data set), the predictive models appear consistent (errors bars in figure 4A and C).

To better appreciate the meaning of the value of RMSECV obtained, it can be expressed as percentage of the average concentration of the range used for the study (2.75% w/w). For instance, before EMSC correction, the RMSECV is found to be 0.106% w/w, which represents a mean error equal to 3.86%. In comparison, following EMSC correction, the RMSECV is slightly lower (0.0936% w/w), resulting in a mean error of 3.4%. Taking the RMSECV as a parameter to evaluate the quality of PLSR outcome, preprocessed and raw data perform similarly.

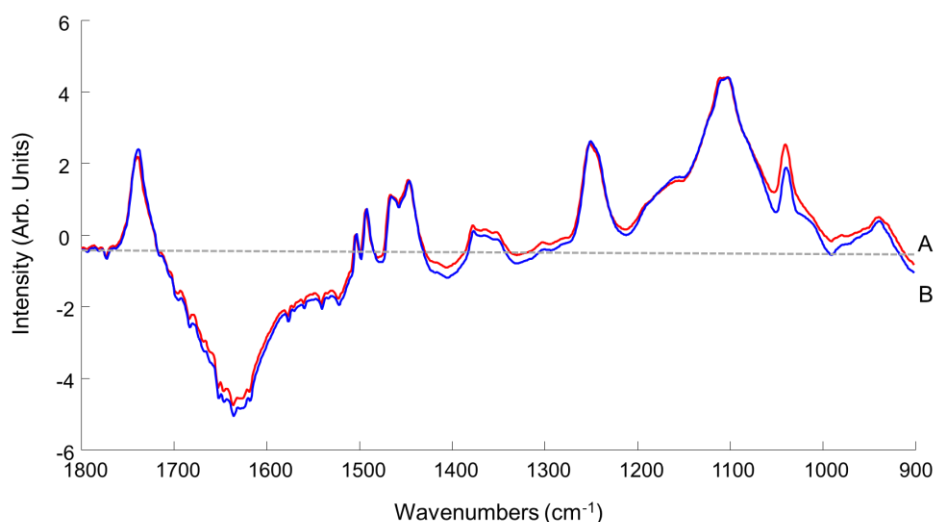


Figure 5. Regression coefficient obtained from the PLSR with Raw data (A - red) and following EMSC (B - blue).

IR spectra are known to be molecular fingerprints of the samples analysed, and therefore the question of specificity is often raised when performing advanced statistical analysis. Features used or selected to construct the predictive models can be monitored to ensure the analysis is based on relevant information contained in the spectra. For instance, the regression coefficient of the PLSR before and after EMSC correction (figure 5A and 5B) exhibits a number of positive and negative features over the spectral range 1800-900 cm^{-1} . Interestingly the ACI specific features at 1739 cm^{-1} , 1504 cm^{-1} , 1493 cm^{-1} , 1464 cm^{-1} , 1447 cm^{-1} , 1378 cm^{-1} and 1039 cm^{-1} , previously identified in the PCA (figure 3A) are also found. The broad negative band at 1637 cm^{-1} is attributed to water and appears negatively correlated with increase in ACI concentrations.

b) Estimation of accuracy

Evaluation of new analytical tools for quantification of specific molecules in a formulation environment requires calibration with reference concentrations which ultimately serve as a comparator for experimentally determined (or predicted) concentrations. In the present case, studying cosmetic formulations can simplify the process of establishing the calibration set, as the different steps of the sample preparation are precisely monitored and all ingredients precisely weighed before introduction to the preparation. However, although the RMSECV and the R^2 are relevant indicators of the quality of the regression model and reflect the overall accuracy of the predictive model, they fail to provide a clear estimation of the accuracy of quantification achieved. In order to obtain this information, it is essential to consider the samples independently and directly compare the predicted concentration with the true concentration.

This can be achieved using the LKOCV method illustrated above. Due to the high number of samples in the data set, it is difficult to provide a graphical representation summarizing the results, but rather they are gathered in table 6. The results are expressed as relative error in %, indicating the number of samples exhibiting relative errors below 1%, between 1% and 2.5%, between 2.5% and 5% and finally above 5%. Additionally, the mean relative error is given and the min-max values provided.

Before applying EMSC, the PLSR unambiguously delivers the best accuracy in predicted concentrations using the raw data, with a mean relative error of 1.82% w/w. It is also interesting to compare the sample distribution according to the relative error, by which only 3 samples are seen to be above 5% w/w relative error and 26 samples below 1% w/w. After applying EMSC, PLSR is consistent with mean relative error of about 2.22% w/w and now 4 samples are above 5% relative error.

Table 6. Summary of quantitative results for PLSR applied to the training set samples using the LKOCV approach. (Before and after applying EMSC)

	Pre-processing	Relative error (%)				Mean	Min-Max
		<1%	1-2.5%	2.5-5%	>5%		
PLSR	Raw data	26	19	12	3	1.82	0.03 - 11.76
	EMSC	18	24	14	4	2.22	0.12 -13.11

Ultimately, a quantitative technique, used as a formulation monitoring tool, should be able to deliver accurate concentrations in unknown samples. Reasoning by comparing relative errors provides better insights into the reliability of the methods, with surprisingly similar outcomes for raw and EMSC corrected data. The spectra collected with ATR-IR contain the specific spectral information sought for quantitative analysis of Omegalight® in the model studied. More importantly, the 60 samples of Set_01 were gathered at the early stage of the study to constitute the training set of data collected to estimate concentrations in blind samples. The intrinsic characteristic of multivariate calibration methods, such as PLSR, associated with spectroscopic techniques, is the requirement for a reliable training set to construct the prediction model. The results obtained with the LKOCV demonstrate the reliability of Set_01 and validate the relevancy to use this data set for further investigation on the Unknown_set_01, Unknown_set_02 and Unknown_tet_03.

3.4 – Quantification of Omegalight® in unknown samples

The final stage of the study is to demonstrate the validity of the quantification of the ACI by means of ATR-IR using a number of unknown samples categorized as Unknown_set_01, Unknwon_set_02 and Unknown_set_03. In this approach, the 3 additional blind sets (n = 3 x 15 samples) of data have been included and considered as unknown samples. Taking the preparation and analysis of Set_01 as T₀, Unknwon_set_01, Unknwon_set_02 and Unknwon_set_03 have been analysed at respectively T₁= 30 days (month+1), T₂= 60 days (month+2) and T₃= 90 days (month+3). Set_01 is used as training set for

the quantitative analysis, while Unknown_set_01, Unknown_set_02 and Unknown_set_03 are then input in the models as blind samples. In the conception of the study, no samples were specifically identified as calibration or test in Set_01 for the prediction of unknown samples. Therefore, the approach proposed is based on a random selection to objectively split the data to avoid adding any bias in the outcome of the analysis (i.e. not arbitrary select the best combination of samples). PLSR is performed using the Leave-K-out routine described above, with 2/3 of samples of training Set_01 randomly selected as calibration and the other 1/3 used as test. It is important to stress that the test sets are always composed of Unknown_set_01, Unknown_set_02 and Unknown_set_03 (n= 45). The number of optimal latent variables to be used is determined by the calibration/test sets, thus optimal parameters are fully defined independently from the unknown samples. Once the model is optimized, the 45 blind samples are predicted. Set_01 contains 60 samples offering an extremely large number of combinations of calibration/test sets which cannot all be computed. In order to provide an overview of the results and to maintain objectivity in the use of the algorithm, an iterative routine (n =100) has also been implemented in this approach. The 100 loops only concern the reset of the calibration and test sets with constant blind sets. It means that for each iteration all the 45 unknown samples are predicted for a given model and subsequently 100 different predictive models have been tested. The concentrations presented in table 7 are means resulting from the 100 loops.

Moreover, no noticeable link has been observed between the Unknown_set number (i.e. the time parameter) and the accuracy of the analysis outcome when performing PLSR, and therefore the results from the 3 blind sets have been merged and summarised in table 7.

Results from the PLSR following EMSC correction are quite satisfying, yielding mean relative error of Out of the 45 unknown samples, only 5 exhibit relative errors above the 5% threshold, with values respectively found at 5.52 % w/w (Unknown_set_01, month+1), 7.70% w/w (Unknwon_set_01, month+1), 8.69% w/w (Unknwon_set_02, month+2), 9.01% w/w (Unknown_set_02, month+2) and 13.20% w/w (Unknown_set_01, month+1). Despite a lack of accuracy from a few blind samples, all results are encompassed by a reduced window of relative error close to the 5% threshold. Results from the raw data remain quite similar with mean relative error of 2.77% and small difference in samples distribution, although only 6 are above 5% relative error. The ATR-IR set up is particularly suited for analysis of semi-solid formulations. The viscous samples behave as liquids, leading to perfect coverage of the ATR crystal, and a constant volume sampled by the evanescent field of the IR radiation propagating in the crystal. In such conditions, the IR spectra are strikingly reproducible with relevant features from the active ingredients detectable in the range of concentration tested. The similarity between raw and EMSC corrected data reflects the reliability of IR measurements of such samples. However, the crucial aspect demonstrated in the present study is the potential of ATR-IR spectroscopy when spectra collected from 4 sets of samples prepared independently and analysed over a 4 months period, in total, are compared and used in the same quantitative model. Using the initial set of 60

samples (month 0) as training set, the quantification can be achieved with high accuracy for the Unknown sets, prepared up to 3 months later. Although the approach for the PLSR is not the most conventional for prediction of the unknown samples, the random selection of samples to define the calibration and test sets further supports the reliability of the data collected in Set_01. Implementing an iterative loop combined to the randomness the algorithm used demonstrate that the outcome of the analysis is naturally satisfying without necessarily attempting to optimised the models to reach the best accuracy. It also suggests that the freshly made formulations generate IR signatures perfectly compatible with the first batch of samples, despite the quite large periods of time separating the data collection. Indeed, for numerous industrial analytical protocols, daily or weekly calibrations are performed from defined standards to account for the instrumental drift (most common example is for HPLC). However, the present study demonstrates that the regression models constructed using an initial dataset, are pertinent for the analysis of datasets recorded up to 3 months later, with no model update. This is highly encouraging and the consistency of the results obtained highlights the power of ATR-IR for this type of application.

Table 7. Summary of results for the 45 test samples with PLSR

Pre-processing	Relative error (%)					
	<1%	1-2.5%	2.5-5%	>5%	Mean	Min-Max
RAW	10	11	18	6	2.77	0.0421 – 8.34
PLSR	<hr/>					
EMSC	14	15	11	5	2.51	0.08- 13.20

4. Conclusion

ATR-IR spectroscopy is a reliable technique for label free molecular characterisation of samples. The low volume requirements and the reproducibility in the data collected are great advantages for quality control of pharmaceutical or cosmetic products. Additionally, the molecular specificity enables to detect and quantify the active ingredient in complex mixtures such as semi-solid formulations without requirement for pre-analytical protocols. Although the ACI is encapsulated in core-shell nanocarriers, IR analysis coupled with advanced data mining methods such as PLSR leads to robust quantitative models with RMSECV of 0.097% (w/w) (R^2 of 0.995) and lower LOD = 0.067% (w/w). Moreover, the results collected from 45 unknown samples clearly demonstrate the accuracy of the approach to predict sample concentrations with error below 5%. Considering the number samples included (n = 105) and

the quality of the results obtained in the present study, the potential of ATR-IR as tool for developing an Omegalight® assay in cosmetic product is illustrated. Combined with emerging high throughput technologies in the IR field, further developments could lead to short term recognition and adaptation of the methodologies in cosmetic and pharmaceutical processing. Validation of a quantitative technique is a long process, however, although this study has advanced the field through more extensive investigation of large samples set.

Conflicts of interest

There are no conflicts to declare

Acknowledgements

This work is part of the MISTIC 2017- 00118114. We thank Conseil Régional Centre Val de Loire and the Cosmétosciences program for financial support.

Bibliography

1. Larkin PJ. Chapter 2 - Basic Principles. In: Larkin PJ, editor. *Infrared and Raman Spectroscopy (Second Edition)*: Elsevier; 2018. p. 7-28.
2. Bunaciu AA, Aboul-Enein HY, Fleschin S. Application of Fourier Transform Infrared Spectrophotometry in Pharmaceutical Drugs Analysis. *Applied Spectroscopy Reviews*. 2010;45(3):206-19.
3. Wooley KL, Hawker CJ, Lee R, Fréchet JMJ. One-Step Synthesis of Hyperbranched Polyesters. *Molecular Weight Control and Chain End Functionalization*. *Polymer Journal*. 1994;26(2):187-97.
4. Herkert T, Prinz H, Kovar K. One hundred percent online identity check of pharmaceutical products by near-infrared spectroscopy on the packaging line. *Eur J Pharm Biopharm*. 2001;51(1):9-16.
5. *Infrared Spectroscopy for Process Analytical Applications*. *Process Analytical Technology*. p. 157-94.
6. Hertrampf A, Sousa RM, Menezes JC, Herdling T. Semi-quantitative prediction of a multiple API solid dosage form with a combination of vibrational spectroscopy methods. *J Pharm Biomed Anal*. 2016;124:246-53.
7. Ali SM, Bonnier F, Lambkin H, Flynn K, McDonagh V, Healy C, et al. A comparison of Raman, FTIR and ATR-FTIR micro spectroscopy for imaging human skin tissue sections. *Analytical Methods*. 2013;5(9):2281-91.
8. Wartewig S, Neubert RH. Pharmaceutical applications of Mid-IR and Raman spectroscopy. *Adv Drug Deliv Rev*. 2005;57(8):1144-70.
9. Ramer GaL, B. *Attenuated Total Reflection Fourier Transform Infrared Spectroscopy*. *Encyclopedia of Analytical Chemistry*: R. A. Meyers and R. A. Meyers 2013.
10. Palo M, Kogermann K, Genina N, Fors D, Peltonen J, Heinämäki J, et al. Quantification of caffeine and loperamide in printed formulations by infrared spectroscopy. *Journal of Drug Delivery Science and Technology*. 2016;34:60-70.
11. Bonnier F, Blasco H, Wasselet C, Brachet G, Respaud R, Carvalho LFCS, et al. Ultra-filtration of human serum for improved quantitative analysis of low molecular weight biomarkers using ATR-IR spectroscopy. *Analyst*. 2017;142(8):1285-98.
12. Ali SM, Bonnier F, Tfayli A, Lambkin H, Flynn K, McDonagh V, et al. Raman spectroscopic analysis of human skin tissue sections ex-vivo: evaluation of the effects of tissue processing and dewaxing. *J Biomed Opt*. 2013;18(6):061202.
13. Kalinkova GN. Infrared spectroscopy in pharmacy. *Vibrational Spectroscopy*. 1999;19(2):307-20.
14. Schulz H, Quilitzsch R, Krüger H. Rapid evaluation and quantitative analysis of thyme, origano and chamomile essential oils by ATR-IR and NIR spectroscopy. *Journal of Molecular Structure*. 2003;661-662:299-306.
15. Sandasi M, Kamatou GPP, Gavaghan C, Baranska M, Viljoen AM. A quality control method for geranium oil based on vibrational spectroscopy and chemometric data analysis. *Vibrational Spectroscopy*. 2011;57(2):242-7.
16. Tankeu S, Vermaak I, Kamatou G, Viljoen A. Vibrational Spectroscopy as a Rapid Quality Control Method for *Melaleuca alternifolia* Cheel (Tea Tree Oil). *Phytochemical Analysis*. 2014;25(1):81-8.
17. Miloudi L, Bonnier F, Bertrand D, Byrne HJ, Perse X, Chourpa I, et al. Quantitative analysis of curcumin-loaded alginate nanocarriers in hydrogels using Raman and attenuated total reflection infrared spectroscopy. *Anal Bioanal Chem*. 2017;409(19):4593-605.

18. Masmoudi H, Dreau YL, Piccerelle P, Kister J. The evaluation of cosmetic and pharmaceutical emulsions aging process using classical techniques and a new method: FTIR. *Int J Pharm.* 2005;289(1-2):117-31.
19. Mihranyan A, Ferraz N, Strømme M. Current status and future prospects of nanotechnology in cosmetics. *Progress in Materials Science.* 2012;57(5):875-910.
20. Muller RH, Petersen RD, Hommoss A, Pardeike J. Nanostructured lipid carriers (NLC) in cosmetic dermal products. *Adv Drug Deliv Rev.* 2007;59(6):522-30.
21. Muller RH, Radtke M, Wissing SA. Solid lipid nanoparticles (SLN) and nanostructured lipid carriers (NLC) in cosmetic and dermatological preparations. *Adv Drug Deliv Rev.* 2002;54 Suppl 1:S131-55.
22. Davidov-Pardo G, McClements DJ. Resveratrol encapsulation: Designing delivery systems to overcome solubility, stability and bioavailability issues. *Trends in Food Science & Technology.* 2014;38(2):88-103.
23. Nguyen HT, Munnier E, Souce M, Perse X, David S, Bonnier F, et al. Novel alginate-based nanocarriers as a strategy to include high concentrations of hydrophobic compounds in hydrogels for topical application. *Nanotechnology.* 2015;26(25):255101.
24. Nguyen HTP, Munnier E, Perse X, Vial F, Yvergnaux F, Perrier T, et al. Qualitative and Quantitative Study of the Potential of Lipid Nanocapsules of One Hundred Twenty Nanometers for the Topical Administration of Hydrophobic Molecules. *J Pharm Sci.* 2016;105(10):3191-8.
25. Nguyen HTP, Souce M, Perse X, Vial F, Perrier T, Yvergnaux F, et al. Lipid-based submicron capsules as a strategy to include high concentrations of a hydrophobic lightening agent in a hydrogel. *Int J Cosmet Sci.* 2017;39(4):450-6.
26. Roberts MS, Mohammed Y, Pastore MN, Namjoshi S, Yousef S, Alinaghi A, et al. Topical and cutaneous delivery using nanosystems. *J Control Release.* 2017;247:86-105.
27. Managuli RS, Kumar L, Chonkar AD, Shirodkar RK, Lewis S, Koteswara KB, et al. Development and Validation of a Stability-Indicating RP-HPLC Method by a Statistical Optimization Process for the Quantification of Asenapine Maleate in Lipidic Nanoformulations. *J Chromatogr Sci.* 2016;54(8):1290-300.
28. Woodhouse EJ, Breyman S. Green Chemistry as Social Movement? *Science, Technology, & Human Values.* 2005;30(2):199-222.
29. Miloudi L, Bonnier F, Barreau K, Bertrand D, Perse X, Yvergnaux F, et al. ATR-IR coupled to partial least squares regression (PLSR) for monitoring an encapsulated active molecule in complex semi-solid formulations. *Analyst.* 2018;143(10):2377-89.
30. Gerretzen J, Szymanska E, Bart J, Davies AN, van Manen HJ, van den Heuvel ER, et al. Boosting model performance and interpretation by entangling preprocessing selection and variable selection. *Anal Chim Acta.* 2016;938:44-52.
31. Heraud P, Wood BR, Beardall J, McNaughton D. Effects of pre-processing of Raman spectra on in vivo classification of nutrient status of microalgal cells. *Journal of Chemometrics.* 2006;20(5):193-7.
32. Wold S, Sjöström M, Eriksson L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems.* 2001;58(2):109-30.
33. Byrne HJ, Knief P, Keating ME, Bonnier F. Spectral pre and post processing for infrared and Raman spectroscopy of biological tissues and cells. *Chem Soc Rev.* 2016;45(7):1865-78.
34. Butler HJ, Ashton L, Bird B, Cinque G, Curtis K, Dorney J, et al. Using Raman spectroscopy to characterize biological materials. *Nat Protoc.* 2016;11(4):664-87.
35. Afseth NK, Kohler A. Extended multiplicative signal correction in vibrational spectroscopy, a tutorial. *Chemometrics and Intelligent Laboratory Systems.* 2012;117:92-9.

36. Wold S, Ruhe A, Wold H, W. J. Dunn I. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J Sci Stat Comput.* 1984;5(3):735-43.
37. Bonnier F, Byrne HJ. Understanding the molecular information contained in principal component analysis of vibrational spectra of biological systems. *Analyst.* 2012;137(2):322-32.
38. Hibbert David B. Vocabulary of concepts and terms in chemometrics (IUPAC Recommendations 2016). *Pure and Applied Chemistry*2016. p. 407.
39. Allegrini F, Olivieri AC. IUPAC-consistent approach to the limit of detection in partial least-squares calibration. *Anal Chem.* 2014;86(15):7858-66.
40. Lee KY, Mooney DJ. Alginate: properties and biomedical applications. *Prog Polym Sci.* 2012;37(1):106-26.