RSC Advances

PAPER

View Article Online
View Journal | View Issue



Cite this: *RSC Adv.*, 2018, **8**, 27037

# Real-time simultaneous detection of microbial contamination and determination of an ultra low-content active pharmaceutical ingredient in tazarotene gel by near-infrared spectroscopy

Qian Xie,† Ruanqi Wu,† Xiaoxiao Zhong,† Yanhong Dong and Qi Fan *

This paper proposes and proves a real-time and non-destructive strategy for sensitive and simultaneous detection of microbial contamination and determination of an ultra low-content active pharmaceutical ingredient in tazarotene gel by near-infrared (NIR) spectroscopy. In this experiment, 88 samples of tazarotene gel (0.41–0.65 mg $g^{-1}$ of tazarotene) were prepared using the standard addition method. Among them, 47 samples were inoculated with 50 μl of different concentrations of *Escherichia coli* (*E. coli*) DH5a in Luria–Bertani (LB) broth to give 1–4 log CFU $g^{-1}$ of *E. coli* DH5a in the gel, 6 samples with 50 μl of LB broth, and 35 samples with nothing. Based on the gel NIR transflectance spectra, *E. coli* DH5a in the gel was detected by the counter propagation artificial neural network (CP-ANN) model with a classification accuracy of 100.0%, while tazarotene in the gel was simultaneously determined by the partial least squares regression (PLS) model with a root mean square error of cross-validation of 0.0232 mg $g^{-1}$. Furthermore, 9 samples of real tazarotene gel were used to verify the practicality of the established NIR spectroscopy. The developed NIR strategy can be used to correctly and quickly release the pharmaceutical gels, required for sensitive and simultaneous control of microbial contamination and the active pharmaceutical ingredient (API) content, to the next stage.

Received 10th April 2018
Accepted 25th June 2018

DOI: 10.1039/c8ra03079k

rsc.li/rsc-advances



## Introduction

Pharmaceutical products (PPs) are at risk of microbial contamination until they are used. Therefore, the detection of microbial contamination in PPs is of fundamental interest for pharmaceutical analysts. Up to now, the pharmacopoeia methods for the detection of microbial contamination in PPs require an incubation period of at least three days.[1–5] Because of the rapid proliferation of microorganisms, the pharmacopoeia methods cannot be used to monitor in real time the microbial contamination in PPs, especially in the manufacturing process. In addition, the pharmacopeia methods also have other drawbacks, such as tedious sample preparation and testing, sample destruction, microbial pollution, and being one-parameter.

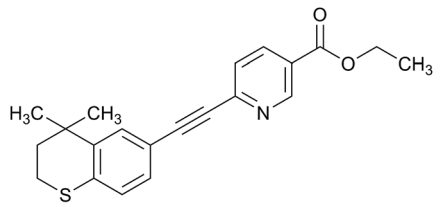
Near-infrared (NIR) spectroscopy, which can characterize a variety of properties of an analyte containing the X–H groups (X = C, N, O, S), is becoming a popular analytical method for drug quality control because it is sample-nondestructive, reagent-free, real-time, available for process control, and multi-parameter.[6–8] The possibility to detect and quantify bacterial contaminations in liquid PPs with fiber optic NIR spectroscopy has been explored by Quintelas *et al.* (*E. coli* lowest content 9.6 CFU $ml^{-1}$).[9]

However, we could not find, so far, the real-time sensitive NIR spectroscopy for the detection of microbial contamination in pharmaceutical gels, although the active pharmaceutical ingredient (API) in pharmaceutical gels has been determined using NIR spectroscopy by Blanco *et al.*[10] (API content 1.25–3.75% w/w), Rosas *et al.*[11] (API content 1.0–1.5% w/w), and Donoso *et al.*[12] (API content about 0.5–2% w/w). And the tazarotene content in gels has been determined using high performance liquid chromatography (HPLC) in Chinese Pharmacopoeia[5] (API content 0.05–0.1% w/w), using high performance thin layer chromatography (HPTLC) by Patel *et al.*[13] (API content 0.05% w/w), using ultraviolet (UV) spectrophotometry by Jogarami *et al.*[14] (API content 0.1% w/w), but not using NIR spectroscopy. Pharmaceutical gels are one of the important semi-solid PPs for topical application to the skin or the body cavity.[5] The unique advantages of pharmaceutical gels include reducing greatly the systemic adverse effects, escaping from the first-pass metabolism, and noninvasive and convenient use.[15] For example, tazarotene gel (tazarotene chemical structure in Fig. 1; labeled content 0.05% w/w, that is, 0.5 mg $g^{-1}$) is well accepted for treatment with psoriasis.[16]

Accordingly, in this paper, the feasibility of real-time, non-destructive, sensitive and simultaneous analysis of *E. coli*

*School of Pharmacy, Chongqing Medical University, Chongqing 400016, China. E-mail: fanqi787@cqmu.edu.cn; Tel: +86 23 6848 5161*

† These authors contributed equally to this article.




tazarotene

Fig. 1 The chemical structure of tazarotene.

DH5a contamination and ultra low-content tazarotene in tazarotene gel was evaluated by using Fourier transform NIR spectroscopy with chemometric techniques, mainly on the basis of the characteristic absorptions of the X–H groups of *E. coli* DH5a and tazarotene. Using this real-time simultaneous NIR approach, pharmaceutical gels can be rapidly analyzed and released to the next stage. This work can extend the NIR applications in the simultaneous analysis of both microbial contamination and ultra low-content API in a pharmaceutical gel.

# Experimental

### Samples and reference values

A total of 88 sterile samples of tazarotene gel, including nine API contents 0.41, 0.43, 0.46, 0.48, 0.51, 0.54, 0.58, 0.61, and 0.65 mg $g^{-1}$ (82–130% of 0.5 mg $g^{-1}$), were prepared with tazarotene (assay 99.7%) and jojoba oil by using standard addition method, filled into the lacquer-coated aluminum tubes, and sealed by an authorized manufacturer, Chongqing Huapont Pharmaceutical Co., Ltd. (Chongqing, China).

Before the NIR spectra acquirement, about 7.5 g of each sterile sample of translucent tazarotene gel in a lacquer-coated aluminum tube was squeezed into an empty disposable sterile ziplock transparent plastic bag (8 cm × 12 cm) and inoculated with 50 μl of different concentrations of *E. coli* DH5a in Luria–Bertani (LB) broth, 50 μl of LB broth, or nothing. Both *E. coli* DH5a and LB broth were supplied by the Key Laboratory of Molecular Biology on Infectious Diseases, Chinese Ministry of Education, Chongqing Medical University (Chongqing, China). After in time and gently pressing to remove the most of trapped air, the bag was zipped immediately and kneaded rapidly and gently to mix the contents. The above-mentioned operations were conducted in a class A clean bench in a class B clean room. Finally, 47 positive samples with 50 μl of different concentrations of *E. coli* DH5a in LB broth (1–4 log CFU $g^{-1}$ of *E. coli* DH5a in the gel) and 41 negative samples without *E. coli* DH5a (including 6 samples with 50 μl of LB broth and 35 samples with nothing) were obtained.

Furthermore, 9 samples of real tazarotene gel produced by Chongqing Huapont Pharmaceutical Co., Ltd. (Chongqing, China) were used to verify the practicality of the established NIR spectroscopy. And 9 real samples were proven uncontaminated using the Chinese pharmacopoeia method.[5] The tazarotene contents were determined as 0.49–0.52 mg $g^{-1}$ (98–104% of 0.5 mg $g^{-1}$) using the HPLC.[5]

### Near-infrared instrument and spectral measurements

The near-infrared instrument, used in this experiment, was Antaris II FT-NIR analyzer (Thermo Fisher Scientific, USA), which was equipped with an indium gallium arsenide detector, an integrating sphere, and a transflectance accessory. The instrument was controlled with the software RESULT 3.0. The chemometric analysis software was TQ Analyst 8.0 (Thermo Fisher Scientific, USA) and Matlab 8.0 (The Math Works, USA).

After the sample in bag was laid flat and squeezed to fill up the space between the window of integrating sphere and the transflectance accessory, the sample NIR spectrum was immediately measured within the range of 10 000–4000 $cm^{-1}$ using the selected resolution, co-averaged scans, and specimen holder (transflectance accessory) thickness. The resolution was selected from 4, 8 and 16 $cm^{-1}$ for more sample information in less time; the co-averaged scans from 16, 32, 48, 64, and 128 for less noise in less time; and the specimen holder thickness from 0.5, 1, and 2 mm for moderate absorption. The spectrum of ambient background, prior to sample, was measured under the selected conditions to eliminate the interferences from $H_2O$ and $CO_2$ in the air on the sample spectrum.

In addition, the NIR spectra of *E. coli* DH5a, LB broth, tazarotene, jojoba oil, and the blank bag were also measured in the same way as the tazarotene gel.

### Detection of *E. coli* DH5a contamination in tazarotene gel

In the detection, both linear and nonlinear models, the discriminant analysis (DA) and counter propagation artificial neural network (CP-ANN) models, were used to validate each other. In the above-mentioned two models, 48 calibration samples were consisted of 24 positive samples with 50 μl of different concentrations of *E. coli* DH5a in LB broth (1–4 log CFU $g^{-1}$ of *E. coli* DH5a in the gel) and 24 negative samples, which include 3 negative samples with 50 μl of LB broth and 21 negative samples with nothing. And 40 validation samples were consisted of 23 positive samples with 50 μl of different concentrations of *E. coli* DH5a in LB broth (1–4 log CFU $g^{-1}$ of *E. coli* DH5a in the gel), 3 negative samples with 50 μl of LB broth, and 14 negative samples with nothing.

The sample spectra were preprocessed separately for the DA and CP-ANN models. The preprocessing techniques were selected from no preprocessing (NP), multiplicative scatter correction (MSC) or standard normal variate (SNV) for eliminating the interference of particle size and compactness, derivative for deducting the background and separating the overlapping signals, and smoothing for denoising. The spectral sub-ranges, respectively for the DA and CP-ANN models, were selected mainly on the basis of the characteristic absorptions of the X–H groups of *E. coli* DH5a in order to reduce irrelevant variables and data redundancy. The DA and CP-ANN models were established separately by using the scores of selected principal components (PCs) of spectral data for calibration samples and the reference values. The cumulative contribution rate of the selected PCs was over 85%.[17]

The DA and CP-ANN models were validated respectively by using the scores of selected PCs of spectral data for validation

samples and the reference values. The DA model was evaluated with the classification accuracy of calibration (CAC) and validation (CAV), and the CP-ANN model with the classification accuracy of calibration (CAC), validation (CAV), and cross-validation (CACV).

### Determination of ultra low-content tazarotene in tazarotene gel

In the determination, the partial least squares regression (PLS) model was used. 44 calibration samples, including 20 positive samples with 50 μl of different concentrations of *E. coli* DH5a in LB broth, 3 negative samples with 50 μl of LB broth, and 21 negative samples with nothing, were distributed in the range of 0.41–0.65 mg $g^{-1}$ of tazarotene in the gel. And 44 validation samples, including 27 positive samples with 50 μl of different concentrations of *E. coli* DH5a in LB broth, 3 negative samples with 50 μl of LB broth, and 14 negative samples with nothing, were also distributed in the range of 0.41–0.65 mg $g^{-1}$ of tazarotene contents.

The sample spectra were preprocessed by the chemometric techniques selected from NP, MSC or SNV, derivative, smoothing, mean centering (MC), and variance scaling (VS). The spectral sub-ranges were selected mainly on the basis of the characteristic absorptions of C–H groups in tazarotene. The PLS model was built by using the scores of selected factors of spectral data for calibration samples and the reference values. The used factors were confirmed by the smallest root mean square error of cross-validation (RMSECV).

The PLS model was validated by using the factor scores of spectral data for validation samples and the reference values. And the PLS model was evaluated with the correlation coefficient of calibration ($R_c$), validation ($R_v$), and cross-validation ($R_{cv}$) between reference values and predicted values; the root mean square error of calibration (RMSEC), validation (RMSEV), and RMSECV; and the bias.

## Results and discussion

### Spectral measurements

The NIR transflectance spectra of 97 translucent samples of tazarotene gel, including 88 simulated samples and 9 real samples, were measured within the range of 10 000–4000 $cm^{-1}$ using the resolution of 8 $cm^{-1}$, 64 co-averaged scans, the specimen holder thickness of 1 mm, and the data collection time of about 31.65 seconds, as shown in Fig. 2a. In Fig. 2a, the brown lines represent 9 real samples, the blue lines 35 samples with nothing, the green lines 6 samples with 50 μl of LB broth, and the red lines 47 samples with 50 μl of different concentrations of *E. coli* DH5a in LB broth.

And the NIR transflectance spectra of *E. coli* DH5a, LB broth, tazarotene, jojoba oil, and the blank bag are given in Fig. 2b respectively in red, green, blue, purple, and dark yellow. It has been demonstrated that the above-mentioned five materials are NIR responsive and can be characterized with the NIR transflectance spectroscopy.

### Detection of *E. coli* DH5a contamination in tazarotene gel

For the detection, the vital preprocessing techniques are listed in Table 1. It can be found from Table 1 that the performances of both DA and CP-ANN models are improved with MSC or SNV. That is, the influence of scattering on the transflectance spectra cannot be ignored in the detection for the nature of *E. coli* DH5a and the translucent gel. Moreover, the performances of both DA and CP-ANN models are not improved with smoothing. We speculate that the application of 64 co-averaged scans has greatly reduced noise in the NIR spectra.

The red line in Fig. 2b indicates that the informative spectral sub-ranges for *E. coli* DH5a are 9000–8000 $cm^{-1}$, 7300–4800 $cm^{-1}$, and 4450–4100 $cm^{-1}$. Since two strong overtones of O–H in $H_2O$ are respectively around 6900 $cm^{-1}$ and 5150 $cm^{-1}$, two portions 7300–6400 $cm^{-1}$ and 5400–4800 $cm^{-1}$ are removed from 7300–4800 $cm^{-1}$. On the other hand, the sub-range of

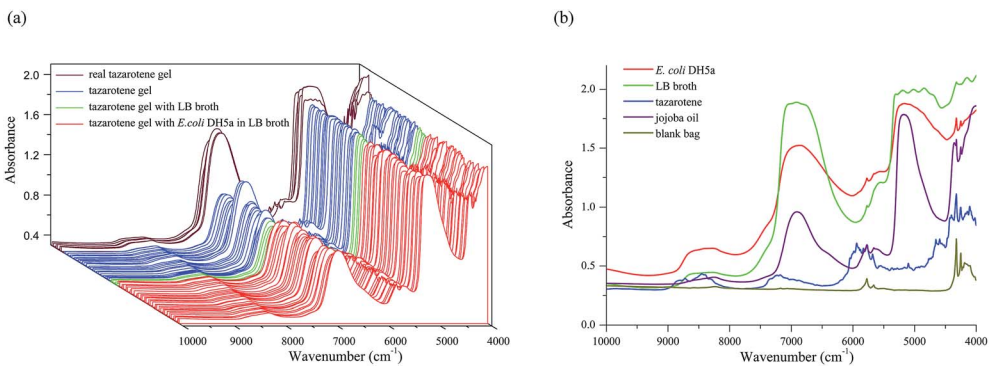



Fig. 2 Raw near-infrared (NIR) transflectance spectra. (a) For 97 samples of tazarotene gel in the transparent plastic bag: the brown lines show 9 real samples, the blue lines 35 samples with nothing, the green lines 6 samples with Luria–Bertani (LB) broth, and the red lines 47 samples with *Escherichia coli* (*E. coli*) DH5a in LB broth. (b) For *E. coli* DH5a, LB broth, tazarotene, jojoba oil, and nothing respectively in the transparent plastic bag: the red line represents *E. coli* DH5a, the green line LB broth, the blue line tazarotene, the purple line jojoba oil, and the dark yellow line the blank bag.

Table 1 Vital preprocessing techniques, spectral sub-ranges and corresponding performances of the DA and CP-ANN models[a]

| Item | Model | Preprocessing | Spectral sub-ranges ($cm^{-1}$) | Number of PCs/cumulative contribution rate (%) | CAC (%) | CAV (%) | CACV (%) |
|---|---|---|---|---|---|---|---|
| DA | 1 | NP | **9000–8000** | 3/100.0 | 95.8 | 95.0 | |
| | **2** | **MSC** | **6400–5400** | **3/99.9** | **100.0** | **100.0** | |
| | **3** | **SNV** | | **3/99.9** | **100.0** | **100.0** | |
| | 4 | FD | | 3/97.7 | 58.3 | 62.5 | |
| | 5 | SD | | 3/70.3 | 72.9 | 82.5 | |
| | 6 | SGS | | 3/100.0 | 95.8 | 95.0 | |
| | 7 | NDS | | 3/98.7 | 58.3 | 65.0 | |
| | 8 | MSC + FD + SGS | | 3/96.1 | 83.3 | 82.5 | |
| | 9 | MSC | 9000–8000<br>6400–5400<br>4450–4100 | 3/99.5 | 97.9 | 95.0 | |
| | 10 | MSC | 9000–8000<br>7300–4800<br>4450–4100 | 3/99.8 | 91.7 | 87.5 | |
| CP-ANN | 11 | NP | **9000–8000** | 2/99.9 | 100.0 | 92.5 | 96.0 |
| | **12** | **MSC** | **6400–5400** | **2/99.7** | **100.0** | **100.0** | **100.0** |
| | 13 | SNV | | 2/99.6 | 100.0 | 100.0 | 100.0 |
| | 14 | SGS | | 2/99.9 | 100.0 | 90.0 | 96.0 |
| | 15 | NDS | | 2/98.0 | 98.0 | 67.5 | 73.0 |
| | 16 | MSC + FD + SGS | | 2/94.7 | 100.0 | 92.5 | 83.0 |

[a] DA for discriminant analysis; CP-ANN counter propagation artificial neural network; PC principal component; CAC classification accuracy of calibration; CAV classification accuracy of validation; CACV classification accuracy of cross-validation; MSC multiplicative scatter correction; SNV standard normal variate; NP no preprocessing; FD first derivative; SD second derivative; SGS Savitzky–Golay smoothing; NDS Norris derivative smoothing.

4450–4100 $cm^{-1}$ is also not used for modeling because it may include the strong combination bands of alkyl, aromatic and the polymer of the plastic bag. Finally, the spectral sub-ranges for modelling were 9000–8000 $cm^{-1}$ and 6400–5400 $cm^{-1}$. The above-mentioned sub-ranges can be mainly attributed to the overtones and combination bands of C–H in *E. coli* DH5a, such as the second overtones of alkyl near 8696–8264 $cm^{-1}$ and aromatic near 8834 $cm^{-1}$, and the first overtones of alkyl near 5882–5555 $cm^{-1}$ and aromatic near 6000 $cm^{-1}$.[18] It is noted that the selected spectral sub-ranges overlap partially with 7000–4000 $cm^{-1}$ used by Rodriguez-Saona *et al.*[19] to analyze bacterial contaminations in liquids, and 6000–5400 $cm^{-1}$ used by Quintelas *et al.*[9] to analyze *E. coli* contaminations in liquid PPs. Moreover, both the DA model 9 (using 4450–4100 $cm^{-1}$) and 10 (using 7300–6400 $cm^{-1}$, 5400–4800 $cm^{-1}$, and 4450–4100 $cm^{-1}$) have lower classification accuracies than the model 2 although all of them use MSC. This result proves that it is correct to remove 7300–6400 $cm^{-1}$, 5400–4800 $cm^{-1}$, and 4450–4100 $cm^{-1}$ from the spectral sub-ranges for modelling.

The performances of the DA and CP-ANN models are listed in Table 1. The optimized DA model is the model 2 or 3. It reaches the CAC 100.0% and the CAV 100.0% by using MSC or SNV, small spectral sub-ranges 9000–8000 $cm^{-1}$ and 6400–5400 $cm^{-1}$, and fewer PCs the first three (cumulative contribution rate 99.9%). Similarly, the optimized CP-ANN model is the model 12. It reaches the CAC 100.0%, the CAV 100.0%, and the CACV 100.0% by using MSC, 9000–8000 $cm^{-1}$ and 6400–5400 $cm^{-1}$, and the first two PCs (cumulative contribution rate 99.7%). Fig. 3 indicates that the model 2 and 12 distribute respectively the positive samples and the negative samples (with or without LB broth) of tazarotene gel logically and distinctly in two separate two-dimensional zones in both calibration and validation.

Therefore, the *E. coli* DH5a contamination in tazarotene gel (1–4 log CFU $g^{-1}$) can be in time, nondestructively, and sensitively detected with the linear DA model or the nonlinear CP-ANN model (using less PCs; with a better representational ability) based on the NIR spectra. And the specificity of the models is confirmed by three experimental results as below. First, the positive samples (with 50 μl of *E. coli* DH5a in LB broth) and the negative samples (with 50 μl of LB broth), with or without *E. coli* DH5a, are classified into two subgroups. Second, all negative samples (with 50 μl of LB broth or nothing), without *E. coli* DH5a, are classified in one subgroup. Third, two stronger overtones of LB broth than *E. coli* DH5a, respectively near 7300–6400 $cm^{-1}$ and 5400–4800 $cm^{-1}$, are not used to build the DA and CP-ANN models. That is, the recognition of *E. coli* DH5a in tazarotene gel is mainly based on the characteristic NIR absorptions of *E. coli* DH5a rather than LB broth and/or tazarotene gel. Moreover, in the validation and prediction, the false positive or false negative results were reduced because the number of positive and negative samples in the calibration samples is equal.

Comparing with the NIR method reported previously by Quintelas *et al.*[9] for analyzing *E. coli* contaminations in liquid PPs (the spectral sub-range for modelling 6000–5400 $cm^{-1}$),

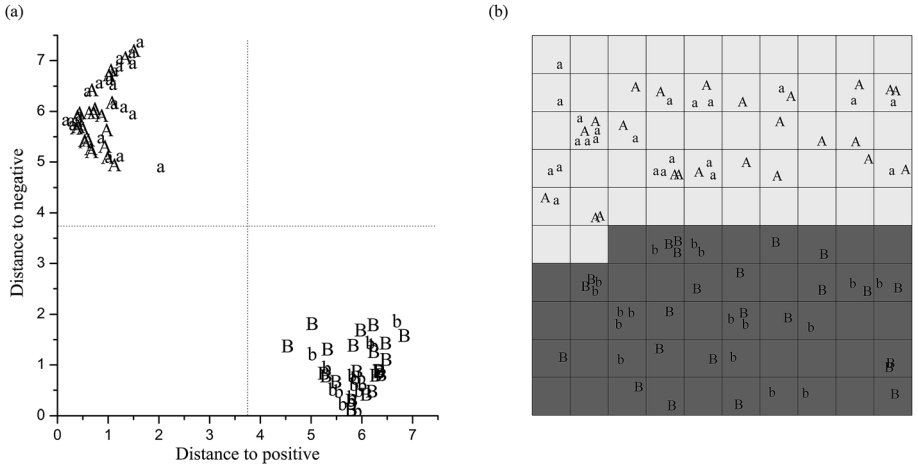



Fig. 3 Projection maps for the discriminant analysis (DA) and counter propagation artificial neural network (CP-ANN) models. The uppercase letter "A" and the lowercase letter "a" indicate positive samples in calibration and validation, separately; "B" and "b" negative samples in calibration and validation, respectively. (a) Illustrates the DA model. (b) Shows the CP-ANN model.

both the DA and CP-ANN models use the spectral information at first overtones 6400–5400 $cm^{-1}$ with the second overtones 9000–8000 $cm^{-1}$. Obviously, high-frequency regions are less affected by the overlapping bands than low-frequency regions. We speculate that it is beneficial to detect the *E. coli* DH5a contamination as low as 1 log CFU $g^{-1}$ using the second overtones of alkyl and aromatic 9000–8000 $cm^{-1}$ but not to use 5400–4800 $cm^{-1}$ and 4450–4100 $cm^{-1}$. The established DA and CP-ANN models are very important in practice because the pharmaceutical gel is inhomogeneous and its sample preparation requires much tedious operations than the liquid PPs.

## Determination of ultra low-content tazarotene in tazarotene gel

For the determination, the important preprocessing techniques are listed in Table 2. As can be seen from Table 2, the PLS model performances are significantly improved with the combination of SNV and MC, but not improved by smoothing. If not considering MC, the preprocessing technique for quantitative analysis, this situation is similar to that in the detection of *E. coli* DH5a contamination in tazarotene gel.

The blue line in Fig. 2b indicates that the informative spectral sub-ranges for tazarotene are 9100–7900 $cm^{-1}$, 7600–

Table 2 Vital preprocessing techniques, spectral sub-ranges and corresponding performances of the PLS model[a]

| Model | Preprocessing | Spectral sub-ranges ($cm^{-1}$) | Factor number | $R_c$ | $R_v$ | $R_{cv}$ | RMSEC (mg $g^{-1}$) | RMSEV (mg $g^{-1}$) | RMSECV (mg $g^{-1}$) | Bias (mg $g^{-1}$) |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | NP | **9100–7900** | 8 | 0.9688 | 0.9373 | 0.9345 | 0.0187 | 0.0257 | 0.0273 | 0 |
| 2 | MSC | **7600–6500** | 8 | 0.9761 | 0.9554 | 0.9506 | 0.0162 | 0.0226 | 0.0240 | 0 |
| 3 | SNV | **6400–5600** | 8 | 0.9761 | 0.9549 | 0.9477 | 0.0162 | 0.0231 | 0.0244 | 0 |
| 4 | SGS | **4700–4500** | 8 | 0.9688 | 0.9373 | 0.9345 | 0.0187 | 0.0257 | 0.0273 | 0 |
| 5 | MC | | 8 | 0.9743 | 0.9589 | 0.9343 | 0.0173 | 0.0212 | 0.0277 | 0 |
| 6 | FD + NDS | | 8 | 0.9566 | 0.9491 | 0.8653 | 0.0231 | 0.0327 | 0.0393 | 0 |
| 7 | SD + NDS | | 8 | 0.9182 | 0.7478 | 0.5669 | 0.0308 | 0.0572 | 0.0684 | 0 |
| 8 | MC + VS | | 7 | 0.9725 | 0.9358 | 0.9478 | 0.0178 | 0.0262 | 0.0247 | 0 |
| 9 | MSC + MC | | 7 | 0.9757 | 0.9542 | 0.9525 | 0.0160 | 0.0222 | 0.0236 | 0 |
| **10** | **SNV + MC** | | **7** | **0.9780** | **0.9667** | **0.9491** | **0.0154** | **0.0207** | **0.0232** | **0** |
| 11 | MSC + MC + VS | | 7 | 0.9739 | 0.9521 | 0.9464 | 0.0171 | 0.0233 | 0.0246 | 0 |
| 12 | SNV + MC + VS | | 7 | 0.9790 | 0.9571 | 0.9507 | 0.0162 | 0.0215 | 0.0235 | 0 |
| 13 | SNV + MC | 9100–7900<br>7600–7200<br>6150–5600<br>4700–4500 | 8 | 0.9745 | 0.9363 | 0.9336 | 0.0163 | 0.0267 | 0.0279 | 0 |

[a] PLS for partial least squares regression; $R_c$ correlation coefficient of calibration; $R_v$ correlation coefficient of validation; $R_{cv}$ correlation coefficient of cross-validation; RMSEC root mean square error of calibration; RMSEV root mean square error of validation; RMSECV root mean square error of cross-validation; NP no preprocessing; SNV standard normal variate; MSC multiplicative scatter correction; SGS Savitzky–Golay smoothing; FD first derivative; NDS Norris derivative smoothing; SD second derivative; MC mean centering; VS variance scaling.

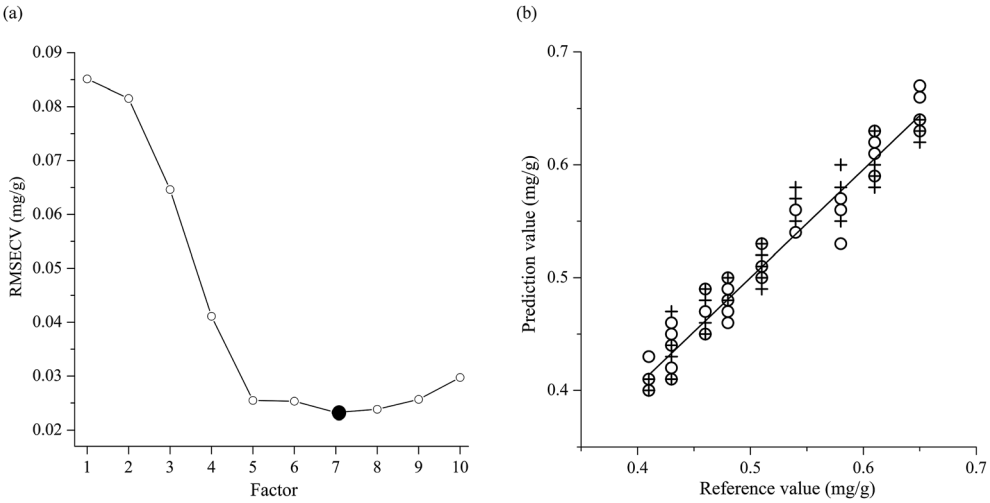



Fig. 4 Scatter plots for the partial least squares regression (PLS) model. (a) Shows the relationship between root mean square error of cross-validation (RMSECV) and factor number of the PLS model; the optimal factor number is highlighted by ●. (b) Shows the relationships between reference values and predicted values for calibration (○) and validation (+).

6500 $cm^{-1}$, 6400–5600 $cm^{-1}$, and 4700–4000 $cm^{-1}$. Because the sub-range of 4500–4000 $cm^{-1}$ may include the strong combination bands of alkyl, aromatic, and the polymer of the plastic bag, it is removed from 4700–4000 $cm^{-1}$. Finally, the spectral sub-ranges for modelling were 9100–7900 $cm^{-1}$, 7600–6500 $cm^{-1}$, 6400–5600 $cm^{-1}$, and 4700–4500 $cm^{-1}$. The above-mentioned sub-ranges can be mainly attributed to the overtones and combination bands of C–H in tazarotene. For example, the second overtones of $CH_3$ and $CH_2$ are about 8696–8264 $cm^{-1}$; and phenyl about 8834 $cm^{-1}$. The first overtones of $CH_3$ appear near 5905, 5876, and 5872 $cm^{-1}$; $CH_2$ in linear near 5680 $cm^{-1}$; $CH_2$ in cyclic near 6060, 5791, and 5697 $cm^{-1}$; and phenyl near 6000 $cm^{-1}$. The combination bands of $CH_3$ appear near 7355, 7263, and 4545–4500 $cm^{-1}$; $CH_2$ near 7186 and 7080 $cm^{-1}$; and phenyl and pyridyl near 4700–4500 $cm^{-1}$.[18] It is noted that the spectral sub-range of 7600–6500 $cm^{-1}$ might include the overtone of O–H in $H_2O$ around 6900 $cm^{-1}$. However, we can find from Table 2 that the portion around 6900 $cm^{-1}$ is useful to determine tazarotene in the gel, although it might overlap with the overtone of O–H in $H_2O$ around 6900 $cm^{-1}$, because the model 10 (use of the portion around 6900 $cm^{-1}$) has higher $R_c$, $R_v$, and $R_{cv}$, and smaller RMSEC, RMSEV, and RMSECV than the model 13 (no use of the portion around 6900 $cm^{-1}$).

The performances of the PLS model are listed in Table 2. It can be seen from Table 2 that the PLS model reaches the RMSECV 0.0232 mg $g^{-1}$ by optimizing into the model 10, as in Fig. 4. It is built by using the combination of SNV and MC, four spectral sub-ranges 9100–7900 $cm^{-1}$, 7600–6500 $cm^{-1}$, 6400–5600 $cm^{-1}$, and 4700–4500 $cm^{-1}$, and seven factors. It has the $R_c$ 0.9780, the $R_v$ 0.9667, the $R_{cv}$ 0.9491, the RMSEC 0.0154 mg $g^{-1}$, the RMSEV 0.0207 mg $g^{-1}$, the RMSECV 0.0232 mg $g^{-1}$, and the bias 0 mg $g^{-1}$. Compared with the model 10, the model 12 has the lower $R_v$ 0.9571 and the larger RMSEC 0.0162 mg $g^{-1}$, RMSEV 0.0215 mg $g^{-1}$, and RMSECV 0.0235 mg $g^{-1}$ although it has the higher $R_c$ 0.9790 and $R_{cv}$ 0.9507. Therefore, the model 12 is not as good as the model 10. Fig. 4a illustrates that the smallest RMSECV 0.0232 mg $g^{-1}$ of the model 10 is obtained by using seven factors. And Fig. 4b shows the linear relations between reference values and predicted values of tazarotene content in the gel in both calibration and validation of the model 10.

Consequently, the ultra low-content tazarotene in the gel (0.41–0.65 mg $g^{-1}$) can be sensitively, in time, and nondestructively determined by the PLS model based on the NIR spectra. The specificity of the model is demonstrated by the following experimental results. The ultra low-content tazarotene in the gel can be determined by the PLS model no matter if the gel is inoculated with 50 μl of *E. coli* DH5a in LB broth, 50 μl of LB broth, or nothing. That is, the determination of tazarotene in the gel is mainly based on the characteristic NIR absorptions of tazarotene. And the precision of the model is described with the relative standard deviations (RSD). The RSDs ($n = 3$) for the three API levels (0.41, 0.51, and 0.65 mg $g^{-1}$; that is, 82%, 102%, and 130% of 0.5 mg $g^{-1}$) are 1.43%, 1.15%, and 1.79%, respectively.

Comparing with the previously reported NIR methods for determining the API in pharmaceutical gel (about 0.5–2% w/w),[10–12] the PLS model uses the second overtones of $CH_3$, $CH_2$, and aromatic 9100–7900 $cm^{-1}$ but does not use 4500–4000 $cm^{-1}$. The high-frequency regions are less affected by the overlapping bands than low-frequency regions. On the other hand, the low-frequency band, such as the combination band of phenyl near 4050 $cm^{-1}$, has the very strong absorptions outside the linear range of the detector. We speculate that the combined use of the spectral information at 7600–6500 $cm^{-1}$ and 6400–5600 $cm^{-1}$ with the second overtones 9100–7900 $cm^{-1}$ is the reason for the ultra low-content tazarotene 0.041–0.065% w/w can be determined by the PLS model. The established PLS model is very important in practice because it can be used to determine the ultra low-content API in the pharmaceutical gel.

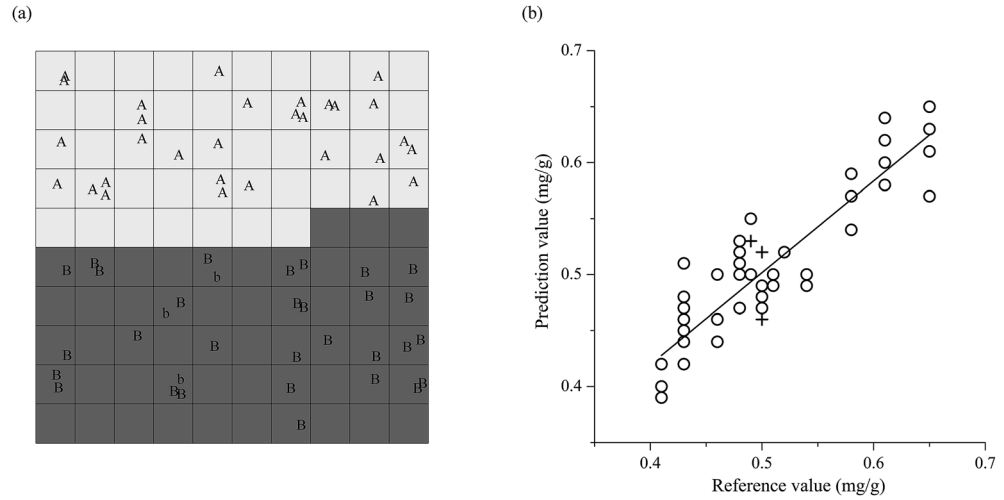



Fig. 5 (a) Projection map for the counter propagation artificial neural network (CP-ANN) model. The uppercase letter "A" and "B" indicate positive and negative samples in calibration, separately; the lowercase letter "b" real samples uncontaminated in validation. (b) Scatter plot for the partial least squares regression (PLS) model. It shows the relationships between reference values and predicted values for calibration (○) and validation (+).

### Detection of *E. coli* DH5a contamination and determination of tazarotene in real tazarotene gel

9 samples of real tazarotene gel (uncontaminated, API content 0.49–0.52 mg $g^{-1}$), produced by Chongqing Huapont Pharmaceutical Co., Ltd. (Chongqing, China), were used to verify the practicality of the established NIR spectroscopy. Among them, 6 samples were used as calibration samples and 3 samples as validation samples.

In the detection of *E. coli* DH5a contamination, 60 calibration samples were consisted of 30 positive samples with 50 μl of different concentrations of *E. coli* DH5a in LB broth (1–4 log CFU $g^{-1}$ of *E. coli* DH5a in the gel) and 30 negative samples, which include 3 negative samples with 50 μl of LB broth, 21 negative samples with nothing, and 6 real samples uncontaminated. Consequently, 3 real samples uncontaminated were correctly distributed by the CP-ANN model in the zone of negative samples, as shown in Fig. 5a.

In the determination of tazarotene in real tazarotene gel, 50 calibration samples were consisted of 44 simulated samples (API content 0.41–0.65 mg $g^{-1}$), which include 20 positive samples with 50 μl of different concentrations of *E. coli* DH5a in LB broth, 3 negative samples with 50 μl of LB broth, and 21 negative samples with nothing, and 6 real samples (API content 0.49–0.52 mg $g^{-1}$). Consequently, the tazarotene contents in 3 real samples were determined by the PLS model, as illustrated in Fig. 5b. And the three relative errors (REs) 8%, −8%, and 4% are acceptable.[5]

## Conclusions

A real-time and non-destructive method was developed to sensitively and simultaneously detect *E. coli* DH5a contamination and determine ultra low-content tazarotene in tazarotene gel (1–4 log CFU $g^{-1}$ of *E. coli* DH5a; 0.41–0.65 mg $g^{-1}$ of tazarotene) based on the NIR spectra of tazarotene gel. Using the real-time and multiparametric NIR approach, pharmaceutical gels, typically requiring tedious extractions, can be sensitively and quickly analyzed and released into the next stage. This work can extend the NIR applications in the sensitive and simultaneous monitoring of both microbial contamination and an ultra low-content API in a pharmaceutical gel.

## Conflicts of interest

There are no conflicts of interest to declare.

## Acknowledgements

We are grateful for the financial supports from the Municipal Science and Technology Committee of Chongqing [grant number cstc2012gg – yyjs10039] and from the District Science and Technology Committee of Yuzhong District of Chongqing [grant number 20120206] and the supports from Chongqing Huapont Pharmaceutical Co., Ltd. and from the Key Laboratory of Molecular Biology on Infectious Diseases, Chinese Ministry of Education, Chongqing Medical University.

## References

1 *The United States Pharmacopeia*, The United States Pharmacopeial Convention, Rockville, MD, 2015.
2 *The European Pharmacopoeia*, European Directorate for the Quality of Medicines & Healthcare, Strasbourg, 2013.
3 *The British Pharmacopoeia 2015*, The British Pharmacopoeia Commission, London, 2014.
4 *The Japanese Pharmacopoeia*, The Society of Japanese Pharmacopeia, Tokyo, 2016.
5 *The Chinese Pharmacopoeia*, Chinese Pharmacopoeia Commission, Beijing, 2015, vol. 4.

6 F. Shikata, S. Kimura, Y. Hattori and M. Otsuka, *RSC Adv.*, 2017, **7**, 38307–38317.
7 Q. Y. Luo, Y. H. Yun, W. Fan, J. H. Huang, L. X. Zhang, B. C. Deng and H. M. Lu, *RSC Adv.*, 2015, **5**, 5046–5052.
8 M. Otsuka, A. Koyama and Y. Hattori, *RSC Adv.*, 2014, **4**, 17461–17468.
9 C. Quintelas, D. P. Mesquita, J. A. Lopes, E. C. Ferreira and C. Sousa, *Int. J. Pharm.*, 2015, **492**, 199–206.
10 M. Blanco, M. Alcalá and M. Bautista, *Eur. J. Pharm. Sci.*, 2008, **33**, 409–414.
11 J. G. Rosas, M. Blanco, J. M. González and M. Alcalá, *J. Pharm. Sci.*, 2011, **100**, 4442–4451.
12 M. Donoso and E. S. Ghaly, *Pharm. Dev. Technol.*, 2006, **11**, 389–397.
13 M. R. Patel, R. B. Patel, J. R. Parikh and B. G. Patel, *Anal. Methods*, 2010, **2**, 275–281.
14 R. Jogarami, P. Jain and S. Sharma, *J. Pharm. Res.*, 2012, **5**, 2273–2275.
15 A. Vintiloiu and J. C. Leroux, *J. Controlled Release*, 2008, **125**, 179–192.
16 R. H. Foster, R. N. Brogden and P. Benfield, *Drugs*, 1998, **55**, 705–711.
17 W. Z. Lu, *Modern near infrared spectroscopy analytical technology*, China Petrochemical Press, Beijing, 2006.
18 J. Workman and L. Weyer, *Practical guide to interpretive near-infrared spectroscopy*, CRC Press, USA, 2008.
19 L. E. Rodriguez-Saona, F. M. Khambaty, F. S. Fry and E. M. Calvey, *J. Agric. Food Chem.*, 2001, **49**, 574–579.