2018

ATR-IR Coupled to Partial Least Square Regression (PLSR) for Monitoring an Encapsulated Active Molecule in Complex Semi-Solid Formulation

Lynda Miloudia
Université François-Rabelais de Tours, France

Franck Bonnier
Technological University Dublin, Franck.Bonnier@tudublin.ie

Kevin Barreaua,
Université François-Rabelais de Tours, France

See next page for additional authors

Recommended Citation

Miloudi, L., Bonnier, F. & Barreau, K. (2018). ATR-IR Coupled to Partial Least Square Regression (PLSR) for Monitoring an Encapsulated Active Molecule in Complex Semi-Solid Formulation. *The Analyst*, vol. 143, no. 4. 10.1039/C8AN00547H

ATR-IR coupled to Partial Least Square Regression (PLSR) for monitoring an encapsulated active molecule in complex semi-solid formulation

Lynda Miloudi[a], Franck Bonnier*[, a], Kevin Barreau[a], Dominique Bertrand[b], Xavier Perse[a], Florent Yvergnaux[c], Hugh J Byrne[d], Igor Chourpa[a], Emilie Munnier[a]

[a]Université François-Rabelais de Tours, EA 6295 Nanomédicaments et Nanosondes, 31 avenue Monge, 37200 Tours, France

[b]Data_Frame, 25 rue Stendhal, 44300 Nantes, France

[c] Bioeurope - Solabia Group, Anet, France

[d] FOCAS Research Institute, Dublin Institute of Technology, 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, Partial Least Squares Regression, Hydrogels, label free quantification

Abstract:

Attenuated Total Reflectance – Infrared (ATR-IR) spectroscopy holds great promise for industrial applications as a quality control tool for complex galenic formulations. Although the technique is often promoted for the molecular information it delivers in a label free and cost effective fashion, other advantages can emerge compared to gold standard analytical tools such as liquid chromatography coupled to mass spectrometry. The present study demonstrates how ATR-IR measurements enable accurate quantitative analysis of an Active Cosmetic Ingredient such as Omegalight® encapsulated in a complex alginate based nano-capsule. The study demonstrates how precise concentrations can be obtained without the requirement for fastidious extraction and separation protocols prior to ATR-IR analysis. However, the data mining remains a crucial aspect with particular emphasis on the preprocessing of the data that will be subjected to Partial Least Square Regression (PLSR) analysis. Therefore, different pre-processing methods have been evaluated to investigate the relationship between corrections applied and PLSR outcomes (i.e. precision, ratio of performance to deviation (RPD) and accuracy of the analysis). Ultimately, it has been found that, against all expectations, some of preprocessing methods do not necessarily lead to improvements in the end result, while Extended Multiplicative Scattering Correction (EMSC) is the only one which delivers satisfying results, as defined by a Root Mean Square Error (RMSEV) of 0.07 % (w/w) and RPD greater than 30 when performing analysis over the range 0.4-8.2% (w/w). Despite presence of large amounts of additives such as glycerol and preservatives in the formulation, implementing Leave One Out Cross Validation (LOOCV), further validate the method with RPD of 18 and relative errors for the predicted concentrations below the 5% (w/w) threshold, hence demonstrating that ATR-IR has analytical capabilities for applications in the cosmetic industry.

1. Introduction

Recent evolutions in formulation strategies have seen increasing numbers of vectorisation or encapsulation approaches for Active Cosmetic Ingredients (ACI) or Active Pharmaceutical Ingredient (API) administration. In in products intended for the cutaneous route, the use of nanocarriers (NC) can be either motivated by increased penetration of the ACI through the *stratum corneum* leading to improved efficacy [1–3], enhanced stability in formulas [4,5], or protection against exogenous agents such as light [6]. Core-shell NC are the most commonly documented for dermatological application [3,7] with systems as lipid nanocarriers already used in commercialised cosmetic products[8,9]. Alginate is natural polysaccharide extracted from seaweeds has recently been proposed as ingredient to developed nano-systems, named ANC (Alginate Nano-Carriers), for active cosmetic ingredients encapuslation.[5,10] ANC present great interest in the cosmetic industry and recent optimization in preparation protocols are encouraging for fast implementation in final products in the near future [11]. In general, cosmetic products in their final form are usually presented in a semi-solid form (cream or gel), itself supplemented with additives like antimicrobial ingredients, dyes and moisturizing agents [12,13] to ensure comfort when applied in the skin, and limiting spreading from the site of application [14].

Quantification of the ACI loaded in nanocarriers in those complex mixtures can be achieved by a combination of chemical extraction protocols followed by separation analytical techniques such as High Performance Liquid Chromatography (HPLC) coupled to mass spectrometry [5,15]. However, this method remains fastidious, expensive, uses large volumes of solvent and consequently generates considerable amounts of chemical waste which is in conflict with current concerns aiming to develop green chemistry alternatives [16].

Fourier Transform Infrared (FTIR) spectroscopy is a non-destructive and label free alternative with low instrumental requirements thus generally lower cost [17]. The main advantage remains the minimal sample preparation and lack of requirement for separation steps prior to analysis [18], in contrast to other gold standards such as HPLC. The specific spectral signatures provide information on molecular vibrations that can be used for chemical characterisation of ACIs or APIs[19]. The wealth of information has been exploited for quality control and quantitative analysis of ACIs such as essential oils [20] and other plants extracts[21,22]. Although monitoring of ACIs can be achieved when in semi-solid forms such as gels[23], somehow translation to industrial applications remain limited and only few referenced studies relate to quantitative analysis in cosmetic products[24]. However, promising use of infrared spectroscopy for study of formulation stability have been reported suggesting an increasing popularity of the technique in the cosmetic field[25]. However, there is no evidence in the literature that once the ACI has been loaded in a nanocarriers such as ANC and dispersed in a complex formulation IR analysis

will perform similarly and be able to detect and quantify the molecule of interest. Therefore, present study aims to further investigate the potential of FTIR spectroscopy for the rapid quantification of ACI loaded in ANC and dispersed in a complex cosmetic product-like models.

While Infrared imaging has been reported as suitable tool for tablets or other solid forms characterisation in order to study distribution of API [26], when aiming for quantitative analysis from liquid or viscous samples other instrumental set up such as Attenuated Total Reflectance (ATR) should be preferred [23]. Moreover, deposition of a sample directly onto the ATR crystal enable to collect IR spectra free from abnormal background distortion due to Mie-scattering effect while preserving the experimental conditions for direct application of the Beer Lambert law [27].

In such conditions, the IR spectra collected from complex cosmetic mixtures reflect specifically the chemical composition including contributions of the different ingredients to the spectral features, directly correlated to the their concentration in the mixture [23]. In the present study, Omegalight® (Bioeurope, France), a commercialised lightening agent with melanocytes as cellular targets [11], has been selected. Current strategies aim to encapsulated Omegalight® (OL) in ANC to increase its penetration through the skin in order to reach the dermo-epidermal junction. Despite, at present commercialised cosmetic products containing Omegalight® don't incorporate ANC in their formulations yet, the present work presently described is based on samples exactly mirroring the final form. Carboxymethyl-cellulose (CMC), humectant as glycerol and a preservative mixture of dehydroacetic acid and benzyl alcohol (Cosgard®), are the most common ingredients found in gel-based formulations, have been logically added to the preparation at relevant concentrations in order to perform the IR measurements and address current challenges for the application of FTIR on future commercialised products.

Partial Least Squares Regression (PLSR) has been used on the infrared spectral data sets to construct predictive model to assess the precision and accuracy of the analysis.

Recent improvements in experimental setups and data handling including data pre-processing [28,29] and multivariate statistical analysis [30] have demonstrated the importance of taking into account physical and optical phenomena. Consequently, a comparison of different pre-processing strategies has been conducted, aiming to identify the most suitable method to deliver best precision and accuracy. Pre-processing steps are used to improve the subsequent multivariate regression, classification model or exploratory analysis [31,32].Ultimately, the capabilities of FTIR spectroscopy as analytical tool for potential online screening and monitoring of encapsulated active cosmetic molecules during the production process industrial environment has been highlighted and discussed.

2. Materials and methods

2.1. 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 a model of cosmetic gels, current 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 to the nanocarrier shell.

2.2 Preparation and characterization of Omegalight®-loaded alginate-based nanocarriers (ANC_OL)

2.2.1 Preparation of ANC_OL

Alginate-based nanocarriers are composed of an oily core, surrounded by an alginate shell. The particularity of the system used in this study is that the oily core is the ACI itself. ANC_OL were prepared following the method of oil-in-water emulsification and ionic gelation, described in detail by Nguyen et al. [5]. Briefly, sodium alginate is placed soaking 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 (0.06 g/L) to form the aqueous phase of the emulsion. Secondly, the oil phase was prepared by mixing Omegalight® (0.16 g/g) and sorbitan monooleate (0.01g/L). Finally, the nano-emulsification is achieved by mixing the two phases under sonication (Vibra-cell ultrasonic processor, Sonics, 20 kHz) during 3 minutes. The gelation of the surface of the nanocarriers is achieved by addition of a solution of calcium ions (0.6 g/L). The final concentration of the ANC_OL suspension is 0.21 g/g.

2.2.2. Physico-chemical characterization of ANC_OL suspensions

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 633 nm

laser source and detection angle set at 13°. All measurements have been performed in triplicate and at 25°C. The ANC_OL display an average hydrodynamic diameter of 207 ± 7 nm with a polydispersity index below 0.2, indicating a narrow distribution. The ζ value of ANC_OL was −23 ± 0.4 mV. The physico-chemical properties are in accordance with those observed in previous studies [5] and comply with use of nano-systems in skin cosmetic products.

2.2.3 Preparation of Cosmetic product-like models

The spectroscopic analysis described in the present study has been conducted on 3 gel-based models with increasing complexity. All samples have been prepared to mimic cosmetic products in their final form in order to evaluate the capabilities of ATR-IR spectroscopy in real industrial conditions. For instance, all the samples contain CMC, which is the ingredient providing the gel aspect to product (i.e. hydrogel). For all samples, the final concentration in CMC is 1.5% (w/w). Model_1 was prepared by mixing a CMC-based hydrogel with increasing amounts of ANC_OL suspension (prepared as described above) in order to build a range of 8 different concentrations (see table1). In addition to CMC, cosmetic products usually also contain additives such as humectant and preservatives which are mixed to the hydrogel during preparation. Therefore, the second model aims to investigate how the presence of additives can affect the outcome of the quantitative analysis performed with PLSR. Thus, in Model_2, Cosgard® and glycerol were added to samples at final concentrations of respectively 1% (w/w) and 20% (w/w), which represent the most commonly used concentrations in commercialised cosmetic products. Finally, Model_3 has been prepared to further support the specificity of the analysis performed with ATR-IR spectroscopy. Due to its high concentration and strong IR absorbance, glycerol is expected to be the most strongly interfering ingredient, possibly partially swamping the contribution of Omegalight® in the IR spectra collected. Importantly, in Model_3, extreme conditions have been used, whereby glycerol, at various concentrations for each sample, is randomly added to ensure no correlation can be observed between glycerol and Omegalight® concentrations. For clarity, the list of samples and corresponding concentrations are summarised in table 1. Due to the viscosity of few ingredients, the concentrations provided in the table reflect the exact weight introduced during preparation, aiming to be as closed as possible to targeted values. Also not mentioned in the table, and as a reminder, all samples are hydrogels containing 1.5 % of CMC.

Table 1: List of samples for the 3 gel-based models with corresponding concentrations of Omegalight® (OL), glycerol and Cosgard®. (All concentrations are expressed as % w/w).

		Sample N°							
		1	2	3	4	5	6	7	8
Model_1	[OL]	0,42	0,82	1,63	2,45	3,29	4,10	6,15	8,20
	[Glycerol]	0	0	0	0	0	0	0	0
	[Cosgard®]	0	0	0	0	0	0	0	0
Model_2	[OL]	0,41	0,82	1,64	2,46	3,28	4,09	6,14	8,19
	[Glycerol]	20	20	20	20	20	20	20	20
	[Cosgard®]	1	1	1	1	1	1	1	1
Model_3	[OL]	0,41	0,82	1,64	2,46	3,27	4,10	6,14	8,20
	[Glycerol]	12,5	2,5	0	15	5	10	1	17,5
	[Cosgard®]	1	1	1	1	1	1	1	1

2.3 ATR-FTIR data collection analysis data handling

2.3.1. ATR-FTIR data collection

IR spectra were acquired using a Bruker Vector 22 FT-IR spectrometer (Bruker, Germany) equipped with a Golden Gate single reflection diamond attenuated total reflectance (ATR) accessory (Specac). 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 liquid but with higher viscosity therefore thus 200 µL were deposited directly onto the diamond surface. Prior to sample measurement, a background spectrum was recorded in air (64 scans) and automatically rationed with the sample spectrum (16 averaged scans) by the software. For each sample, 5 deposits have been measured and 3 spectra per drop have been collected. Hydrogels samples behave similarly to liquid therefore no pressure is 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.

2.3.2 Data Handling

The data pre-processing and analyses were performed using Matlab (Mathworks, USA). Partial Least Squares Regression (PLSR) is probably the most used methods to extract quantitative information from spectral data sets. However, the approach remains strongly influenced by any interferences (instrumental response, artefact, sample heterogeneity) possibly affecting the overall quality of the data collected. Consequently, data pre-processing

holds is critical in ensuring the best outcome from the data analysis which, in the case of PLSR, is increased precision and accuracy [33]. Therefore, in the present study, different pre-processing methods have been used including baseline correction with rubber band model [29,34–37], min-max normalisation (MMN) applied to dominant bands at 3315 cm^{-1} or 1635 cm^{-1} [23], vector normalisation (VN) [38,39], Standard Normal Variate (SVN) [32,40] or Extended Multiplicative Signal Correction (EMSC)[41]. The latter has been applied using the EMSC toolbox for Matlab freely available from Nofima Data Modelling (http://nofimamodeling.org/people/). The toolbox enables data uploading and processing in a user friendly interface. A fifth order polynomial EMSC has been preferred and found to be optimal. Full details about the correction and the use of the interface can be found in the published tutorial proposed by Asfeth and Kohler [41]. The impact of the different methods on the PLSR analysis have been compared in order to evaluate the best capabilities of ATR-IR spectroscopy as tool for Omegalight® monitoring in nano-encapsulated based cosmetic products.

PLSR analysis is a robust and well recognised quantitative approach extensively referenced in the literature [42]. The statistical robustness has been evaluated by implementing a 100 fold - cross validation in the routine with 50 % of the data set selected randomly as calibration, while the remaining 50 % are also selected randomly as validation. The output gives an estimation of the model robustness in terms of precision (Root Mean Square Error – RMSE), ratio of performance to deviation (RPD), linearity between the experimental and predicted concentrations (R^2) and accuracy (relative error of the predictive concentration compared to the true value). The RPD informs about the overall prediction capacity of a model is determined as follow[43,44]:

$$RPD = \frac{SD}{RMSEV}$$

Equation (1):

With SD as the standard deviation of the references concentrations and RMSEV as the Root Mean Square Error Validation calculated from the PLSR analysis. Higher RPD values indicates an excellent predictive ability and vice versa. For example, values greater than 3 are useful for screening, values greater than five can be used for quality control, and values greater than eight for any application[45].

Additionally, the data have been also processed using Leave One Out Cross Validation (LOOCV) to illustrate how the PLSR analysis behaves when presented with unknown samples. According to this approach, 7 concentrations are used as calibration set and the 8^{th} is only input at the validation step. This ensures that the

tested sample does not influence the calibration of the regression analysis in any manner. This analysis has been repeated 8 times, each time taking a different concentration as the unknown sample.

3. Results and discussions

3.1 Omegalight®-loaded alginate-based nanocarriers (ANC_OL)

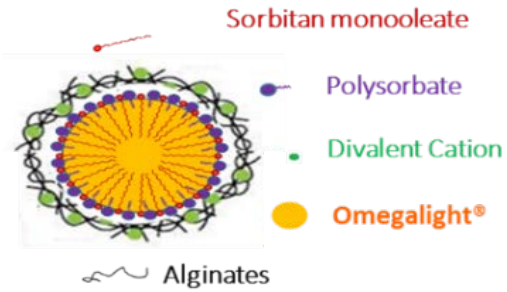

Figure 1: *Illustration of Omegalight® loaded alginate-based nanocarriers (ANC_OL)*

Encapsulation of ACI in core–shell NC is particularly interesting for the inclusion of strongly hydrophobic molecules in gel-based cosmetic end products. For instance, Omegalight® cannot be directly solubilised in a hydrophilic matrix such as a CMC gel. However, once loaded in ANC, the surface property conferred by the alginate based shell is compatible with preparation of homogenous suspensions. Although the alginate (a natural biodegradable polysaccharide extracted from brown algae) can be commonly used to form NC shells [46], the originality of the ANC_OL prepared in this study resides in the 100% organic core composed of the ACI without any dilution in a neutral oil (figure 1). One should keep in mind that, in the final product, typical concentrations of 1%-2% (w/w) of Omegalight® are recommended by the provider. Considering the encapsulating protocol used in the present study, such concentrations are reached when only 6%-12% (w/w) of the ANC_OL suspension is added to the formulation. Consequently, roughly 90% of the sample is composed of the gelling agent (CMC), water and additional additives (glycerol and Cosgard®). This illustrates perfectly the challenge when performing IR analysis to detect the ACI in this molecular soup and achieving quantitative analysis without interferences from other ingredients (of the NC and/or hydrogel).

3.2 Quantitative analysis of Omegalight® in the Model_1

3.2.1 Examination of ATR-IR spectra

IR spectroscopy is a powerful characterization technique, although, as a large majority of organic molecules exhibit strong bands in the spectra collected, it can be difficult to interpret the signatures collected from complex mixtures. In order to better understand the origin of the features observed in the IR spectra, initial work has been conducted on the Model_1 (See material and methods), a simplified system to which the additives (glycerol and Cosgard®) have not been added. Consequently, the possible interference from the ANC shell ingredients and the gelling agent itself can be better evaluated. Although water has a strong absorbance in the mid-IR, it has been demonstrated that working on solutions is achievable if the concentrations involved are sufficiently high to overcome the broad contribution of the water bands located at 3302 cm^{-1} (ν OH mode) and 1637 cm^{-1} (δ OH mode) [17]. The spectrum collected from sample 8 (highest Omegalight® concentration) suggests that the main contribution observed originates from the water contained in the sample (Figure 2A and B). Interestingly, apart from the water bands, it seems no further features from the CMC can be visualized indicating that the concentration is too low (1.5 % w/w) to contribute to the spectra collected (Figure 2B). Omegalight®, Alginate, sorbitan monooleate and polysorbate 80 are characterized by the presence of numerous relatively sharps bands across the high wavenumber range, 3000-2700 cm^{-1}, the 1800-900 cm^{-1} spectral window providing more specific signatures (Figure 2 C-F).

To ease the comparison, table 2 summarizes the different band positions as found in the spectra with corresponding assignments. While the high wavenumber region does not appear to be specific, as possible matches are found in all ingredients, the fingerprint region allows better evaluation of the specific contributions. 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 1491 cm^{-1}, unambiguously identify a strong contribution of the Omegalight® in the spectra collected from sample 8. The bands at 1041 cm^{-1} (ν (C-O), ν (C-C)), 1377 cm^{-1}(δ C-C, δ C-C-H) and 1464 cm^{-1} (δ (CH_3), δ (CH_2) δ (C-C)) are shared with the sorbitan monooleate, but subtle contributions from this ANC shell ingredient are suggested.

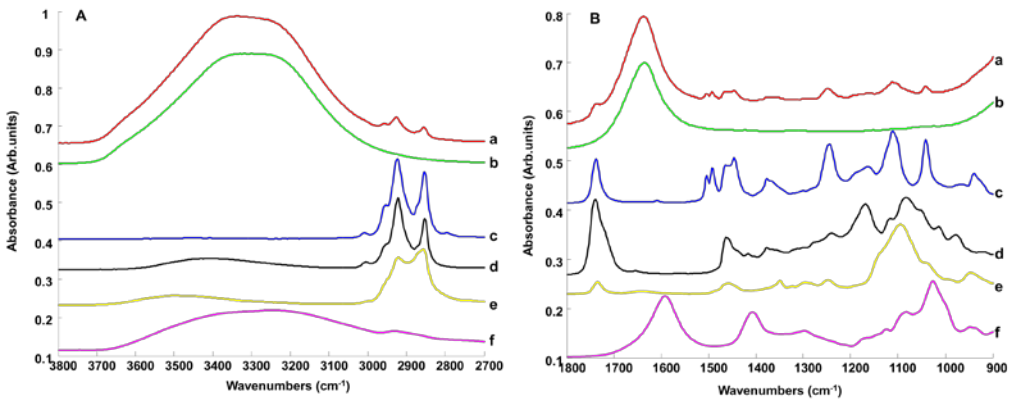

Figure 2: Mean Infrared spectrum collected from Sample 8 (a) compared to spectra collected from CMC hydrogels (b), Omegalight® (c), Sorbitan monooleate (d), Polysorbate 80 (e) and Alginates (f). High wavenumber (A) and fingerprint region (B) have been plotted separately. Spectra are offset for clarity

Table 2: Band positions and assignments of Omegalight®-loaded alginate-based nanocarriers (ANC_OL), Omegalight® (OL), Polysorbate 80, sorbitan monooleate and alginates [23,47]

Wavenumbers (cm^{-1})					Assignments
ANC_OL	OL	Polysorbate 80	Sorbitan monoleate	Alginates	
_	_	946	_	948	ν (C-C)
_	_	_	977	_	ν (C-C)
_	937	_	_	_	ν (C-O)
_	_	_	_	1026	ν (C-O) , ν (C-C), δ (COH)
1041	1041	_	1041	_	ν (C-O) , ν (C-C)
_	_	_	1082	_	ν (C-O)
_	_	_	_	1084	ν(C-O), δ (CCO), δ (C-C)
_	_	1093	_	_	ν (C-O)
1111	1111	_	_	_	ν (C-O)
_	_	_	1117	_	ν (C-O)
_	_	_	_	1120	ν (C-O) , ν_{sym} (C-C)
1163	1163	_	_	_	ν (C-O)
_	_	_	1169	_	ν (C-O)
_	_	_	1238	_	ν (C-O)
_	1244	_	_	_	ν (C-O)
1250	_	1248	_	_	ν (C-O)
_	_	1296	_	1298	δ(OH), δ(CH), τ(CH), *w* (CH)
_	_	1350	_	_	δ (OH), δ (CH3)
1377	1377	_	1377	_	δ C-C-C, δ C-C-H
_	_	_	_	1408	ν_{asym} (COO-)
1446	1446	_	_	_	ν (C=C) (aromatic ring), δ (CH_2)
1464	1464	_	1464	_	δ (CH_3), δ (CH_2) δ (C-C)
1491	1491	_	_	_	ν (C-C)
1504	1504	_	_	_	ν (C=C) (aromatic ring)
_	_	_	_	1593	ν_{sym} (COO-)
1736	1738	1734	1739	_	ν (C=O)
2854	2854	2856	2852	_	ν (CH_2)
2924	2924	2922	2922	2922	ν_{asym} (CH_2)
2956	2954	_	2956	_	ν_{asym} (CH_3)
-	-	-	3006	_	ν (C=CH)

Figure 3 displays the set of mean spectra collected from the Model_1 samples prepared as described in Material and methods (2.1.3). Although the final product concentrations for commercialized Omegalight® product are generally between 1% and 2% (w/w), the range has been extended between 0.42% (w/w) to 8.2% (w/w) in order to evaluate the quantitative capabilities of the ATR-IR with highly and lowly concentrated samples. The features being quite weak compared to the strong contribution of the water bands, the spectra have been represented separately in the high wavenumber and finger print regions. As expected, the bands located at 2956 cm^{-1}, 2924

cm^{-1}, 2854 cm^{-1}, 1736 cm^{-1}, 1504 cm^{-1}, 1491 cm^{-1}, 1464 cm^{-1}, 1446 cm^{-1}, and 1377 cm^{-1}, 1163 cm^{-1}, 1111 cm^{-1}, and 1041 cm^{-1} have decreasing intensities, correlated with the decrease in Omegalight® content in the samples.

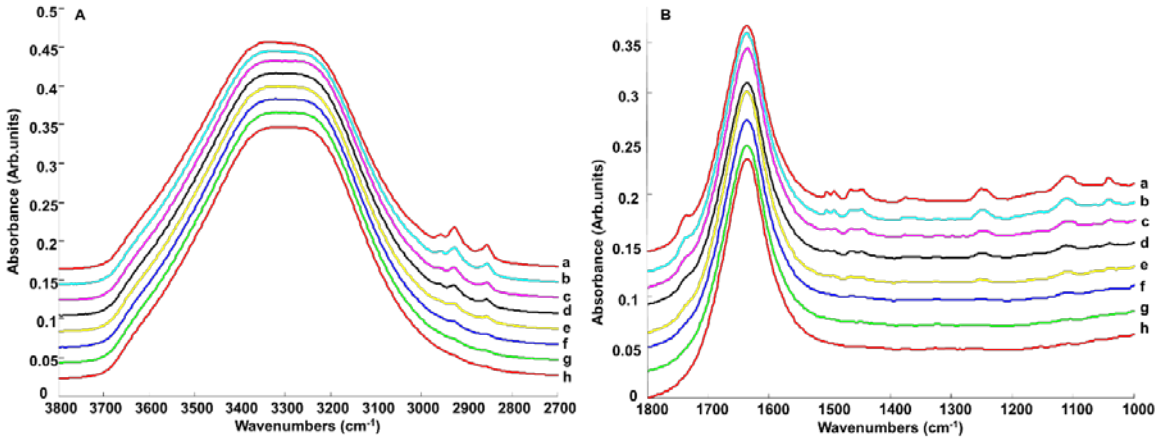

*Figure 3 Mean ATR-FTIR spectra recorded from samples from Model_1 in the high wavenumber region (A) and fingerprint region (B). Spectra are organized according to their concentrations as follow: 0.42% **(a)**, 0.82% **(b)**, 1.63 % **(c)**, 2.45 % **(d)**, 3.29 % **(e)**, 4.10 % **(f)**, 6.15 % **(g)**, and 8.20 % **(h)**. Percentages are expressed in w/w. Spectra are offset for clarity*

3.2.2 Regression results

Examination of data and comparison of mean spectra allows characterization of the different contributions observed, but, in order to evaluate the linear relationship existing between the intensity of the Omegalight® features and the concentrations in the sample, advanced multivariate methods such as PLSR have been tested [48]. In order to evaluate the effect of the preprocessing methods on the PLSR outcome, a general procedure has been established and implemented similarly for all different conditions. As an illustrative example of the PLSR process, Figure 4 presents the different steps using the raw data collected from Model_1. Cross validation is essential to ensure the analysis is not biased, but also to test the robustness of the predictive model used [49]. As a reminder, in the present study a 100-fold cross validation has been employed with 50% spectra used as calibration and the remaining 50% as validation. The calibration and validation sets being independent the statistical relevancy is preserved through the process. The calibration step usually results in a graph representing the Root Mean Square Error Calibration (RMSEC) according to the number of dimension used to regress the data (Figure 4A). As a supervised method, the PLSR algorithm will optimize the model to deliver the lowest RMSEC, leading to a gradual decrease towards 0. However, to ensure the highest predictability of the PLSR model while avoiding over fitting of the data, the number of dimensions should be very carefully selected. For this reason, the remaining data corresponding to the validation set are input to the calibration model in order to

evaluate the RMSEV (validation). As seen in Figure 4B, the curve exhibits a different pattern, with first a steep decrease before stabilizing around 0.2% (w/w) Omegalight®. This indicates that all dimensions calculated from the calibration set are not relevant to the quantification and that, above 3 dimensions, spectral variations captured are no longer correlated to concentrations of Omegalight®. With a RMSEV of 0.1873% (w/w), the number of 3 dimensions has been selected from the subsequent steps of the analysis.

The predictive model is constructed by regression of the observed concentration (true concentrations) against the predicted concentrations (experimental concentrations), as illustrated in figure 4C. This representation allows to graphically observe the linearity of the regression model, which is further supported by the R^2 values, 0.9949 for the present case. Considering the 100-fold cross validation, a very large number of spectra are used for each sample. Thus it has been preferred to use the mean predicted concentrations with corresponding standard deviations to reduce the number of objects plotted in the graph.

Finally, it is essential to validate the molecular selectivity of the analysis performed by means of PLSR. To this end, the weighting vector can be visualized in order to highlight the wavenumbers with most contribution to the regression model. A comparison of the weighting vector (Figure 4.D.a) with the spectrum of Omegalight® (Figure 4.D.b) confirms the analysis is only based on the ACI features without any interference from the other ingredients due to positive bands at 2924 cm^{-1}, 2854 cm^{-1}, 1739 cm^{-1}, 1504 cm^{-1}, 1377 cm^{-1}, 1252 cm^{-1}, 1111 cm^{-1} and 1041 cm^{-1}. Sorbitan monooleate, polysorbate 80 and alginate have specific IR signatures and none of the bands listed in table 1 can be found in the weighting vector. Although Sorbitan monooleate has shared vibrational modes at 1041 cm^{-1}, 1377 cm^{-1} and 1464 cm^{-1}, the absence of a peak or deformation at 1050 cm^{-1} suggests its contribution is negligible. Interestingly, the 2 negative bands are assigned to H_2O vibrations which are logically anti-correlated to increase in Omegalight® due to proportional reduction in water content.

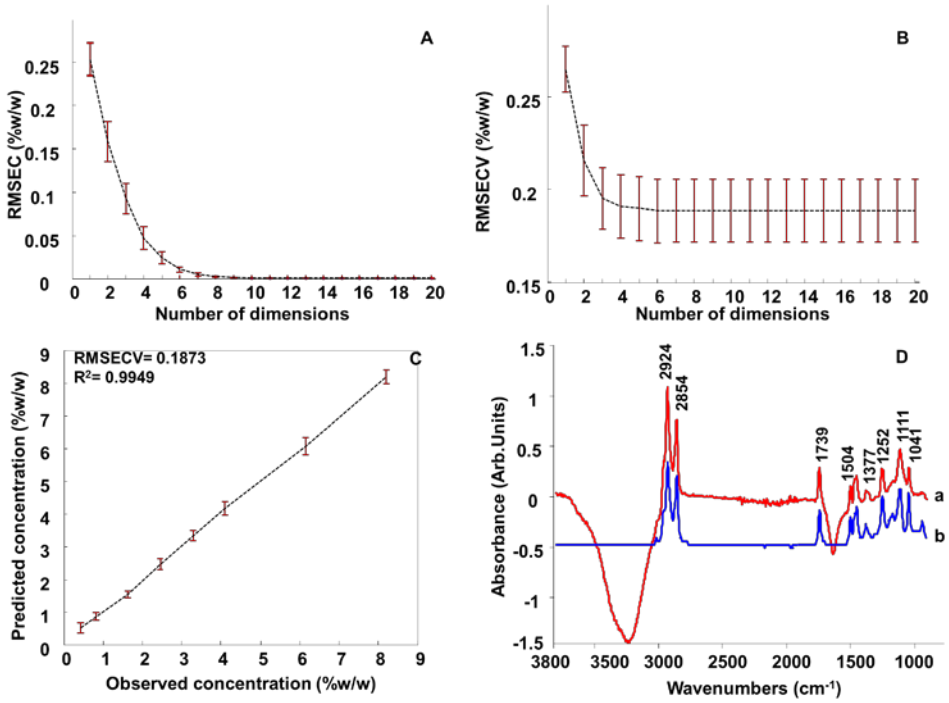

Figure 4: PLSR analysis performed on Model_1 spectra. A: Estimation of the Root Means Square Error Calibration (RMSEC), B: Selection of the number of dimension based on the Root Mean Square Error Validation (RMSEV), C: Regression model constructed using 3 dimensions and D: Weighting vector (a) compared to the spectrum of Omegalight® (b) *(Spectra are offset for clarity).*

The PLSR routine described has been implemented on data sets preprocessed using all different methods, in order to compare the respective analysis performance in terms of precision and accuracy. It has been reported that the selection of an appropriate pre-processing method can greatly improve the prediction ability of the model [28,31,50]. Table 3 gathers the RMSEV and the RPD values calculated after subjecting the spectra to Min-Max normalization, SNV, rubber band followed by vector normalization and EMSC.

Although, the spectra collected from the samples exhibit numerous bands on the spectral range 3800-1000 cm^{-1}, it has been highlighted in table 3 that the finger print region (1800-1000 cm^{-1}) remains the most specific, while the vibrational modes observed in the high wavenumbers (3800-2700 cm^{-1}) seem to be either found in all ingredients or strongly overlap, limiting specific identification of compounds. Consequently, the PLSR analysis has also been performed either on the full range or restricted to the finger print region with corresponding predictive performance evaluated on the basis of the RMSEV and the RPD. Interestingly, the number of optimal dimensions is not affected by the pre-processing method used (data not shown) and the number of 3 dimensions was found optimal for the conditions tested. Considering the experimental set up coupling ATR-IR measurements with liquid samples, it is expected the raw data to already deliver a high precision due to direct

link between the absorbance and concentrations.[51] Moreover, the PLSR method aims to reduce the dimensionality of the data sets to only use the most relevant features for the construction of the predictive models. Although the use of preprocessing methods is a well-established practice, the comparison of the RMSEV and the RPD with the data subjected to MMN, SNV and rubber band coupled to vector normalization suggest that performing the PLSR on the raw data leads to the best outcome. At best, the MMN on the water band at 1635 cm^{-1} produces RMSEV and RPD of respectively 0.189% (w/w) and 24.4 which remains slightly higher than the 0.1683% (w/w) and 15.99 obtained when performing the analysis on the finger print of raw spectra. MMN, RB&VN and SNV appear to all perform similarly, resulting in RMSEV around 0.19% (w/w) with RPD values of about 14.

The EMSC is the only approach which delivers improved predictive ability compared to the raw data. With respective RMSEV which is less than 0.07 % (w/w) and RPD higher than 30 for both full range and finger print region, demonstrating an improvement of a factor 2 to 3 can be achieved.

Table 3 RMSEV obtained with the different pre-processing (MMN: Min Max Normalisation, SNV: Standard Normal Variate, RB&VN: Rubberband & vector normalization and a fifth order polynomial EMSC model conducted on the ATR-FTIR spectra of Model_1. Results are expressed as % (w/w) Omegalight®.

	Wavenumbers (cm^{-1})					
	3800-1000			1800-1000		
Pre-processing	RMSEV (% w/w)	RPD	R^2	RMSEV (% w/w)	RPD	R^2
Raw data	0.1873	14.37	0,9949	0.1683	15.99	0,9959
MMN	0,2000	13.45	0,9940	0,1890	14.24	0,9949
SNV	0,1944	13.84	0,9945	0,1930	13.94	0,9947
RB & VN	0,2140	12.57	0,9933	0,1910	14.09	0,9948
EMSC	0.0727	37.01	0.9992	0.0752	35.78	0.9992

While the RMSEV and RPD informs about the overall prediction capacity of the model constructed from the PLSR, the accuracy of the analysis is rather defined by the difference between the true and experimental concentrations. Table 4 gives, for each sample, the mean concentrations, and respective standard deviations, which have been predicted from the full range analysis performed on EMSC, corrected spectra. The last column of table 4 is the relative error expressed as % and can be considered as the best indicator of the model accuracy. Overall, the results are satisfactory, most of the errors in predicted concentrations being below (or close to) the 5% mark. Only sample 1 exhibits a large difference between the true and experimental concentrations, of about

20%. This observation was somehow expected, due to the strong water contribution swamping the features originating from the Omegalight®, which is even more pronounced at decreased concentrations.

Table 4 Predicted concentrations of Omegalight®-loaded alginate-based nanocarriers (ANC_OL) obtained by partial least squares regression (PLS-R) performed on ATR-FTIR spectra corrected by a fifth order polynomial EMSC (3800-1000 cm^{-1}).

Samples	Observed concentration %(w/w)	Predicted concentration % (w/w)	SD %(w/w)	Relative error (%)
1	0,42	0,52	0,05	23,19
2	0,82	0,81	0,03	1,44
3	1,63	1,52	0,04	7,06
4	2,45	2,47	0,03	0,69
5	3,29	3,30	0,03	0,38
6	4,10	4,11	0,03	0,15
7	6,15	6,09	0,04	1,01
8	8,20	8,25	0,06	0,66

3.3 Quantitative analysis of Omegalight® in end product-like models

3.3.1 Examination of ATR-IR spectra

After optimization of the PLSR protocol on a simplistic model, the second step of the study was focused on evaluating the potential of ATR-IR to analyse samples mirroring real cosmetic products. Although the water could be expected to be the main obstacle to overcome, the presence of additives in the formulation adds another level of complexity to the quantitative analysis. For instance, glycerol is characterised by strong IR features [52] . Generally, found at final concentrations in cosmetic products around 20% (w/w), the influence on the IR signatures is obvious, as illustrated in figure 5. Although, the high wavenumber region does not seem to be affected (figure 5B), the fingerprint region now exhibits dominant features at 1111 cm^{-1}, 1041 cm^{-1} and 993 cm^{-1}, specific to this polyol compound. In contrast, although Cosgard® also exhibits numerous specific features in the fingerprint region (Figure 5A.c), its contribution to the IR spectra of the formulations is weak. The absence of obvious bands around 1010 cm^{-1} and 1460 cm^{-1} (figure 5A. a) would even suggest a negligible contribution, probably due to its low final concentration (i.e. 1% w/w)

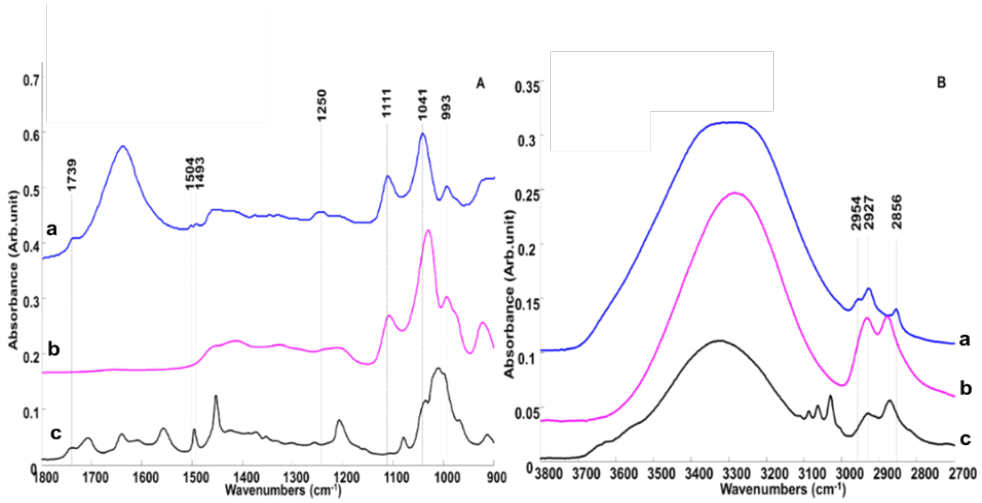

Figure 5 Mean Infrared spectra collected from sample 8 Model_2 (a), glycerol (b) and Cosgard® (c) obtained over finger print region (A) and the high wavelengths range (B). Spectra are offset for clarity

3.3.2 Regression results

Similar to the analysis of Model_1, the data have been subjected to PLSR analysis before and after different preprocessing methods to illustrate how the predictive analysis performs. Model_2 has only one variable changing between samples, the concentration of Omegalight®, while glycerol and Cosgard® are kept at constant levels (see material and methods – table 1). The results gathered in table 5 suggest the presence of additives in the formulation strongly alters the precision of the analysis when raw, MMN, RB&VN and SVN data are considered, while the outcome for EMSC corrected spectra appears to be affected to a lesser extent. However, independent of the pre-processing method, the predictive models are now clearly better while focusing on the finger print region. EMSC remains the most promising approach, having a RMSEV of 0.0923% (w/w) a RPD higher than 30. A third model has been prepared with glycerol concentrations randomly distributed between 0% and 20%. Although Model_3 is an exaggeration of the real life conditions, the variations induced in glycerol concentrations were required to demonstrate that there was no correlation between the presence of glycerol and the predictive performances of the PLSR and also to evaluate the robustness of the analysis performed. Despite the extreme variations in glycerol induced between samples, the PLSR applied to the ATR-IR data is able to extract highly specific information to monitor the Omegalight® without interference from the other ingredients. Interestingly, the RMSEV and RPD obtained with the EMSC correction remains consistent between Model_2 and Model_3 (around 0.09% w/w and 30 respectively).

Table 5 Root mean squares error of cross-validation (RMSEV) obtained for different pre-processing methods (MMN, SNV, RB&VN and a fifth order polynomial EMSC model) applied to ATR-FTIR spectra from Model_2 and Model_3. Results are expressed as % (w/w) Omegalight®

		Wavenumbers (cm^{-1})					
		1800-1000			3800-1000		
	Pre-processing	RMSEV	RPD	R^2	RMSEV	RPD	R^2
Model 2	Raw data	0,2276	11.80	0,9924	0,2261	11.88	0,9925
	MMN	0.2749	9.772	0,9893	0,275	9.77	0,9890
	SNV	0.2241	11.99	0,9928	0.2617	11.99	0.9904
	RB+VN	0.2289	11.74	0,9926	0.2618	10.26	0.9906
	EMSC	0.0858	31.31	0.9989	0.1221	22.00	0.9977
Model 3	Raw data	0,1810	14.86	0,9950	0,2203	12.21	0,9924
	MMN	0.2170	12.39	0,9925	0,1949	13.80	0,9942
	SNV	0.1802	14.92	0,9949	0.1862	14.44	0.9945
	RB+VN	0.1938	13.88	0,9941	0.1960	13.72	0.9939
	EMSC	0.0853	31.53	0.9989	0.1218	22.08	0.9978

The weighting vector presented in figure 6A and 6B correspond to respective PLSR analysis performed on Model_2 and Model_3. Similarly, to Model_1, the comparison with the pure spectrum of Omegalight® supports the molecular septicity of the analysis with all major features from the ACI clearly identified. Bands located at 1739 cm^{-1}, 1504 cm^{-1}, 1493, 1462, 1446, 1375 cm^{-1}, 1363 cm^{-1}, 1246 cm^{-1}, 1111 cm^{-1} and 1041 cm^{-1} suggest neither Sorbitan monooleate, polysorbate 80, or alginate have negligible contributions in the analysis. Despite a slight change in the features amplitudes, the pattern exhibited are closely similar and it is highlighted that the strong variations in glycerol implied in Model_3 could have an impact on the PLSR analysis but the specificity of the analysis towards Omegalight® is preserved. Interestingly, the water band at 1635 cm^{-1} remains a quite strong negative feature indicating that somehow the water content influences the construction of the predictive models. Indeed, variations in the ANC_OL concentrations would lead to different water contents between samples and it is therefore understandable that such band appears in the weighting vector. Moreover, Model_3 further dissipate any doubts about the possible direct link between water contents and precision of PLSR. In that model the glycerol concentrations used generate modifications in water content but in a randomized fashion, therefore ensuring no correlation can be established between water and the predicted concentrations of Omegalight®.

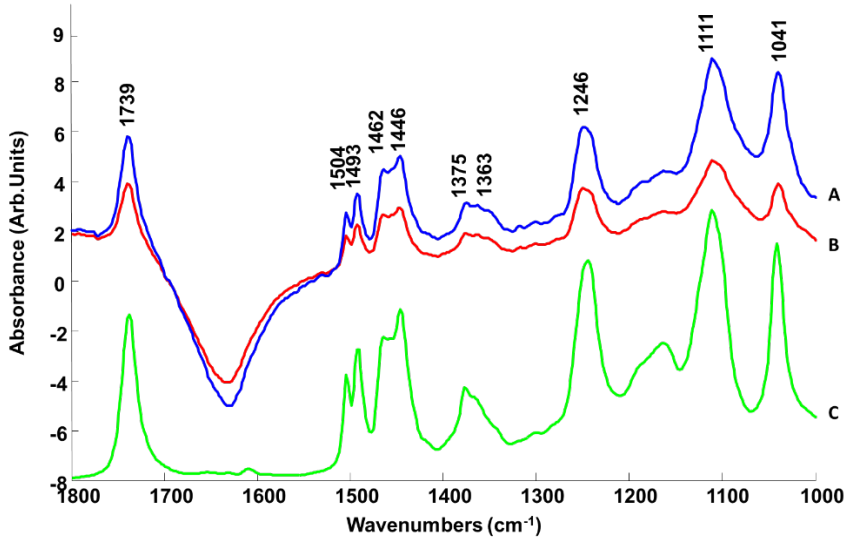

Figure 6 Plot of the first PLSR weighting vector from PLSR analysis obtained with EMSC corrected spectra from Model_2 ***(A)*** *and Model_3* ***(B)*** *compared with infrared spectrum of Omegalight®* ***(C).*** *Spectrum C offset for clarity.*

It is unlikely that glycerol or any other additive would vary that much in final cosmetic products therefore Model_3 has been mainly designed to demonstrate the specificity of the analysis, enabling quantification of an ACI such as *Omegalight*® in a complex semi-solid formulation. Consequently, it has been found more relevant to provide accuracy of the predictive model constructed from Model_2 in table 6. All predicted experimental concentrations are fairly close to the true concentrations leading to relative errors below 5% for almost all samples except number 2. Naturally, the relative error could tend to increase at lowest concentrations and observing a value of 8.3% for sample 2 could be considered as normal, however sample 1 displays a relative error of 2.44%, indicating ATR-IR can also be accurate for lowest concentrations tested. These observations unambiguously reflect the quantitative capabilities of the ATR-IR approach coupled to EMSC correction and PLSR analysis in the case of *Omegalight*® which, remains reasonably close or below to the 5% threshold.

Table 6 Predicted concentrations of Omegalight®-loaded alginate-based nanocarriers (ANC_OL) obtained by partial least squares regression (PLS-R) performed on Model_2 corrected by a fifth order polynomial EMSC on finger print region.

Samples	Observed concentration % (w/w)	Predicted concentration % (w/w)	SD % (w/w)	Relative error (%)
1	0,41	0,42	0,05	2,44
2	0,82	0,89	0,05	8,30
3	1,64	1,66	0,05	1,37
4	2,46	2,40	0,05	2,59
5	3,28	3,15	0,05	3,85
6	4,09	4,17	0,06	1,95
7	6,14	6,16	0,07	0,37
8	8,19	8,19	0,09	0,00

3.4 Validation of the Quantitative analysis of Omegalight® by means of ATR-IR

Performing quantitative analysis using blind samples is ambiguously the most relevant step to demonstrate the reliability of a method[53]. Statistically speaking, a blind sample would be a set a data (i.e. spectra) kept independent from the calibration set and only used as validation material. For instance, to ensure the calibration is not being oriented to deliver the expected outcome it appears necessary to remove all the spectra corresponding to a single concentration. To address such consideration, the Leave One Out Cross Validation (LOOCV) is particularly suited. The calibration model is constructed using 7 concentrations from the sample set tested, while the last is identified as an unknown sample or blind sample. This operation can be repeated for each sample in a sequence to evaluate how the PLSR performs according to the unknown concentration. Table 7 presents the results collected from Model_2 following EMSC. Implementing the LOOCV affects the outcome of the PLSR analysis resulting in relative errors higher than those obtained in section 3.4. While samples 3, 4, 7 and 8 have relative errors below 5%, others are between 6% and 12.72%. The study was initially designed as a proof of concept with a relatively low number of samples, and removing an entire group of spectra (12.5% of entire dataset) from the calibration model clearly affects the predictive performances. The calibration step aims to construct a model encompassing the different spectral variations to then establish the best linear relationship between the concentrations and some specific spectral features. Considering the type of samples analysed, water and glycerol are the 2 molecular species which dominate the spectra collected. Consequently, it is conceivable that some underlying modifications in the spectra are not properly integrated in the PLSR analysis due to the LOOCV. In order to reduce the influence of the corresponding features, MMN has been applied to the main water band at 1637 cm^{-1} and the glycerol dominant peak at 1041 cm^{-1}. While MMN on the water band didn't

show any improvement (data not shown), the same operation applied at 1041cm^{-1} leads to increased accuracy (table 7). Only sample 2 and 6 have relative errors above 5% with respectively 6.24% and 7.49%. Although ATR-IR is a highly sensitive technique to detect and monitor *Omegalight*® in sophisticated cosmetic formulations, the technique is also sensitive to the complex chemical matrix which somehow plays a key role in the quantitative analysis performed. Although glycerol does not affect the quantitative analysis, as demonstrated in section 3.4, the strong features displayed in the spectra collected contribute to the PLSR robustness and therefore need to be taken into account during the analysis.

Ultimately, performing the PLSR analysis using the spectral sets as blind samples further demonstrates the feasibility to perform accurate quantitative analysis by means of ATR-IR spectroscopy. While the shell formed provides relevant carriers with protective and/or enhanced penetration properties, it also contributes to the consideration of the approaches to be employed for chemical characterisation of the loaded ACI. ATR-IR appears to be adapted for the detection and quantification of the encapsulated active ingredient (for instance *Omegalight®)* in a non-destructive manner offering a suitable platform for direct analysis without fastidious and time consuming extraction protocols coupled to more advanced chromatography techniques.

Table 7 Root mean square error of validation (RMSEV) and predicted concentrations of Omegalight®- obtained with LOOCV on Model_2 after a fifth order polynomial EMSC correction applied to the finger print.

		EMSC *RMSEV = 0.1439% (w/w)* *RPD= 18.67*		EMSC + MMN (1041 cm^{-1}) *RMSEV= 0.1430 % (w/w)* *RPD= 18.79*	
Samples	Observed concentration (%w/w)	Predicted concentration % (w/w)	Relative error (%)	Predicted concentration % (w/w)	Relative error (%)
1	0,41	0,43	6,07	0,40	3,34
2	0,82	0,92	12,72	0,87	**6,24**
3	1,64	1,72	5,09	1,67	1,93
4	2,46	2,36	3,96	2,42	1,54
5	3,28	3,08	6,14	3,13	4,46
6	4,09	4,35	6,42	4,40	**7,49**
7	6,14	6,17	0,51	6,26	1,89
8	8,19	8,16	0,40	7,89	3,61

4. Conclusion

The molecular specificity of the IR analysis performed on complex chemical mixtures offers numerous perspectives for it positioning as a powerful, cost effective and rapid analytical tool. Moreover, coupled to ATR set-up the absence of sample separation protocols prior to analysis remains the greatest advantage compared to current gold standard approaches. While separation techniques such as chromatography coupled to mass spectrometry play a key role in the medical field, IR spectroscopy presents huge potential when it comes to quality control application and in particular the cosmetic industry. Considering the concentrations involved IR spectra can be directly recorded from liquid samples while the coupling with advanced data mining methods such as EMSC and PLSR leads to accuracy below 5%. Ultimately, combined with emerging high throughout technologies in the IR field, further developments could lead to short term recognition and adaptation of the methodologies in cosmetic and pharmaceutical processing to address concerns towards core-shell nanocarriers based commercialized products.

Conflicts of interest

There are no conflicts to declare

Acknowledgements

This work is part of the COSMICC project (2015-00103497). We thank Conseil Régional Centre Val de Loire and the Cosmetosciences program for financial support.

References

1 A. Mihranyan, N. Ferraz and M. Strømme, *Prog. Mater. Sci.*, 2012, **57**, 875–910.

2 R. H. Müller, M. Radtke and S. A. Wissing, *Adv. Drug Deliv. Rev.*, 2002, **54**, S131–S155.

3 R. H. Müller, R. D. Petersen, A. Hommoss and J. Pardeike, *Adv. Drug Deliv. Rev.*, 2007, **59**, 522–530.

4 *Trends Food Sci. Technol.*, 2014, **38**, 88–103.

5 H. T. P. Nguyen, E. Munnier, M. Souce, X. Perse, S. David, F. Bonnier, F. Vial, F. Yvergnaux, T. Perrier and I. Chourpa, *Nanotechnology*, 2015, **26**, 255101.

6 H. T. P. Nguyen, E. Munnier, X. Perse, F. Vial, F. Yvergnaux, T. Perrier, M. Soucé and I. Chourpa, *J. Pharm. Sci.*, 2016, **105**, 3191–3198.

7 M. Roberts, Y. Mohammed, M. Pastore, S. Namjoshi, S. Yousef, A. Alinaghi, I. Haridass, E. Abd, V. Leite-Silva, H. Benson and J. Grice, *J. Control. Release*, 2017, **247**, 86–105.

8 J. Pardeike, A. Hommoss and R. H. Müller, *Int. J. Pharm.*, 2009, **366**, 170–84.

9 L. Montenegro, F. Lai, A. Offerta, M. G. Sarpietro, L. Micicchè, A. M. Maccioni, D. Valenti and A. M. Fadda, *J. Drug Deliv. Sci. Technol.*, 2016, **32**, 100–112.

10 H. T. P. Nguyen, E. Allard-Vannier, C. Gaillard, I. Eddaoudi, L. Miloudi, M. Soucé, I. Chourpa and E. Munnier, *Colloids Surfaces B Biointerfaces*, 2016, **142**, 272–280.

11 H. T. P. Nguyen, M. Soucé, X. Perse, F. Vial, T. Perrier, F. Yvergnaux, I. Chourpa and E. Munnier, *Int. J. Cosmet. Sci.*, 2017, **39**, 450–456.

12 H. Chaudhary, K. Kohli and V. Kumar, *Int. J. Pharm.*, 2014, **465**, 175–86.

13 A. Manosroi, C. Chankhampan, W. Manosroi and J. Manosroi, *Eur. J. Pharm. Sci.*, 2013, **48**, 474–83.

14 H. T. P. Nguyen, E. Munnier, M. Souce, X. Perse, S. David, F. Bonnier, F. Vial, F. Yvergnaux, T. Perrier, S. Cohen-Jonathan, I. Chourpa, , *Nanotechnology*, 2015, **26**, 255101.

15 R. S. Managuli, L. Kumar, A. D. Chonkar, R. K. Shirodkar, S. Lewis, K. B. Koteshwara, M. S. Reddy and S. Mutalik, *J. Chromatogr. Sci.*, 2016, **54**, 1290–300.

16 E. J. Woodhouse and S. Breyman, *Sci. Technol. Hum. Values*, 2005, **30**, 199–222.

17 F. Bonnier, H. Blasco, C. Wasselet, G. Brachet, R. Respaud, L. F. C. S. Carvalho, D. Bertrand, M. J. Baker, H. J. Byrne and I. Chourpa, *Analyst*, 2017.

18 S. M. Ali, F. Bonnier, A. Tfayli, H. Lambkin, K. Flynn, V. McDonagh, C. Healy, T. Clive Lee, F. M. Lyng and H. J. Byrne, *J. Biomed. Opt.*, 2012, **18**, 61202.

19 G. N. Kalinkova, *Vib. Spectrosc.*, 1999, **19**, 307–320.

20 H. Schulz, R. Quilitzsch and H. Krüger, *J. Mol. Struct.*, 2003, **661**–**662**, 299–306.

21 S. Tankeu, I. Vermaak, G. Kamatou and A. Viljoen, *Phytochem. Anal.*, 2014, **25**, 81–88.

22 M. Sandasi, G. P. P. Kamatou, C. Gavaghan, M. Baranska and A. M. Viljoen, *Vib. Spectrosc.*, 2011, **57**, 242–247.

23 L. Miloudi, F. Bonnier, D. Bertrand, H. J. Byrne, X. Perse, I. Chourpa and E. Munnier, *Anal. Bioanal. Chem.*, 2017, **409**, 4593–4605.

24

25 H. Masmoudi, Y. Le Dréau, P. Piccerelle and J. Kister, *Int. J. Pharm.*, 2005, **289**, 117–131.

26 A. A. Gowen, C. P. O'Donnell, P. J. Cullen and S. E. J. Bell, *Eur. J. Pharm. Biopharm.*, 2008, **69**, 10–22.

27 P. Bassan, H. J. Byrne, F. Bonnier, J. Lee, P. Dumas and P. Gardner, *Analyst*, 2009, **134**, 1586.

28 J. Gerretzen, E. Szymańska, J. Bart, A. N. Davies, H.-J. van Manen, E. R. van den Heuvel, J. J. Jansen and L. M. C. Buydens, *Anal. Chim. Acta*, 2016, **938**, 44–52.

29 P. Heraud, B. R. Wood, J. Beardall and D. McNaughton, *J. Chemom.*, 2006, **20**, 193–197.

30 H. J. Byrne, P. Knief, M. E. Keating and F. Bonnier, *Chem. Soc. Rev.*, 2016.

31 *TrAC Trends Anal. Chem.*, 2009, **28**, 1201–1222.

32 *TrAC Trends Anal. Chem.*, 2013, **50**, 96–106.

33 H. Butler, L. Ashton, B. Bird, G. Cinque and K. Curtis, *Nat. Protoc.*, 2016.

34 M. J. Baker, J. Trevisan, P. Bassan, R. Bhargava, H. J. Butler, K. M. Dorling, P. R. Fielden, S. W. Fogarty, N. J. Fullwood, K. A. Heys, C. Hughes, P. Lasch, P. L. Martin-Hirsch, B. Obinaju, G. D. Sockalingum, J. Sulé-Suso, R. J. Strong, M. J. Walsh, B. R. Wood, P. Gardner and F. L. Martin, *Nat. Protoc.*, 2014, **9**, 1771–91.

35 O. Ibrahim, A. Maguire, A. D. Meade, S. Flint, M. Toner, H. J. Byrne and F. M. Lyng, *Anal. Methods*, 2017, **9**, 4709–4717.

36 A. Cao, A. K. Pandya, G. K. Serhatkulu, R. E. Weber, H. Dai, J. S. Thakur, V. M. Naik, R. Naik, G. W. Auner, R. Rabah and D. C. Freeman, *J. Raman Spectrosc.*, 2007, **38**, 1199–1205.

37 L. Kerr, 2016.

38 R. Gautam, S. Vanga, F. Ariese and S. Umapathy, *EPJ Tech. Instrum.*, 2015, **2**, 8.

39 N. K. Afseth, V. H. Segtnan and J. P. Wold, *Appl. Spectrosc.*, 2006, **60**, 1358–1367.

40 M. S. Dhanoa, R. J. Barnes and S. J. Lister, *Appl. Spectrosc. Vol. 43, Issue 5, pp. 772-777*, 1989, **43**, 772–777.

41 *Chemom. Intell. Lab. Syst.*, 2012, **117**, 92–99.

42 S. Wold, M. Sjöström and L. Eriksson, in *Chemometrics and Intelligent Laboratory Systems*, 2001, vol. 58, pp. 109–130.

43 *Eur. J. Soil Sci.*, 2012, **63**, 141–151.

44 Y. Kim, M. Singh, S. K.-F. chemistry and undefined 2007, *Elsevier*.

45 D.-W. Sun, *Computer vision technology for food quality evaluation*, Elsevier/Academic Press, 2016.

46 *Prog. Polym. Sci.*, 2012, **37**, 106–126.

47 L. Pereira, A. Sousa, H. Coelho, A. M. Amado and P. J. A. Ribeiro-Claro, *Biomol. Eng.*, 2003, **20**, 223–228.

48 M. E. Keating and H. Byrne, 2015, **2492**, 2482–2492.

49 L. Eriksson, T. Byrne, E. Johansson, J. Trygg and C. Vikström, *Multi- and megavariate data analysis : basic principles and applications*, .

50 M. Palo, K. Kogermann, N. Genina, D. Fors, J. Peltonen, J. Heinämäki and N. Sandler, *J. Drug Deliv. Sci. Technol.*, 2016.

51 F. Bonnier, H. Blasco, C. Wasselet, G. Brachet, R. Respaud, L. F. C. S. Carvalho, D. Bertrand, M. J. Baker, H. J. Byrne and I. Chourpa, *Analyst*, 2017.

52 Franck Bonnier, M. J. Baker and H. J. Byrne, *Anal. Methods*, 2014, **6**, 5155–5160.

53 S. Guo, T. Bocklitz, U. Neugebauer and J. Popp, *Anal. Methods*, 2017, **9**, 4410–4417.