Journal of Innovative Optical Health Sciences
Vol. 7, No. 4 (2014) 1450005 (10 pages)
© The Authors
DOI: 10.1142/S1793545814500059

World Scientific
www.worldscientific.com

Variable selection in near infrared spectroscopy for quantitative models of homologous analogs of cephalosporins

Yan-Chun Feng[*,‡], Zhen Ni[†] and Chang-Qin Hu[*,§]
[*]*National Institutes for Food and Drug Control (NIFDC)*
Beijing 100050, P. R. China

[†]*Tibet Institute for the Control of Food and Pharmaceutical Products*
Lhasa 850000, P. R. China
[‡]*fyc@nifdc.org.cn*
[§]*hucq@nifdc.org.cn*

Received 28 June 2013
Accepted 29 October 2013
Published 12 December 2013

Two universal spectral ranges (4550–4100 cm^{-1} and 6190–5510 cm^{-1}) for construction of quantitative models of homologous analogs of cephalosporins were proposed by evaluating the performance of five spectral ranges and their combinations, using three data sets of cephalosporins for injection, i.e., cefuroxime sodium, ceftriaxone sodium and cefoperazone sodium. Subsequently, the proposed ranges were validated by using eight calibration sets of other homologous analogs of cephalosporins for injection, namely cefmenoxime hydrochloride, ceftezole sodium, cefmetazole, cefoxitin sodium, cefotaxime sodium, cefradine, cephazolin sodium and ceftizoxime sodium. All the constructed quantitative models for the eight kinds of cephalosporins using these universal ranges could fulfill the requirements for quick quantification. After that, competitive adaptive reweighted sampling (CARS) algorithm and infrared (IR)–near infrared (NIR) two-dimensional (2D) correlation spectral analysis were used to determine the scientific basis of these two spectral ranges as the universal regions for the construction of quantitative models of cephalosporins. The CARS algorithm demonstrated that the ranges of 4550–4100 cm^{-1} and 6190–5510 cm^{-1} included some key wavenumbers which could be attributed to content changes of cephalosporins. The IR–NIR 2D spectral analysis showed that certain wavenumbers in these two regions have strong correlations to the structures of those cephalosporins that were easy to degrade.

Keywords: Near infrared spectroscopy; cephalosporins; quantitation; spectral range selection.

[§]Corresponding author.

1. Introduction

Variable or spectral range selection is a critical step for model construction using near infrared (NIR) spectroscopy. It is now widely accepted that better prediction can be obtained using the selected spectral ranges rather than the full spectrum.[1] There are a multitude of approaches available for variable selection such as simulated annealing (SA),[2] genetic algorithms (GAs),[3] uninformative variable elimination (UVE),[4] interval partial least squares (iPLS),[5] successive projections algorithm (SPA),[6] competitive adaptive reweighted sampling (CARS),[7] ant colony optimization algorithm (ACO),[8] etc. Zhou *et al.* reviewed the work in the area of variables selection methods using NIR spectroscopy.[9] Most of these algorithms are based on statistics related to the model's performance, e.g., root mean square error of cross validation (RMSECV). As the NIR spectra emerge from overtones and combinations of fundamental mid-infrared (MIR) absorptions, the NIR bands become naturally broad and overlapping, while their intensity is 10–50 times less than their corresponding MIR bands; thus it is difficult to accurately determine relationships between the spectral ranges and the chemical information of the analytes. Therefore, only a few analysts use the structural information of analytes to select NIR spectral variables. Hao *et al.*[10] proved that in a complex mixture, it is impossible to entirely eliminate the interferences from other constituents by variable selection; however, if the selected spectral ranges correlate with the analyte of interest, the predictive ability of the NIR model can be dramatically improved. Liu *et al.*[11] found that the combined use of specific NIR spectral regions, related to both structural characteristics and content, could lead to better NIR prediction when they constructed a quantitative model of gentamicin injection. Obviously, incorporating chemical information by representing NIR regions for different analytes is helpful to select spectral ranges.

Cephalosporins (Fig. 1) are a major group of β-lactam antibiotics, which are the most frequently prescribed antibiotics in clinical medicine. However, the instability of cephalosporins makes them easily degradable during the process of production and storage. Their degradation reactions mainly take place at β-lactam ring (ring cleavage reaction) or its 3-substituent (R_3 hydrolysis).[12,13] Hence, a series of universal NIR quantitative models of cephalosporins were setup to monitor their quality during their circulation in China.[14–16] After studying the performance of these universal quantitative models, certain empirical NIR spectral ranges were determined for quantitative models constructed for cephalosporins. In this study, we show these spectral regions, and subsequently use the IR–NIR two-dimensional (2D) correlation spectroscopy to determine the relationship between the chemical structure of cephalosporins and their NIR spectra. CARS method was also used to validate the relationship between these spectral ranges and the percent contents of cephalosporins.

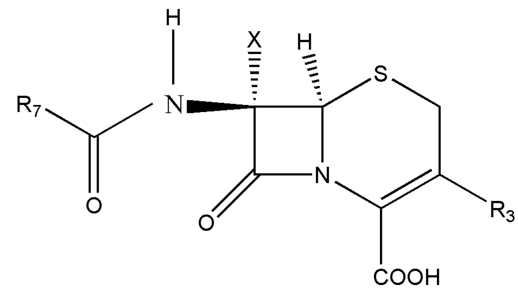


Fig. 1. The chemical structure of cephalosporins.

2. Experimental

2.1. *Samples*

All the samples (cefuroxime sodium, ceftriaxone sodium, cefoperazone sodium, cefmenoxime hydrochloride, ceftezole sodium, cefmetazole, cefoxitin sodium, cefotaxime sodium, cefradine, cephazolin sodium and ceftizoxime sodium for injection) used in this study were collected from Chinese market by the national institutes for food and drug control in 2006. Three batches of cefuroxime sodium, three batches of ceftriaxone sodium and four batches of cefoperazone sodium were used for the acceleration experiment. The concentrations of all these samples were determined by HPLC method according to Chinese Pharmacopoeia 2005 Edition, and expressed as percentage.

2.2. *Acceleration experiment*

All the samples for acceleration experiment were maintained in artificial climate chambers with controlled temperature and humidity. The accelerated storage condition and frequency of sampling for content determination are described in Table 1.

Table 1. Conditions for acceleration experiment.

Powder injection	Storage condition	Sampling frequency (day)
Cefuroxime sodium	T = 60°C, RH = 40%	2, 8, 17
Ceftriaxone sodium	T = 80°C, RH = 50%	2, 4, 8
Cefoperazone sodium	T = 60°C, RH = 45%	0, 11

2.3. *Instruments and data acquisition*

The NIR spectrometer, Model EQUINOX55 (Bruker Optik GmbH, Ettlingen, Germany) was used in this experiment. This instrument is equipped with a 1.5 m fiberoptic diffuse reflectance probe and an extended TE-cooled indium gallium arsenide (InGaAs) detector. Bruker OPUS software version 6.0 was used for all data collection and analysis.

Diffuse reflectance spectra were recorded at $8\,\mathrm{cm}^{-1}$ resolution with 64 co-added scans over the spectral range of 4000–$12{,}000\,\mathrm{cm}^{-1}$. Spectralon 1 (Labsphere, New Hampshire, USA), which was permanently mounted in a holder on the instrument, was used as the reference material for all background spectra. All spectra were non-destructively measured from the bottom of each bottle using a 1.5 m fiber optic probe. One sample was selected at random for each batch to record spectra. Five NIR spectra were recorded per sample and the average spectrum was used for the model construction or the analysis.

2.4. *CARS algorithm*

CARS, a novel strategy for variable selection proposed by Prof Liang and his collaborators,[7] has the potential to select an optimal combination of the wavelengths existing in the full spectrum, coupled with partial least squares (PLS) regression. This process of variable selection is somewhat similar to the "survival of the fittest" principle in Darwin's Evolution Theory and can be accomplished directly using the program of MatLab. In an efficient and competitive way, CARS enables the selection of a combination of key wavelengths which are of great significance. The algorithm is described in details as follows:

The data matrix X $(m \times p)$ is a spectral matrix which contains m samples in rows and p variables (wavenumbers) in columns. Vector $\mathbf{y}$ with order $m \times 1$ denotes the measured property of interest. The scores matrix of X is denoted by T, which is a linear combination of X with W as combination coefficient. "c" is the regression coefficient vector of $\mathbf{y}$ against T by least squares. Thus, we have the following formula:

$$T = XW, \tag{1}$$

$$\mathbf{y} = Tc + e = XWc + e = Xb + e, \tag{2}$$

where e is the prediction error and $b = Wc = [b_1, b_2, \ldots, b_p]^T$ is the p-dimensional coefficient vector. The absolute value of the ith element in b, denoted $|b_i| (1 \le i \le p)$ reflects the ith wavelength's contribution to $\mathbf{y}$. Thus, it is natural to say that the larger $|b_i|$ is, the more important the ith variable is. For evaluating the importance of each wavenumber, we define a normalized weight as:

$$w_i = \frac{|b_i|}{\sum_{i=1}^{p} |b_i|}, i = 1, 2, 3, \ldots, p. \tag{3}$$

Additional attention should be paid to ensure that the weights of the eliminated wavenumbers by CARS are set to zero manually so that the weight vector $\mathbf{w}$ is always p-dimensional.

Suppose the number of Monte Carlo (MC) sampling runs of CARS is set to N. In each sampling run, CARS works in four successive steps:

Step 1: A PLS model is built using the randomly selected samples (usually 80% of the calibration set).

Step 2: Using exponentially decreasing function (EDF) to remove the wavenumbers whose $|b_i|$ are relatively small by force. In the ith sampling run, the ratio of wavenumbers to be used is computed using an EDF defines as:

$$r_i = ae^{-ki}, \tag{4}$$

where a and k are two constants determined by the following two conditions: (I) in the first sampling run, all the p wavenumbers are used for modeling which means that $r_1 = 1$, (II) in the Nth sampling run, only two wavenumbers are reserved such that we have $r_N = 2/p$. With the two conditions, a and k can be calculated as:

$$a = \left(\frac{p}{2}\right)^{1/(N-1)}, \tag{5}$$

$$k = \frac{\ln(p/2)}{N-1}. \tag{6}$$

Step 3: Adaptive reweighted sampling (ARS) is employed in CARS to further eliminate wavenumbers in a competitive way.

Step 4: Using the reserved wavenumbers to reconstruct PLS model and calculate the RMSECV.

After N sampling runs, CARS obtains N subsets of variables and corresponding N RMSECV values. Finally, the subset with lowest RMSECV is selected as optimal subset of variables.

2.5. *IR–NIR 2D correlation spectral analysis*

IR–NIR 2D correlation spectral analysis is a new method for selecting characteristic spectral bands of a given chemical compound in the range of NIR. Based on the functional groups identified by IR spectrum and the correlation analysis of IR–NIR spectrum, the typical characteristic NIR spectral bands that were closely related to the structure of the analyte of interest were identified.[17] During the IR–NIR 2D spectral analysis, the spectra of various samples at 30–90°C were collected at equal intervals of 10°C. For IR analysis, the samples were ground with potassium bromide powder and pressed to form slices for spectral collection. For NIR analysis, diffuse reflectance spectra of samples were recorded using the integrating sphere. NIR raw spectra suffered from a marked scatter effect observed as baseline shift and tilt as a result of varying particle size and varying shape of the powdered samples. In order to best preserve the original features of the NIR/IR spectra and increase spectral resolution, second derivative was performed on all IR and NIR spectra before 2D correlation analysis. The IR–NIR 2D spectral analysis was obtained with software (2Dshinge[c] Shigeaki Morita, Kwansei-Gakuin University, 2004–2005, Japan). The 2D correlation graph was a colored graph with positive correlation in red and negative correlation in blue. The darker the red color, the stronger correlation was suggested.

3. Results and Discussion

3.1. *Selection of spectral ranges for universal quantitative models of cephalosporins*

As shown in Fig. 2, cephalosporins have strong absorption in the following five NIR spectral ranges: 4550–4100 cm^{-1}(Range I), 5390–4650 cm^{-1}(Range II), 6190–5510 cm^{-1}(Range III), 7290–6310 cm^{-1} (Range IV), and 9150–8150 cm^{-1}(Range V). Study of the performance of three accomplished NIR universal quantitative models for cephalosporins from past research (shown in Table 2), suggested that the usage of Ranges I, III and V during model construction was in higher frequency than of other two ranges. Therefore, in order to find the best spectral range for cephalosporins, Ranges I, III and V of these band combinations were used to rebuild the above three quantitative models separately using the same calibration and validation set as the references, while other parameters of these models such as the spectral pre-treatment methods were not changed when reconstructing models. The results indicated that when using Range I combined with Range III, the models gave better results for cross validation and external testing than those using other ranges (Table 3). Subsequently, the spectra of degradation samples obtained from the accelerated test were predicted using these reconstructive models in order to evaluate if the choice of Ranges I and III was suitable. It can be seen from Table 4 that Ranges I and III were good choices for quantitative models although in some cases the model with single range could also provide better prediction results, such as the Model 7 for cefuroxime sodium and Model 6 for ceftriaxone sodium. However, considering the models' robustness, spectral range combination was superior to a single range. Compared to the performance of the reported optimization models (original model) shown in Table 2, the root mean square errors of prediction (RMSEP) values of the new models using Ranges I and III were similar to those of original models; hence, the new models using Ranges I and III can fully meet the needs of actual prediction.

Cefuroxime sodium, ceftriaxone sodium and cefoperazone sodium are three kinds of clinically useful classical cephalosporins and their characteristics are clearly described in Chinese Pharmacopoeia: cefuroxime sodium occurs as crystalline powder, and its water content is no more than 3.5%. Ceftriaxone sodium contains crystalline hydrate, and its water content varies from 8.0% to 11.0%. Cefoperazone sodium occurs in crystalline form and amorphous form. The water content of crystalline form varies between 2% to 5% while no more than 2% exists in amorphous form. From these descriptions, we find that there are significant differences in the water contents of these three kinds of cephalosporins; the production process of these drugs is

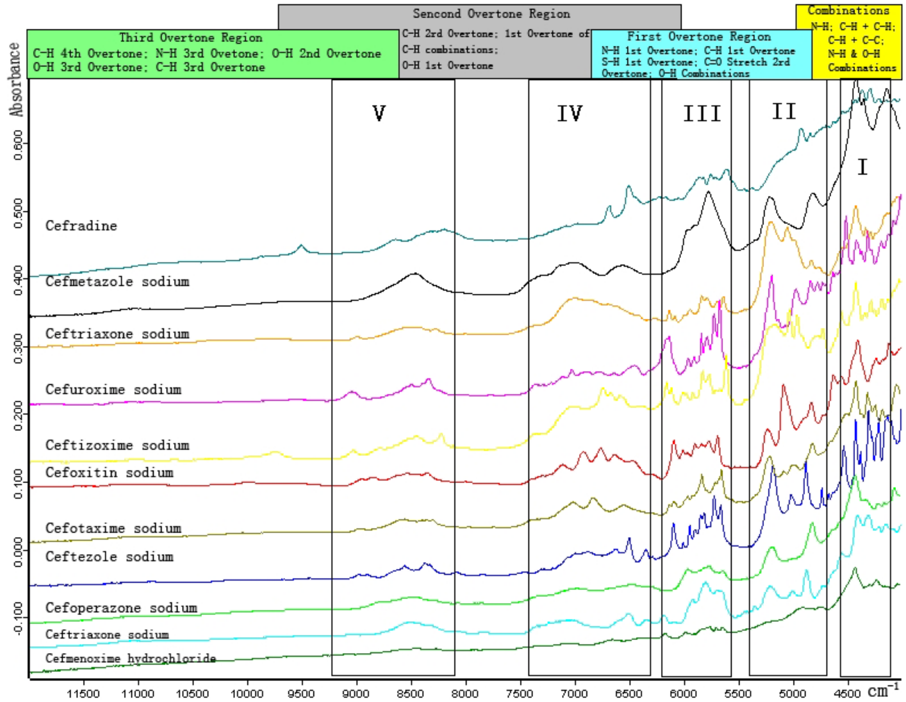


Fig. 2. The NIR spectra of cephalosporins.

Table 2. Characteristics of the initial published calibration models for cephalosporins.

Calibration model	Spectral range (cm^{-1})	Content range (%)	Pretreatment method	RMSECV (%)	Correlation coefficient (r)	Rank	RMSEP (%)
Cefuroxime sodium[14]	11004.6–7494.6 6102.1–4246.8	40.8–96.0	First derivative and vector normalization	1.28	0.9912	8	2.21
Ceftriaxone sodium[15]	6993–5527 4601–4227	72.1–90.6	First derivative and vector normalization	1.21	0.9433	7	1.07
Cefoperazone sodium[16]	4246.7–4424.2 5442.5–6256.3	70.8–89.6	First derivative and vector normalization	1.19	0.9356	10	2.42

either lyophilization (amorphous product) or solvent crystallization (crystalline product). Lyophilization and solvent crystallization are the two major manufacturing processes of cephalosporins. Therefore, these drugs are considered to be representative samples of cephalosporins both for process and structure. The observation that the combination of Ranges I and III is appropriate for the quantitative model construction of cefuroxime sodium, ceftriaxone sodium and cefoperazone sodium indicates that this range may be a globally appropriate range for the quantitative model construction of cephalosporins and could potentially be used as the first choice for other cephalosporins when a calibration model needs to be setup rapidly.

3.2. *Validation for the combination of Ranges I and III*

To further investigate if the Ranges I and III would be a suitable combination for other cephalosporins, samples of cefmenoxime hydrochloride, ceftezole sodium, cefmetazole, cefoxitin sodium, cefotaxime sodium, cefradine, cephazolin sodium and ceftizoxime sodium for injection were collected from different manufacturers in China, and their NIR spectra

Table 3. Characteristics of models obtained with different spectral ranges.

			Cross validation		External test
Calibration model	Model no.	Spectral range (cm^{-1})	R^*	RMSECV	RMSEP
Cefuroxime sodium	1	9207–7494, 6407–5493, 5002–4193	97.97	1.86	1.44
	2	9207–7494, 6407–5493	97.18	2.2	1.44
	3	9207–7494, 5002–4193	97.62	2.02	1.55
	4	6407–5493, 5002–4193	97.94	1.88	1.37
	5	9207–7494	97.1	2.23	1.94
	6	6407–5493	97.08	2.24	1.36
	7	5002–4193	97.93	1.88	1.55
Ceftriaxone sodium	1	10,100–7400, 5400–6500, 4500–4200	91.88	1.45	1.26
	2	10,100–7400, 5400–6500	90.57	1.56	1.21
	3	10,100–7400, 4500–4200	88.6	1.72	1.82
	4	5400–6500, 4500–4200	92.4	1.4	1.17
	5	10,100–7400	75.38	2.53	1.88
	6	5400–6500	92.03	1.44	1.32
	7	4500–4200	91.17	1.51	1.43
Cefoperazone sodium	1	10,200–7700, 6250–5446, 4424–4246	92.42	1.29	2.43
	2	10,200–7700, 6250–5446	91.15	1.39	2.52
	3	10,200–7700, 4424–4246	74.87	2.35	3.49
	4	6256–5443, 4424–4246	93.56	1.19	2.42
	5	10,200–7700	72.96	2.44	3.31
	6	6250–5446	92.76	1.26	2.78
	7	4424–4246	71.65	2.49	3.91

*Determination coefficient ($R = r^2$).

Table 4. Prediction results of different models for accelerated degradation samples.

			Mean difference of prediction $= \frac{\sum_{i=1}^{n} \mathrm{ABS}(\mathrm{NIRvalue} - \mathrm{HPLCvalue})}{n}$						
Powder injection	Batches	Acceleration time (day)	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
Cefuroxime sodium	3	2d, 8d, 17d	1.2	3.7	1.4	1.4	1.4	2.04	1.1
Ceftriaxone sodium	3	2d, 4d, 8d	0.9	0.8	1.3	0.9	1.4	0.7	1.2
Cefoperazone sodium	4	0d, 11d	3.0	3.3	5.8	1.4	4.1	1.5	1.0

were recorded. Subsequently, quantitative models for each drug using the combination of Ranges I and III were constructed. The parameters and validation results for these models are presented in Table 5. Previous analysis of the performance of all the universal quantitative models equipped in mobile labs indicated that if the value of RMSECV was no more than 3.0, the corresponding model could meet our need for rapid screening drugs (data not shown). As we can see in the Table 5, all the RMSECVs for the eight models were less than 3.0, which revealed that the combination of Ranges I and III could be representative of cephalosporins, and using this range would greatly facilitate the process of construction of NIR quantitative models' of cephalosporins.

3.3. *Spectral interpretation of Ranges I and III*

3.3.1. *Correlation analysis between concentration and NIR spectral regions by CARS algorithm*

CARS algorithm uses the absolute values of regression coefficients of PLS model as an index for evaluating the importance of each wavenumber. So it can usually locate an optimal combination of

Table 5. Model characteristics of other cephalosporins using Ranges I and III.

Calibration model for injection	Spectral range (cm^{-1})	Pretreatment method	Content range (%)	RMSECV (%)	R	Rank	Number of spectra in calibration set
Cefmenoxime hydrochloride	6198.3–5507.9 4497.4–4092.4	First derivative and vector normalization	39.0–82.3	2.39	91.06	3	46
Ceftezole sodium	6202.2–5546.5 4852.2–4177.2	First derivative and vector normalization	84.6–93.3	0.968	86.74	5	45
Cefmetazole sodium	7502–5446.2 4246.7–4601.5	Vector normalization	73.6–92.8	1.09	96.47	8	51
Cefoxitin sodium	7502–5446.2 4424.1–4242.8	First derivative and vector normalization	81.6–98.7	2.12	74.98	10	22
Cefotaxime sodium	4293–7201.2	First derivative and vector normalization	54.1–96.2	1.75	97.03	6	89
Cefradine	4223.4–4501.1 5399.8–6502.9	First derivative and vector normalization	13.3–77.2	2.92	97.11	10	53
Cephazolin sodium	6503.8–5396.7 4501.7–4197	First derivative and vector normalization	53.3–95.3	2.82	93.67	11	119
Ceftizoxime sodium	6503.1–5523.4 4948.6–4242.8	First derivative and vector normalization	83.4–92.8	0.58	91.94	11	102

some key wavenumbers which are attributable to the chemical property of interest. In this study, CARS method was used to determine the key wavenumbers with high correlation to the content of two kinds of cephalosporins, namely cefotaxime sodium and cefuroxime sodium. About 28 batches of cefotaxime sodium for injection with the content range from 76.6% to 95.2% and 36 batches of cefuroxime sodium for injection with the content range from 40.8% to 96.0% (some samples were from acceleration experiment) were used for CARS analysis. The key wavenumbers selected by CARS for cefotaxime sodium and cefuroxime sodium are shown in Table 6. As can be seen, most of the selected wavenumbers belong to Range I or Range III, which indicates that Ranges I and III contain some key wavenumbers with high correlation to the content of the analytes.

3.3.2. *Correlation analysis between chemical structure of cephalosporins and their NIR spectral regions using IR–NIR 2D analysis*

In this study, IR–NIR 2D correlation analysis of cephalosporins was used to help determine structural information in NIR spectra. Figure 3 shows the IR spectra of cefmenoxime hydrochloride for injection. The main absorption peaks in the IR spectrum include acylamino (–CONH–, 3265 cm^{-1}), amidocyanogen (–NH, 3204 cm^{-1}), methyl ($–CH_3$, 2996 cm^{-1}) and methylene ($–CH_2–$, 2940 cm^{-1})

Table 6. Results of CARS analysis for cefotaxime sodium and cefuroxime sodium.

Preparation	The selected wavenumbers belonged to range I (cm^{-1})	The selected wavenumbers belonged to range III (cm^{-1})	Other selected wavenumbers (cm^{-1})
Cefotaxime sodium	4351, 4355, 4378, 4605, 4609, 4617	5870, 5874, 5886, 5890	6518, 6522, 6538
Cefuroxime sodium	4532, 4528, 4524, 4517, 4386, 4382, 4366	5655, 5647, 5639, 5635	6658, 6653

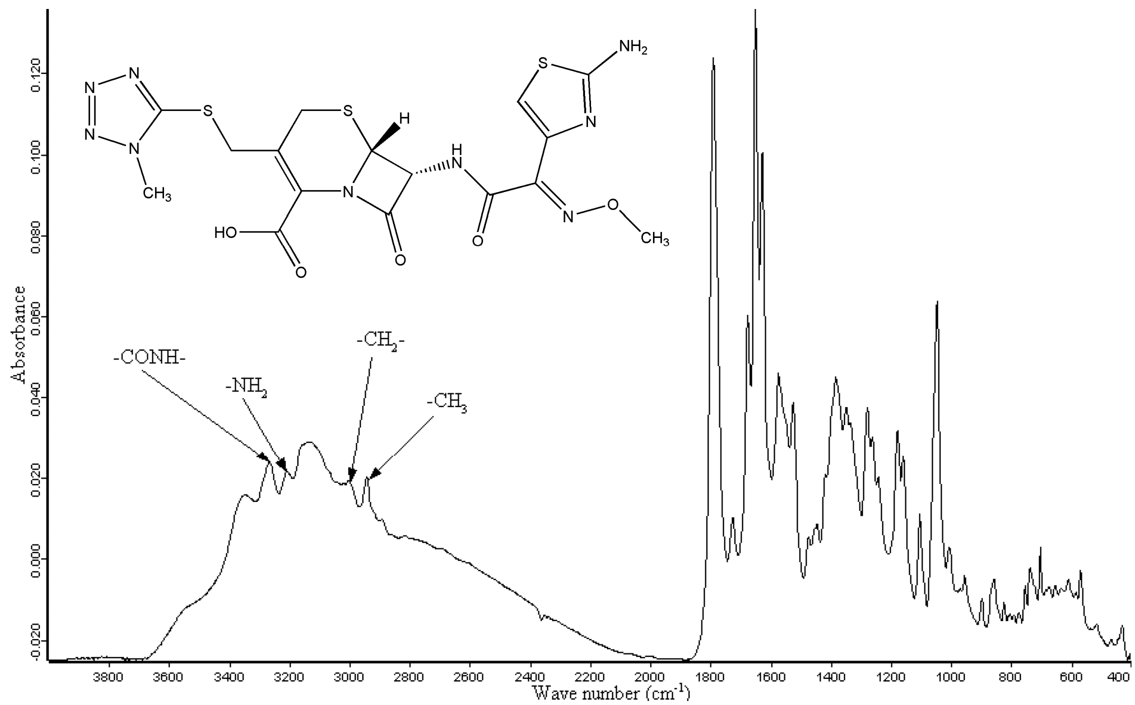
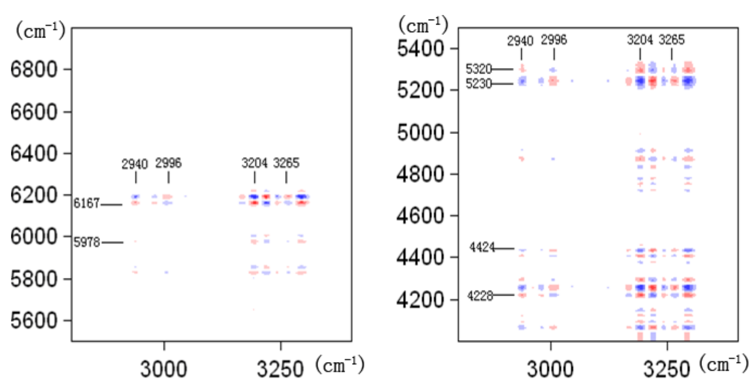


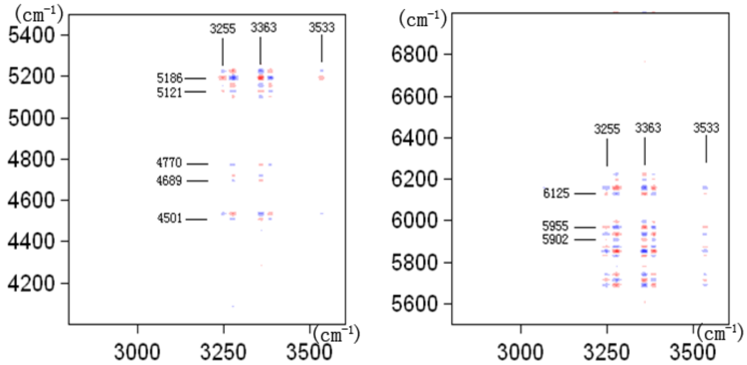
Fig. 3. IR spectrum of cefmenoxime hydrochloride for injection.

groups. Figure 4 shows the synchronized IR–NIR 2D correlation spectra of cefmenoxime hydrochloride for injection as influenced by temperature. The range in Fig. 4(a) was between the IR region of 3400–2800 cm^{-1} and the NIR region of 7000–5500 cm^{-1} and in Fig. 4(b) it was between the IR region of 3400–2800 cm^{-1} and the NIR region of 5500–4000 cm^{-1}. It is obvious that four peaks (at the wavenumbers of 6167, 5978, 5320, 4228 cm^{-1}) in the NIR spectra were all positively correlated with the absorption peaks of methylene at the wavenumber of 2940 cm^{-1} and amidocyanogen at the wavenumber of 3204 cm^{-1}. The correlation analysis shows that variations in the trends of these four peaks in NIR spectrum were identical to those of the absorption peaks at the wavenumber of 2940 cm^{-1} and 3204 cm^{-1} respectively, thus indicating that they were all vibration peaks of methylene and amidocyanogen. Another two peaks in NIR spectrum (5230 cm^{-1}, 4424 cm^{-1}) were positively correlated with the peak of acylamino group at 3265 cm^{-1} in IR spectrum, which revealed that



(a) (b)

Fig. 4. 2D synchronous correlation spectra of cefmenoxime hydrochloride for injection. (a) Spectrum between the IR region of 3400–2800 cm^{-1} and the NIR region of 7000–5500 cm^{-1}. (b) Spectrum between the IR region of 3400–2800 cm^{-1} and the NIR region of 5500–4000 cm^{-1}.



(a) (b)

Fig. 5. 2D synchronous correlation spectra of cefuroxime sodium for injection. (a) Spectrum between the IR region of 3800–2800 cm^{-1} and the NIR region of 4000–5500 cm^{-1}. (b) Spectrum between the IR region of 3800–2800 cm^{-1} and the NIR region of 5500–7000 cm^{-1}.

these two peaks came from the vibration of acylamino group. In these six NIR peaks of cefmenoxime hydrochloride, 4228 cm^{-1} and 4424 cm^{-1} belong to Range I, while other two peaks at 6167 cm^{-1} and 5978 cm^{-1} belong to Range III.

Figure 5 shows the synchronized IR–NIR 2D correlation spectra of cefuroxime sodium for injection as influenced by temperature. Figure 5(a) spectrum was between the IR region of 3800–2800 cm^{-1} and the NIR region of 4000–5500 cm^{-1} and Fig. 5(b) spectrum was between the IR region of 3800–2800 cm^{-1} and the NIR region of 5500–7000 cm^{-1}. As can be seen, the three peaks in the NIR spectrum of cefuroxime sodium (at the wavenumbers of 6125, 5955 and 5186 cm^{-1} were all positively correlated with the peak of the acylamino group at 3533 cm^{-1} and the peak of the amido group at 3363 cm^{-1} and 3255 cm^{-1} respectively; the peaks of 5902, 4770, 4689 and 4501 cm^{-1} in the NIR spectrum were also positively correlated with the peak of the amido group at 3363 cm^{-1} in the IR spectrum; the peak of 5121 cm^{-1} was positively correlated with the peak of the amido group at 3363 cm^{-1} and 3255 cm^{-1}. Among the eight NIR characteristic peaks of cefuroxime sodium, 4501 cm^{-1} belongs to Range I and 5902, 5955 and 6125 cm^{-1} belong to Range III.

The IR–NIR correlation analysis was also performed on other homologous compounds of cephalosporins. Similar results as those of cefmenoxime hydrochloride and cefuroxime sodium were obtained, namely that Ranges I and III contain the wavenumbers which have strong correlations to some functional groups of cephalosporins. According to previous reports, most of these groups were easy to degrade in cephalosporins.[12,13] In other words, Ranges I and III had potential to differentiate cephalosporins and their degradation products either with ring cleavage or hydrolyzed R_3.

4. Conclusions

The empirical universal spectral ranges for developing NIR quantitative calibration models of cephalosporins are recommended in this paper, namely 4550–4100 cm^{-1} (Range I) and 6190–5510 cm^{-1} (Range III). These empirical spectral regions were obtained from detailed analysis of reconstructed NIR calibration models of three kinds of cephalosporins (cefuroxime sodium, ceftriaxone sodium and cefoperazone sodium). Subsequently, the NIR quantitative models of eight other kinds of cephalosporins for injection were built up using Ranges I and III, which demonstrated the universality of this spectral combination to most homologues of cephalosporins. CARS algorithm and IR–NIR 2D correlation spectral analysis provided scientific basis for these two spectral ranges as the universal regions for construction of calibration models of cephalosporins. The above two analytical methods demonstrated that Ranges I and III were not only related to the structural characteristics of cephalosporins, but also to changes in their contents. Therefore, the combination of Ranges I and III can be used as the preferred spectral regions or as the basis of spectral regions optimization for construction of quantitative models of cephalosporins.

Acknowledgment

This work was supported by grant from the National Department Public Benefit Research Foundation (General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China) (Grant No. 2012104008). At the same time, the authors would like to thank Prof Yi zeng Liang (Central South University, PR China) for freely providing us with CARS program.

References

1. N. Boaz, R. C. Ronald, "The prediction error in CLS and PLS: The importance of feature selection prior to multivariate calibration," *J. Chemom.* **19**, 107–118 (2005).
2. J. H. Kalivas, N. Roberts, J. M. Sutter, "Global optimization by simulated annealing with wavelength selection for ultraviolet-visible spectrophotometry," *Anal. Chem.* **61**, 2024–2030 (1989).
3. D. Jouan-Rimbaud, D. L. Massart, R. Leardi, O. E. de Noord, "Genetic algorithms as a tool for wavelength selection in multivariate calibration," *Anal. Chem.* **67**, 4295–4301 (1995).
4. V. Centner, D. L. Massart, O. E. de Noord, S. de Jong, B. M. Vandeginste, C. Sterna, "Elimination of uninformative variables for multivariate calibration," *Anal. Chem.* **68**, 3851–3858 (1996).
5. L. Nøgaard, A. Saudland, J. Wagner, J. P. Nielsen, L. Munck, S. B. Engelsen, "Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near infrared spectroscopy," *Appl. Spectrosc.* **54**, 413–419 (2000).
6. M. C. U. Araújo, T. C. B. Saldanha, R. K. H. Galvão, T. Yoneyama, H. C. Chame, V. Visani, "The successive projections algorithm for variable selection in spectroscopic multicomponent analysis," *Chemom. Intell. Lab. Syst.* **57**, 65–73 (2001).
7. H. Li, Y. Liang, Q. Xu, D. Cao, "Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration," *Anal. Chim. Acta.* **648**, 77–84 (2009).
8. F. Allegrini, A. C. Olivieri, "A new and efficient variable selection algorithm based on ant colony optimization. Applications to near infrared spectroscopy/partial least-squares analysis," *Anal. Chim. Acta.* **699**, 18–25 (2011).
9. X. Zhou, J. Zhao, M. J. W. Povey, M. Holmes, H. Mao, "Variables selection methods in near infrared spectroscopy," *Anal. Chim. Acta.* **667**, 14–32 (2010).
10. Y. Hao, W. Cai, X. Shao, "Construction of the calibration model for near infrared spectral analysis of complex samples," *Chem. J. Chinese U.* **30**, 28–31 (2009).
11. Y. Y. Liu, C. Q. Hu, "Construction of predictive models for gentamicin contents in aqueous solution using specific near infrared spectral regions," *Vib. Spectrosc.* **55**, 241–249 (2011).
12. J. M. Indelicato, T. T. Norvilas, R. R. Pfeiffer, W. J. Wheeler, W. L. Wilham, "Substituent effects upon the base hydrolysis of penicillins and cephalosporins. Competitive intramolecular nucleophilic amina attack in cephalosporins," *J. Med. Chem.* **17**, 523–527 (1974).
13. A. D. Deshpande, K. G. Baheti, N. R. Chatterjee, "Degradation of β-lactam antibiotics," *Curr. Sci. India.* **87**, 1684–1695 (2004).
14. Z. Ni, C. Q. Hu, F. Feng, "Development of a near infrared method for rapid determination of cefuroximes sodium for injection," *Chin. Pharm. J.* **43**, 138–141 (2008).
15. S. R. Hou, Y. C. Feng, C. Q. Hu, "Development of a near infrared method for rapid determination of ceftriaxone and water content in ceftriaxone sodium for injection," *J. Pharm. Anal.* **28**, 936–941 (2008).
16. H. H. Pang, Y. C. Feng, C. Q. Hu, B. R. Xiang, "Construction of universal quantitative models for determination of cefoperazone sodium for injection from different manufacturers using near infrared reflectance spectroscopy," *Spectrosc. Spect. Anal.* **26**, 2214–2218 (2006).
17. T. Liu, Y. C. Feng, D. Q. Song, C. Q. Hu, "Selection of characteristic spectral bands for the analysis by the NIR correlation coefficient method," *J. Chin. Pharmaceu. Sci.* **20**, 83–91 (2011).