IFSCC 2025 CONGRESS
FRANCE_CANNES 15-18 SEPT 25
THE FUTURE IS SCIENCE
ifscc2025.com

IFSCC 2025 full paper (IFSCC2025-963)

Accelerating Natural Cosmetic Innovation:

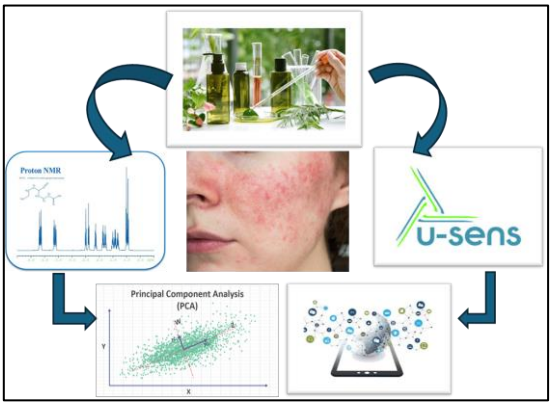
A digital Approach to predicting Skin Sensitization Potential

Francisco ALVAREZ [*1], Cecile PIROIRD [1], Nathalie ALEPEE[1], Laurent BOISAUBERT[1], Brice BONNET[1], Mohamed N. TRIBA[2] and Ratnadeep PAUL CHOUDHURY[3]

[1] Research & Innovation, L'OREAL, Aulnay sous Bois, France,
[2]Université Paris 13, Laboratoire de Spectroscopie des Biomolécules et des Milieux Biologiques (SBMB), 74 rue Marcel Cachin, Bobigny, France
[3] Research & Innovation, L'OREAL, Bengaluru, India

Corresponding Author: Francisco ALVAREZ <francisco.alvarez@loreal.com>

Abstract: The growing demand for naturals in cosmetics presents a challenge: ensuring the safety of complex plant-based ingredients. To address this, we developed a digital tool that predicts the skin sensitization potential of natural extracts, accelerating product development while maintaining high safety standards. This innovative method correlates NMR spectral fingerprints with in vitro U-SENS™ test results. Using a database of 117 samples (plant extracts, fractions, and biomarkers), a custom Matlab application processed NMR data to build a predictive model. Principal Component Analysis (PCA) revealed a correlation between NMR data and U-SENS™ EC150 values (the concentration inducing a positive response). An Orthogonal Partial Least Squares (OPLS) model then classified extracts as "Favorable" (EC150 > 100 µg/mL, low sensitization) or "Unfavorable" (EC150 < 40 µg/mL, high sensitization). The model demonstrated good predictive capability (Kappa 0.7, accuracy 86%). Implementing a 65% probability threshold optimized sensitivity and specificity (both >91%) and introduced an "inconclusive" category. This NMR-based approach efficiently pre-screens plant extracts for sensitization potential, streamlining raw material development and accelerating innovation in



natural cosmetics. While this tool complements, not replaces, confirmatory U-SENS™ and other NAMs testing, it allows for early identification and prioritization of promising candidates.

Keywords: Skin Sensitization, Nuclear Magnetic Resonance (NMR), U-SENS™, Principal Component Analysis (PCA), Orthogonal Partial Least Squares (OPLS),

1. Introduction

The burgeoning demand for naturals in cosmetics presents a dual-edged sword for the industry. Consumers are increasingly drawn to nature-derived products, prompting the incorporation of plant-based extracts rich in bioactive molecules [1-2]. While these ingredients offer exciting potential for innovative cosmetic development, ensuring their safety and preserving the complex, often delicate, balance of their composition presents a significant hurdle [3]. Current safety assessments, often time-consuming and expensive, are further complicated by the inherent variability of these natural extracts.

To accelerate the "green cosmetics revolution" and address the challenge of safely and efficiently screening natural ingredients, we developed an innovative digital tool. This tool correlates the Nuclear magnetic resonance (NMR) spectral fingerprints of natural extracts with their skin sensitization potential, as measured by the in vitro U-SENS™ test. NMR is a versatile tool for determining the structural, chemical and physical properties of small and medium-sized molecules under a wide variety of sampling conditions. Using proton (^{1}H) or carbon (^{13}C) resonance frequencies, it is possible to rapidly obtain information on chemical functions based on chemical shift. Non-destructive, non-invasive and reproducible, NMR is the perfect tool for building a spectral database and correlating chemical information with biological activities. NMR metabolomics studies, often reported in the literature, are good examples of this type of correlation [4]. In this research work, proton (^{1}H)NMR spectra was used to provide a global fingerprint of the sample. This fingerprint is a single spectrum, which is a set of signals specific to each small molecule (metabolite) making up the biological sample. This rapid, cost-effective pre-screening method identifies potentially sensitizing ingredients early in product development, classifying them as either "Favorable" (low potential) or "Unfavorable" (high potential). While not a replacement for validated methods like the U-SENS™ and the KeratinoSens™

test [5,6], this tool can streamline product development, prioritizes resources, enhances consumer safety, and represents a significant advancement in predictive toxicology for more sustainable practices within the natural cosmetics industry. This study focuses on developing this digital tool to predict the sensitizing activity of botanicals using proton (1H) NMR spectroscopy.

2. Materials and Methods

2.1 Botanical Samples

In this study, we focused on 117 samples consisting of 92 undiluted botanical extracts, 5 fractions and 20 biomarkers. The selection of these samples was based on the availability of pre-existing U-SENS data and established chemical profiles of these samples.

2.2 ^{1}H NMR Measurement:

NMR spectra of the 117 botanicals samples were acquired on a Bruker Avance I spectrometer operating at 400.13 MHz 1H NMR resonance frequency equipped with a standard 5-mm BBI probe and automatic sample changer. Tuning and matching of the probe, locking on solvent and resolution setting were automatically done on each sample prior to spectral acquisitions using 1D 1H proton (pulse program “zg”) sequence. A total of 16 free induction decays (FIDs) were recorded into 16K points with a spectral width of 20 ppm. The relaxation delay was set to 1 s. and the acquisition time to 0,95 s. FIDs were multiplied by an exponential weighting function corresponding to line broadening of 0.3 Hz prior to Fourier transform. Experiments were carried out at a temperature of 27°C. The 1H-NMR chemical shifts were referenced to tetramethylsilane (TMS) signal at 0.0 ppm. 50 mg of sample was weighed out and dissolved in 1 mL of DMSO-d6 containing 0.03% TMS. In this study, we generated a total of 117 NMR spectra (1H proton) that was used to prepare our spectral reference database of botanical extracts.

2.3 U-SENS EC150 Measurement

The U-SENS™ method (7) quantifies the change in the expression of a cell surface marker associated with the process of activation of monocytes and dendritic cells (DC) in the human histiocytic lymphoma cell line U937, following exposure to sensitizers. The measured expression levels of CD86 cell surface marker in the cell line U937 is then used for supporting

the discrimination between skin sensitizers and non-sensitizers. The following parameters are calculated in the U-SENS™ test method: CV70 value, i.e. a concentration showing 70% of U937 cell survival (30% cytotoxicity) and the EC150 value, i.e. the concentration at which the test chemicals induced a CD86 stimulation index (S.I.) of 150%. The individual conclusion of an U-SENS™ run is considered Negative if the S.I. of CD86 is less than 150% at all non-cytotoxic concentrations (cell viability ≥ 70%) and if no interference is observed (maximal concentration of 200 µg/mL). In all other cases: S.I. of CD86 higher or equal to 150% and/or interferences observed, the individual conclusion of an U-SENS™ run is considered Positive.

2.4 Data Processing and Statistical analysis:

Phase correction, baseline correction and calibration were performed for each NMR spectrum using PepsNmr [8]. The 117 processed NMR spectra were reduced to 6000 integrated regions of equal width (0.002 ppm) and compiled in a X data matrix of 117 rows and 6000 columns. To control the effect of a possible variability in samples concentration probabilistic quotient normalization [9] was performed on X. Prior to multivariate analysis, centering and Pareto variable scaling of X was applied. The multivariate analysis methods we used in this study are mentioned below. They were performed with an in-house Matlab1 code using the same algorithm as SIMCA-P (Umetrics, Umeå Sweden) to detect any group separation based on signal variability.

Unsupervised analysis: Analysis of spectroscopic data using principal component analysis (PCA) is usually the first step in statistical study [10]. The aim of this technique is to reduce the multidimensional space of the data while preserving the majority of the variance [11]. In general, PCA is used to detect outliers or spontaneous sample grouping.

Supervised analysis: The second step in most statistical analyses is to use an orthogonal latent structure projection (OPLS) to discriminate samples. Compared to the classical latent structure analysis (PLS) projection, this method improves the interpretation of spectroscopic variations between discriminated groups by removing orthogonal information that has no impact on discrimination. OPLS models were calculated using the method of Trygg and Wold [12].

3. Results

3.1 Unsupervised analysis:

A 117 NMR proton (1H) spectra database, was established. The first step in this study was to perform principal component analysis (PCA) in order to detect any relationships between NMR and U-SENS™ data. Observation of the PCA curve (Figure 1) shows that when the T2 component of the PCA (NMR variables) is correlated with the main U-SENS™ quantitative readout (the EC150 value (calculated concentration in µg/mL required to reach the positive threshold for CD86 induction). Two groups are distinguished in this database, that could correspond to the “Favorable” and “Unfavorable” U-SENS profiles of samples. This result confirms that there is a clear correlation between NMR data (chemical functions detected on a proton spectrum) and biological U-SENS EC150 end value measurements (yellow and blue dots, respectively Favorable and Unfavorable values).

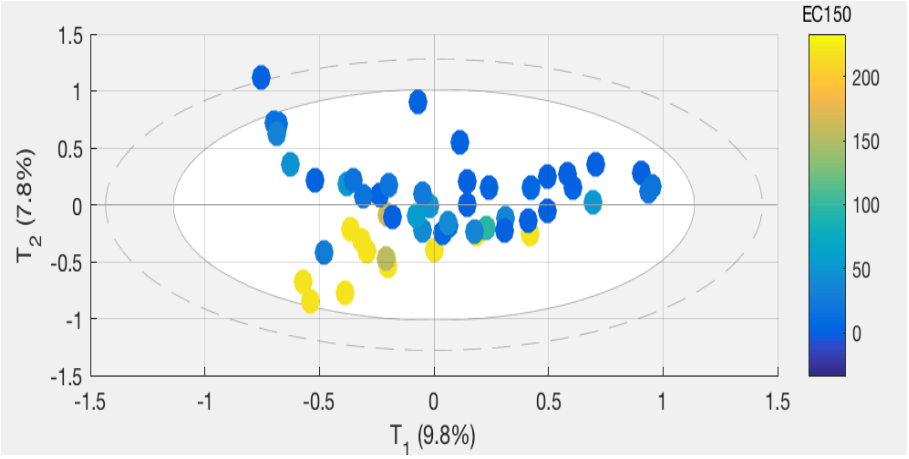


Figure 1. Results of principal component analysis (PCA) of 1 H NMR spectra

3.2 Supervised analysis:

Based on the results obtained from the PCA analysis, we assessed the possibility of building a predictive model of EC150 (U-SENS test) from the NMR data. As a first step, we attempted to build a regression model directly predicting the EC150 value. We found that the predictive capacity of this model was low ($Q2 < 0.5$). One reason for this is that there is no linear relationship between EC150 and NMR signal intensity. Using an Orthogonal Projections to Latent Structures (OPLS) regression model, we therefore performed a predictive model to predict

membership of two classes: "Favorable" (EC150 >100 µg/mL) and "Unfavorable" (EC150<40 µg/mL), based on NMR spectral data. Therefore, we developed a predictive classification model with two categories: Unfavorable (EC 150 < 40 µg/mL) and Favorable (EC150>100 µg/mL) using 117 chemicals. It is important to note that, due to the lack of data, the developed model does not address nor predict the class of chemicals with a 40 ≥ EC150 ≥ 100 value. Model quality parameters were estimated using the cross-validation method. The cross-validation method consists in building several sub-models, each time using only a fraction (6/7 in our case) of the chemicals

For the calculated model, we find the following values for the main estimators of prediction quality:

Kapaa	0.70
Accuracy	0.86
Sensibility	0.88
Specificity	0.82
Negative Predictive Value (NPV)	0.80
Positive Predictive Value (PPV)	0.89

Table 1: Main estimators of prediction quality:

(**Kappa**: consistency between the model prediction and the measured biological value; **Accuracy**: percentage true values; **Specificity:** percentage True NEGATIVE; **Sensitivity** :percentage True POSITIVE; **Negative Predictive Value (NPV):** probability that following a NEGATIVE prediction, that NatEx will be truly NEGATIVE; **Positive Predictive Value (PPV):** probability that following a POSITIVE prediction, that NatEx will truly be POSITIVE.)

The model's measured performance indicators show that the overall prediction is viable. The kappa value obtained (K=0.7) also shows that the concordance rate between predictions and real values is reliable.

3.3 Confidence of the approach:

To enhance prediction reliability, we assessed the model's confidence level for classifying samples into "Favorable" and "Unfavorable" categories. This involved evaluating model performance at various probability thresholds (50%, 60%, 65%, 70%, 75%, 80%) as shown in Table 2. A threshold of 65% proved optimal, balancing high sensitivity and specificity (>91%), while reducing false positives from 20% to 9% and maintaining acceptable accuracy and positive predictive value (PPV). This approach introduced an "inconclusive" call for 17% of the chemicals, resulting in an overall response rate of 83%.

	50% (by default)	60%	65%	70%	75%	80%
N=	117	103	97	91	83	75
Response (%)	100%	88%	83%	78%	71%	64%
Misclassified (%)	15%	14%	5%	5%	4%	4%
False Neg	9	6	2	1	1	1
False pos	9	8	3	3	2	2
Specificity	80%	79%	91%	88%	91%	91%
Sensitivity	88%	91%	97%	98%	98%	98%
Accuracy	85%	86%	95%	95%	96%	96%
Kappa	67%	71%	88%	89%	90%	90%
VPP	88%	88%	95%	95%	96%	96%
VPN	80%	84%	94%	96%	95%	95%

Table 2: Model performance at various probability thresholds (50%, 60%, 65%, 70%, 75%, 80%)

This 65% threshold value on class prediction probabilities has enabled us to establish three levels of confidence: High confidence level [1- 0,65], Low confidence level [0,65-0,51] and No conclusion [0,50; 0,50]. Considering the inclusion of the probabilistic confidence, the predictive model of EC150 (U-SENS test) from the 117 NMR data could be represented as plotted in Figure 2 and interpreted in Table 3.

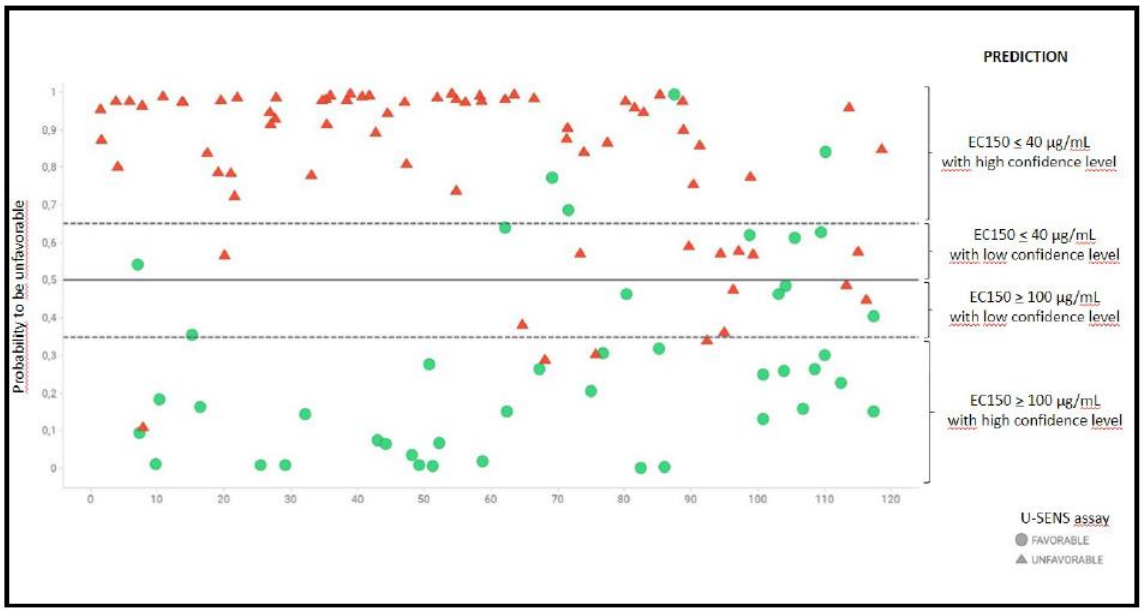


Figure 2: Representation of the predictive model of EC150 (U-SENS test) from the 117 NMR data

Probability to be UNFAVORABLE (EC150 < 40 µg/mL)	**Probability to be FAVORABLE (EC150 > 100 µg/mL)**	**CONCLUSION**
[1; 0,65]	[0,35; 0]	UNFAVORABLE / EC150 < 40 µg/mL with high confidence level (PPV = 95%)
]0,65; 0,51]	[0,49; 0,35[	UNFAVORABLE / EC150 < 40 µg/mL with low confidence level (PPV = 64%)
0,50	0,50	No conclusion
[0,49; 0,35[	]0,65; 0,51]	FAVORABLE / EC150 > 100 µg/mL with low confidence level (NPV = 44%)
[0,35; 0]	[1; 0,65]	FAVORABLE / EC150 > 100 µg/mL) with high confidence level (NPV = 94%)

Table 3: Conclusions of the predictive model of EC150 (U-SENS test) from NMR data

3.4. NMR Spectral Analysis

When we analyse the average proton NMR spectra of the samples (classified with a high degree of confidence >90%) from the ‘Favorable’ and ‘Unfavorable’ classes in the database, five zones of differentiation clearly appear (Figure 4) when comparing the two classes: zone 1 (0.4->2.4 ppm), zone 2 (3.8->5.6 ppm), zone 3 (5.8-7.3 ppm) zone 4 (8.6->9.8 ppm) and zone 5 (11.5->13 ppm). Each zone represents a specific signal emitted by the protons of the different chemical functions present in the molecules of the natural extracts. These chemical functions are probably responsible for alerting skin sensitization to the natural extracts. These

observations clearly show the relationship between NMR spectral data and biological data from U-SENS.

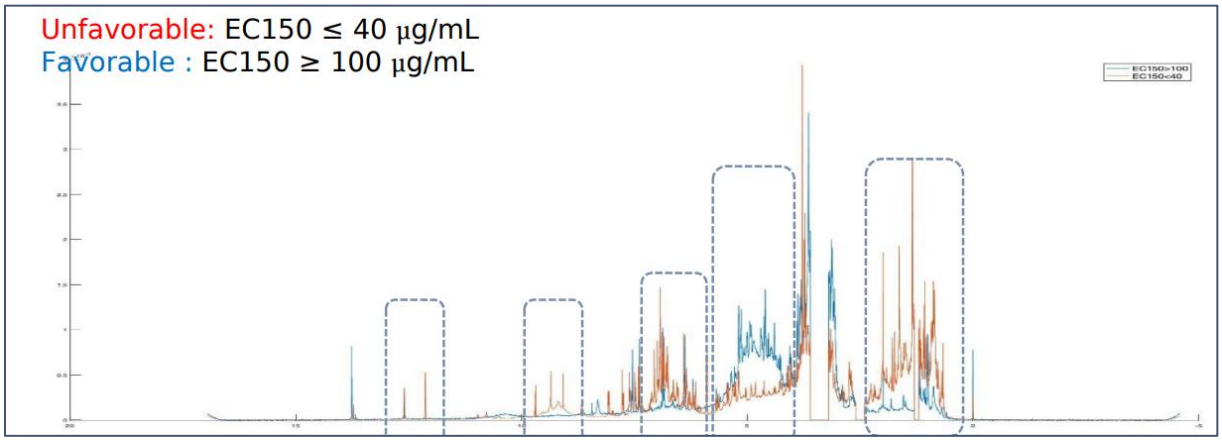


Figure 3: Average NMR proton spectra of the botanical extracts

4. Discussion

This study highlights the promising potential of combining NMR spectral fingerprints and multivariate analysis to predict the skin sensitization potential of natural extracts, effectively streamlining cosmetic ingredient safety assessments. This approach correlates strongly with in vitro U-SENS™ assay results, offering a faster and more efficient alternative. Specifically, PCA analysis demonstrates clear separation between favorable and unfavorable U-SENS™ profiles based on NMR fingerprints, linking chemical composition to sensitization potential. While direct EC150 prediction proved complex, an OPLS classification model successfully categorizes extracts into "Favorable" (EC150 > 100 µg/mL) and "Unfavorable" (EC150 < 40 µg/mL) groups with impressive accuracy (0.86) and sensitivity (0.88). Furthermore, a confidence threshold enhances the model's predictive power, acknowledging the inherent variability of natural extracts and providing a more nuanced risk assessment. This valuable in silico pre-screening tool allows researchers to prioritize promising natural extracts early in product development, significantly reducing the time and cost associated with traditional safety testing. Importantly, it's intended to complement, not replace, the validated in vitro U-SENS™ test for definitive safety assessment.

5. Conclusions

We have developed a new digital screening tool to help predict the sensitization profile of natural plant extracts. This new approach is based on the correlation of NMR fingerprints and biological activity measurements from the U-SENS™ in vitro test. This has allowed us to develop a predictive model to classify the extracts into two categories: the Favorable class (EC150 > 100 µg/mL) and the Unfavorable class (EC150 < 40 µg/mL). It should not be seen

as a substitute of the current U-SENS method but may be performed upfront at an early stage of the ranking of candidates to identify the best candidates.

References

1. Natural Cosmetics (pp. 100–108). (2023). https://doi.org/10.9734/bpi/mono/978-81-19315-51-2/ch11
2. Hirata, D., Rocha, E., Nogueira, R. A., Rezende, L. C. S. H., & Felipe, D. (2022). Natural and organic cosmetics: Beneficial properties for the environment and health. Int. J. Adv. Eng. Res. Sci., 9(11), 276–281. https://doi.org/10.22161/ijaers.911.34
3. Harvey, S. (2014). Producers meet growing appetite for green cosmetics : natural, organic & environmental. S. Af. Pharm. Cosm. Rev., 41(2), 23–26. https://journals.co.za/content/im_sapcr/41/2/EJC149190
4. Dieterle F., Ross A., Schlotterbeck G., and Senn H.. « Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1H NMR Metabonomics ,Anal. Chem. 2006, 78, 4281-4290
5. OECD. (2023b). Test No. 442E: In vitro skin sensitisation: In vitro skin sensitisation assays addressing the key event on activation of dendritic cells on the adverse outcome pathway for skin sensitisation, OECD guidelines for the testing of chemicals, section 4. Editions OCDE. https://doi.org/10. 1787/9789264264359-en
6. OECD. (2022). Test no. 442D: In vitro skin sensitisation: ARE-Nrf2 luciferase test method, OECD guidelines for the testing of chemicals, section 4. Editions OCDE. https://doi.org/10.1787/9789264229822-en
7. Piroird C., Ovigne J.M., Rousset F., Martinozzi-Teissier S., Gomes C., Cotovio J., Alépée N. « The Myeloid U937 Skin Sensitization Test (U-SENS) addresses the activation of dendritic cell event in the adverse outcome pathway for skin sensitization", Toxicology in Vitro, 2015, 29, 901-916
8. Martin M., Legat b., Leenders J., Vanwinsberghe J., Rousseau R., Boulanger B.,Eilers P.H.C., Tullio P.D., Govaerts P., "PepsNMR for 1H NMR metabolomic data pre-processing", Anal. Chim. Acta, 2018,1019, 1-13
9. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Anal Chem. 2006, 78(13):4281-90. doi: 10.1021/ac051632c. PMID: 16808434.
10. Corsaro, C.; Vasi, S.; Neri, F.; Mezzasalma, A.M.; Neri, G.; Fazio, E. NMR in Metabolomics: From Conventional Statistics to Machine Learning and Neural Network Approaches. *Appl. Sci.* **2022**, *12*, 2824. https://doi.org/10.3390/app12062824
11. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philos Trans A Math Phys Eng Sci. 2016, 374(2065):20150202. doi: 10.1098/rsta.2015.0202. PMID: 26953178; PMCID: PMC4792409.
12. Trygg, J. and Wold, S., Orthogonal Projections to Latent Structures (O-PLS). 2002, J. Chemometrics, 16, 119-128. http://dx.doi.org/10.1002/cem.695