IL NUOVO CIMENTO **47 C** (2024) 268
DOI 10.1393/ncc/i2024-24268-7

COMMUNICATIONS: SIF Congress 2023

Decoding complexity: A comprehensive review of MCR-ALS applications in advancing biological insights

V. NOTARSTEFANO

Dipartimento di Scienze della Vita e dell'Ambiente, Università Politecnica delle Marche Ancona, Italy

received 22 January 2024

Summary. — The great potential of vibrational spectroscopic techniques in the analysis of biological materials has been widely proven. The applications are numerous and range from diagnostics to drug development. However, to unveil the needed information from vibrational spectra, multiple chemometric methods have been elaborated, among which the Multivariate Curve Resolution - Alternating Least Squares (MCR-ALS) approach is gaining a growing interest to unmix the complex spectroscopic dataset into smaller matrices of physically meaningful pure spectral components and their relative concentrations. The present paper reviews some recent applications of the coupling of MCR-ALS regression and vibrational spectroscopies to biological samples and its potential in disentangling complex biological information within the analysed matrices.

1. – Introduction

Vibrational spectroscopies, including Fourier Transform InfraRed (FTIR) and Raman spectroscopies, are widely and successfully applied in biophysical and biomedical research to unravel molecular details of biological samples [1-4]. Raman and IR bands/peaks, with their position, intensity and width, describe lipids, proteins, nucleic acids and sugars within cells and tissues. These spectral fingerprints are useful to apply these techniques to diagnostic purposes [5, 6] and for the assessment of drug toxicity and mode of action [7, 8]. However, in order to relate the complex biologically derived spectrochemical data to the biological processes of interest, numerous chemometric approaches have been developed and exploited [9, 10]. In fact, the vibrational spectral profiles of most biological samples drastically overlap, making it challenging to identify the information of interest. In particular, the present review overviews the applications of the Multivariate Curve Resolution – Alternating Least Squares (MCR-ALS) approach on vibrational spectroscopic dataset collected from complex biological samples, including cells and tissues. The MCR-ALS approach, firstly developed for chemical analyses to study the evolution of mixtures with multiple concentrations, is based on a matrix approximation method, which has shown to be a valuable tool to extract kinetic and biophysical information from complex biological samples, like cells and tissues [10, 11]. The bases of this multivariate statistical approach will be depicted in sect. **2**, and some applications of MCR methods

on spectroscopic datasets will be exemplified in sect. **3**. Finally, sect. **4** will portray some studies highlighting the big potential of MCR-ALS in the biological and biomedical field.

2. – Multivariate curve resolution

Multivariate Curve Resolution (MCR) has been widely employed to address the so-called mixture analysis problem, which, citing de Juan and Tauler, "*comes down to realize that all what is observable or measurable in a real system seldom responds to a single cause or source*" [12]. All the variations that are observed in a system, whether influenced by time and other conditions, reflect the combined effects of numerous individual varying components present in the mixture. Hence, the challenge is to determine these individual factors based on the measured data, even in the case of overlapping signals gathered from the single components. MCR in this sense is applied to comprehend these factors by representing data as a bilinear model of pure component contributions [13]. In various cases, the observed response of the system is an instrumental measurement, hence an additive model can be employed to describe the individual signal contributions [12]. Equation (1) expresses the mixture analysis model: $\mathbf{D}$ is a data table collecting in row all the mixed responses gathered at a particular stage; n represents the sources of variation; $\mathbf{c}_i$ reflects qualitatively the nature of the source of variation $\mathbf{s}_i$; $\mathbf{E}$ corresponds to the unexplained variation:

$$\mathbf{D} = \sum_{i=1}^{n} \mathbf{c}_i \mathbf{s}_i^T + \mathbf{E}. \tag{1}$$

Equation (2) expresses in a more general way the bilinear MCR model (fig. 1): $\mathbf{S}^T$ is a matrix with n rows $\mathbf{s}_i$; $\mathbf{C}$ is a matrix with n columns containing the distribution profiles [12]:

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E}. \tag{2}$$

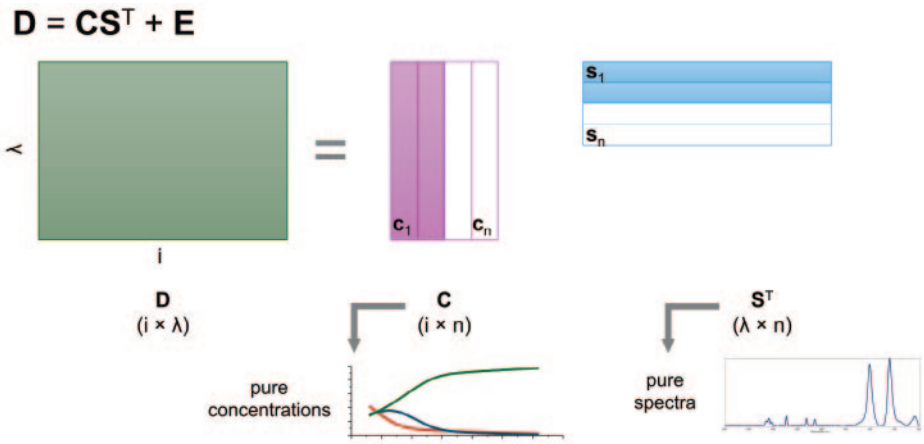

Fig. 1. – Decomposition by multivariate curve resolution of a spectra/concentrations dataset. Adapted from literature [12, 14, 15].

3. – MCR-ALS and vibrational spectroscopic datasets

MCR methods have been successfully employed to analyse spectroscopic data. First, UV spectroscopic data from two-compound mixtures were modelled under non-negativity constraints [16]. Then, data deriving from other spectroscopic measurements also on more complex systems were analysed by MCR. Here, some examples of the application of MCR-ALS on vibrational spectroscopic datasets are reported (more examples are available at [17]).

Considering the paradigmatic application of MCR on spectroscopic data, the bilinear model can be considered as follows: the matrix $\mathbf{D}$ $(i \times \lambda)$ contains the spectral data of i samples and λ variables, the matrix $\mathbf{C}$ $(i \times n)$ contains the concentrations of the system's n components for all the spectra, the matrix $\mathbf{S}^T$ $(\lambda \times n)$ contains the transposed spectra of the pure components, and $\mathbf{E}$ $(i \times \lambda)$ represents the residual matrix [11].

In order to optimize the set of initial estimates of spectra or concentration profiles, iterative methods are widely implemented upon selection of constraints; these methods are useful, given their flexibility and the lack of need of assumption in the model [13, 17]. One of the most common iterative methods is Multivariate Curve Resolution-Alternating Least Squares (MCR-ALS), proposed by Tauler [18], which works with either the full matrix $\mathbf{C}$ or $\mathbf{S}^T$: in every iteration, under the action of specific constraints, it calculates both matrices of the bilinear model, until the $\mathbf{CS}^T$ model satisfactorily reproduces the original data $\mathbf{D}$ [13]. In MCR-ALS, it is crucial to start with a set of reasonable initial estimates, which follow the constraints that will be applied in the optimization process and can be derived from prior knowledge or be gathered with some methods [13]. Among these, the local rank analysis method Evolving Factor Analysis (EFA) is the most applied: by repeatedly applying Principal Component Analysis (PCA), it highlights the components' emergence and decay within the dataset and, assuming a sequential emergence-decay order for the components, provides an estimation of the concentration profiles [13, 14, 17].

Equation (1) would have an ambiguous solution without supplementary information [17]. In fact, as mentioned above, the MCR optimization process needs to be driven by selected constraints, which are mathematical/chemical conditions, also related to the intrinsic properties of the system, that reduce the infinite number of solutions of the equation, by shaping the row and column profiles in the bilinear model [13,17-19]. To pursue the example of MCR application to spectroscopic dataset, a widely applied constraint is non-negativity, which forces only positive values, since only positive concentration values are meaningful; moreover, the equality constraint is useful when a prior knowledge of pure spectra/concentration profiles exists [13, 17].

As mentioned above, spectroscopic studies have extensively integrated MCR methods, which let unravel the pure spectra and concentrations profiles from a measured dataset, by determining the number of significant components.

One example of this approach is reported in [20], where MCR-ALS was applied on a dataset composed of hetero-spectral two-dimensional correlation of FTIR and FT-Raman data obtained from the measurement of vegetable oils, with the aim of evaluating the degree of lipid oxidation. In this work, the authors employed singular value decomposition (SVD) and PCA to assess the number of components/contributions to include; then, the SIMPLISMA (SIMPLe-to-use Interactive Self-modelling Mixture Analysis) approach was used to determine the purest variable and the ALS method (under some constraints, including the non-negativity one) was employed as iterative optimization method. Although the spectroscopic dataset collected from the oil matrix was complex, given to the co-existence of a variety of oxidative reaction products, MCR-ALS provided

useful information related to some lipid-oxidation steps, such as the initial formation of conjugated double bonds, the formation of saturated carbonylic compounds, the specific oxidation behaviour of the analysed samples [20].

A peculiar application of the combination of Raman spectroscopy and MCR methods is reported in [21]: this approach was exploited to study the hydration properties of aggregates composed of three different cationic ammonium surfactants by monitoring the transitions of micelle shapes from spherical to rod-like to worm-like. Thanks to the MCR approach performed on the Raman spectroscopic dataset, it was possible to understand how the change of shape induced variations in weaker water tetrahedral order and weaker H-bonding, both in the headgroup and in the hydrophobic chain. This contributed to highlight the important links between the structure of self-assemblies and their hydration states.

MCR-ALS was also used to study the hydrolysis of bovine serum albumin with protease K. In this study, the dataset was composed of infrared spectra, and the chemometric approach was combined with two-dimensional correlation spectroscopy (2DCoS). SVD was used to estimate the number of components; EFA was selected to estimate either the $\mathbf{C}$ or $\mathbf{S}^T$ matrices, and the iterative ALS process was used for optimization. Peculiarly in this study, a special attention was dedicated to the residuals obtained from the MCR analysis [$\mathbf{E} = \mathbf{D} - (\mathbf{CS}^T)$]: by this approach, the model could be implemented, by including components that were discarded by the initial model, reaching a 99.99% of explained accumulative variance [22].

It has been reported that MCR-ALS outperforms other more commonly employed regression approaches, but it is noteworthy to pinpoint some limitations of its application, which need to be taken into account. Some of these limitations are related to peculiar data conditions, including a small number of samples, background signals, and low signal-to-noise ratio [23]. A notable limitation is the phenomenon of the so-called rotational ambiguity, linked to the non-unicity of the solution: in fact, several combinations of concentration/response fulfilling the model exist [24, 25]. Another known limitation is given by the way of considering noise: it is reported that in MCR the noise is mostly considered independent from the signal and ideally distributed (iid) [25]; hence, MCR will inevitably direct to over-fitting, whenever a signal-related noise is present.

4. – MCR-ALS and biospectroscopy

A peculiar application of MCR-ALS in biophysics is to disentangle some dynamic modifications in molecular components of biological samples [10] and extract kinetic (also non-linear) and biochemical information from spectra. Some examples of noteworthy applications of MCR-ALS on biospectroscopic datasets are here reported.

A recent work reported the application of MCR-ALS to reveal some aspects related to time-dependent cellular processes, like drug uptake and cellular responses [11]. The authors were able to distinguish between drug and cell response Raman spectra. As regards the statistical approach, non-negativity and unimodality constraints were employed in the model, and SVD was used prior MCR-ALS, to assess the number of components. MCR-ALS was modelled with and without equality constraints (a constraints used when the signal of a compound is known). Considering the experimental dataset employed, obtained by measuring cells treated with doxorubicin, two components were used: the first one displayed the spectral features of the drug, while the second was in part superimposable with the spectral profile of the cell. The analysis along time of the Raman peaks assigned to doxorubicin evidenced their temporal increase, suggesting its uptake

by cells, while the decrease of the DNA- and RNA-related bands was indicative of the known mode of action of the drug. In general, the authors reported a usefulness of MCR-ALS in studying the progression along time of both drug uptake and cellular responses, also considering the non-linear kinetic nature of these biological processes.

MCR-ALS was also exploited on a multimodal Raman and IR dataset obtained from cells treated with doxorubicin [26]. The use of the two spectroscopic techniques let, on one hand, have a Raman spectral signature of the drug, together with some indications on DNA intercalation, and on the other hand, observe biochemical changes in cells, as a response to the uptake of doxorubicin. In particular, MCR-ALS was used to model the IR spectra collected on treated cells, and the model was constrained to use as the first component the Raman spectrum of the drug, by employing known kinetics of drug uptake.

A phenomenological rate equation approach, was also developed: it is a hard-and-soft modelling approach that consists in the use of chemical kinetic reactions under some rate constants, with the aim of finding chemically interpretable solutions from the analysis of biological systems [27]. These reactions can be very simple, like A $\rightarrow$ B, A $\rightarrow$ B $\rightarrow$ C, A + B $\rightarrow$ C + D, and may reveal some biological processes, such as drug uptake, drug binding, and cell response [27]. Pursuing this approach the set of ordinary differential equations that define this system are employed as kinetic constraints in the iterative optimization process of MCR-ALS. This approach led to a successful modelling of the system, in particular explaining the temporal progress of concentration profiles, and also elucidating the spectral features assigned to drug binding and cellular responses.

With the same approach, we studied oral squamous carcinoma cells treated with cisplatin, which has a fully defined mode of action. In particular, the temporal progression of the cell population from viable (V), to early apoptotic (EA), to dead (D) cells was modelled as V $\rightarrow$ EA $\rightarrow$ D: this system was described by a simple set of ordinary differential equations and the rate constants were calculated, in order to be used to describe by MCR-ALS the evolution of the spectral modifications along time. This approach was able not only to unveil Raman spectral signatures of the temporal progression of cell death, but also to quantify the rates of the cellular response [28]. Figure 2 depicts the steps of the described approach.

MCR-ALS can decompose vibrational spectra also with a diagnostic aim. A recent work proposed this approach to identify the spectral signatures of skin tissues [29]. Raman spectra were recorded on skin samples representing a normal condition, keratosis, basal cell carcinoma, malignant melanoma, and pigmented nevus: MCR-ALS let obtain from this complex spectral dataset the signature of the main skin components, such as water, lipids, melanin, and proteins, also in relation to the specific sample conditions. Then, the relationships between the concentration profiles of the components obtained by MCR-ALS and the Raman spectral profiles were investigated by logistic regression, in particular to compare malignant/benign neoplasms and melanoma/keratosis/pigmented nevus, providing promising results for diagnostic purposes [29].

Also data from imaging vibrational spectroscopy have been analysed by MCR. This class of methods is also referred to as hyperspectral imaging (HSI), which integrates imaging and spectroscopy, providing a morpho-chemical correlation of the examined sample [30]. Briefly, the microscope coupled with the spectrometer allows for the selection of an area of interest, defined by x and y coordinates, expressed in pixels; in each pixel, a spectrum is acquired (λ variables per spectrum), and an experimental data cube ($x \times y \times \lambda$) is generated. Some works in literature explain how to extract data to employ MCR on imaging vibrational data from the data cube [31-33]. First, the data cube needs to be

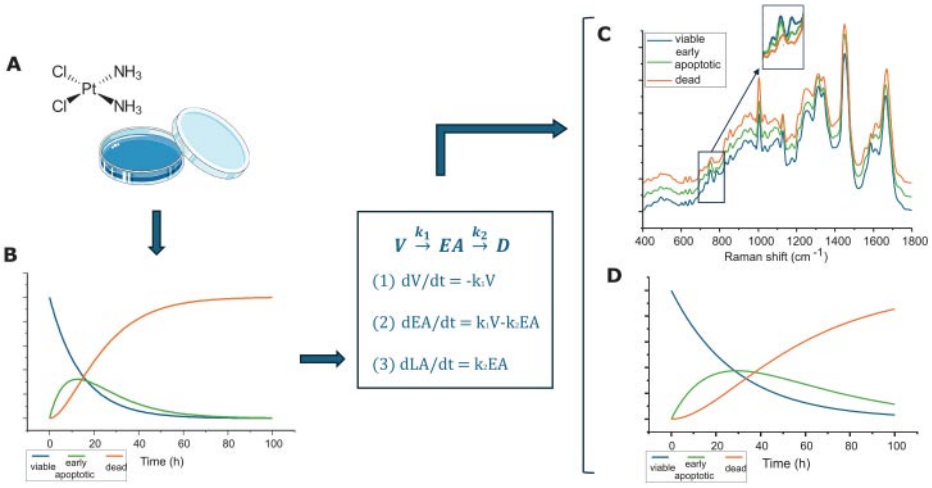

Fig. 2. – Steps followed to apply MCR-ALS for the study of OSCCs response to cisplatin: (A) *in vitro* treatment; (B) assessment of the kinetic progression of cells from viable, to early apoptotic, to dead; employment of the obtained ordinary differential equations and MCR-ALS to identify (C) the three component spectra of the system, corresponding to viable, early apoptotic, and dead cells, and (D) the kinetic evolution of the relative contributions of the three components to the measured Raman spectrum, at each time [28].

unfolded, and the two-way matrix $\mathbf{D}$ is obtained ($xy \times \lambda$). On this matrix, the MCR bilinear model can be applied, as previously discussed (fig. 3).

Respect to other multivariate exploratory approaches on hyperspectral data, MCR-ALS aims at estimating the concentration of the analytes in each pixel. For example, Duponchel and colleagues compared MCR-ALS to other diverse multivariate analysis methods, including Orthogonal Projection Approach (OPA) and SIMPLISMA, and found MCR-ALS leading to a proper extraction of the spectral profiles of the pure compounds [32].

As mentioned above, MCR-ALS is peculiarly suitable to follow the temporal progression of biophysical, chemical, and biological processes. In this light, a recent study applied MCR-ALS to a Stimulated Raman Scattering (SRS) dataset and investigated the rates of penetration of 4-cyanophenol into human skin, with the aim of assessing the concentration of the selected compound at different depths, overcoming the interference of Raman signals of skin itself [34]. After analysing the SRS spectral profile of 4-cyanophenol, the MCR-ALS of skin samples was performed, adding to the SVD-derived components, a component related to the spectrum of the chemical (equality constraint), in order to guide the binary model to find the concentration map of the target species. MCR-ALS was able to isolate the signals of 4-cyanophenol even within a complex biological environment.

Similarly, the transcutaneous permeation of retinol into frozen and living human skin was assessed coupling MCR-ALS and Raman imaging. Three different components were considered as constituents of all Raman spectra: the spectrum of native skin, the spectrum of free retinol and the one of encapsulated retinol. Useful biological results were obtained from this approach, including the assessment of a better transcutaneous penetration of retinol upon encapsulation and in living skin, together with the evidence of

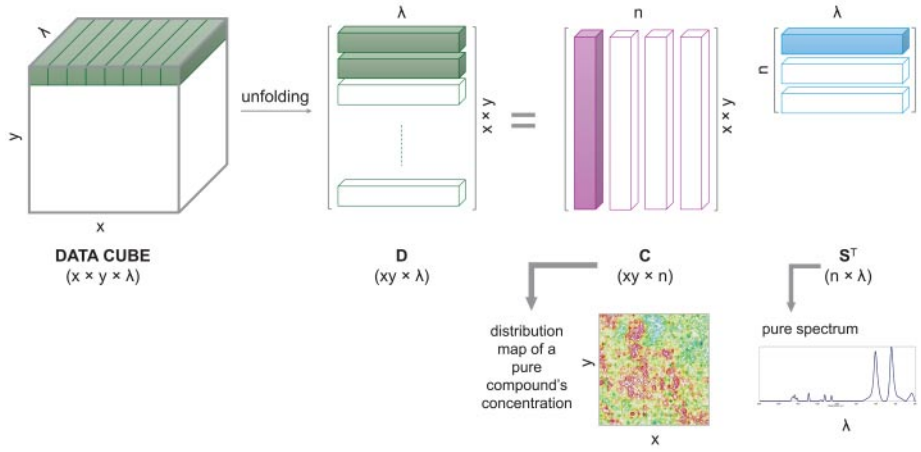

Fig. 3. – Multivariate curve resolution methods applied to hyperspectral datasets. Adapted from literature [13, 32, 35, 36].

retinol accumulation in the *stratus corneum* both in living and frozen tissue. Respect to other fitting techniques, MCR-ALS let take into account the component spectra related to retinol maintaining a certain degree of flexibility, which made it also possible to evaluate the changes in the Raman signal of retinol due to its interactions with skin [37].

5. – Conclusions

The potential of Raman and IR spectroscopies, including hyperspectral imaging, in the biological and biomedical field has been widely and successfully proven. This combined approach showed to be powerful to investigate for example drug mode of action and kinetics, suggesting a great potential in preclinical screening in drug development research. In the specific case of MCR-ALS to vibrational hyperspectral datasets, this chemometric approach is very effective in disentangling, in a label-free way, the distribution of several chemical species, both from the biological matrix and exogenous ones, even in mapped regions where their vibrational signals are overlapped. These chemical results demonstrated to be extremely useful to unravel biological processes within the analysed matrix.

* * *

The author acknowledges for the funding the European Union - NextGenerationEU under the Italian Ministry of University and Research (MUR) National Innovation Ecosystem grant ECS00000041 - VITALITY - CUP I33C22001330007.

REFERENCES

[1] Baker M. J. *et al.*, *Nat. Protoc.*, **9** (2014) 1771.
[2] Belloni A. *et al.*, *BBA-Mol. Basis Dis.*, **1870** (2024) 166873.
[3] Belloni A. *et al.*, *Appl. Sci.*, **13** (2023) 63954.
[4] Butler H. J. *et al.*, *Nat. Protoc.*, **11** (2016) 664.
[5] Belloni A. *et al.*, *BBA-Mol. Basis Dis.*, **1868** (2022) 166494.

[6] Licini C. *et al.*, *BioFactors*, **48** (2022) 1089.
[7] Campagna R. *et al.*, *Int. J. Mol. Sci.*, **24** (2023) 338.
[8] Notarstefano V. *et al.*, *Spectrochim. Acta A*, **269** (2022) 120735.
[9] Morais C. L. M. *et al.*, *Nat. Protoc.*, **15** (2020) 2143.
[10] Noothalapati H. *et al.*, *Anal. Sci.*, **33** (2017) 15.
[11] Perez-Guaita D. *et al.*, *Talanta*, **208** (2020) 120386.
[12] de Juan A. and Tauler R., *Anal. Chim. Acta*, **2021** (1145) 79.
[13] de Juan A. *et al.*, *Anal. Methods*, **6** (2014) 4964.
[14] de Juan A. and Tauler R., *Anal. Chim. Acta*, **500** (2003) 195.
[15] Mazivila S. J. and Santos J. L. M., *TrAC*, **157** (2022) 116698.
[16] Lawton W. H. and Sylvestre E. A., *Technometrics*, **13** (1971) 617.
[17] Garrido M. *et al.*, *Anal. Bioanal. Chem.*, **390** (2008) 2059.
[18] Tauler R., *Chemom. Intell. Lab. Syst.*, **30** (1995) 133.
[19] de Juan A. *et al.*, *Anal. Chim. Acta*, **346** (1997) 307.
[20] Muik B. *et al.*, *Anal. Chim. Acta*, **593** (2007) 54.
[21] Shen Y. *et al.*, *J. Phys. Chem. Lett.*, **11** (2020) 7429.
[22] Domí Ãnguez-Vidal A. *et al.*, *Anal. Chem.*, **78** (2006) 3257.
[23] Debus B. *et al.*, *Chemom. Intell. Lab. Syst.*, **198** (2020) 103941.
[24] Pellegrino Vidal R. B. *et al.*, *Anal. Chim. Acta*, **1003** (2018) 10.
[25] Ruckebusch C. and Blanchet L., *Anal. Chim. Acta*, **765** (2013) 28.
[26] Perez-Guaita D. *et al.*, *J. Biophoton.*, **13** (2020) 12.
[27] Perez-Guaita D. *et al.*, *Cells*, **11** (2022) 1555.
[28] Notarstefano V. *et al.*, *Analyst*, **148** (2023) 4365.
[29] Matveeva I. *et al.*, *Sensors*, **22** (2022) 9588.
[30] Notarstefano V. *et al.*, *J. Biophoton.*, **13** (2020) 4.
[31] Amigo J. M. *et al.*, *TrAC*, **27** (2008) 696.
[32] Duponchel L. *et al.*, *J. Chem. Inf. Comput. Sci.*, **43** (2003) 2057.
[33] Felten J. *et al.*, *Nat. Protoc.*, **10** (2015) 217.
[34] Goel A. *et al.*, *Spectrochim. Acta A*, **296** (2023) 122639.
[35] Offroy M. *et al.*, *Sci. Rep.*, **5** (2015) 12303.
[36] Smith J. P. *et al.*, *Analyst*, **144** (2019) 5425.
[37] Essendoubi M. *et al.*, *Skin Res. Technol.*, **27** (2021) 1100.