J. Pharm. Pharmacol. 2000, 52: 1271–1277
Received April 4, 2000
Accepted June 6, 2000

© 2000 J. Pharm. Pharmacol.

Rapid Identification of *Digitalis purpurea* Using Near-Infrared Reflectance Spectroscopy

MARIKO KUDO, ROBERT A. WATT AND ANTHONY C. MOFFAT

Centre for Pharmaceutical Analysis, The School of Pharmacy, University of London, 29/39 Brunswick Square, London WC1N 1AX, UK

Abstract

Glycosides from *Digitalis* are widely used for the treatment of various cardiac conditions. The potential for near-infrared (NIR) spectroscopy as a technique for the rapid identification of *Digitalis purpurea* was studied. If successful, this method would be advantageous over traditional methods which are destructive and time-consuming.

It was possible to identify *D. purpurea* from other plants using a Maximum Distance in Wavelength Space statistical comparison method on standard normal variate-corrected, second-derivative spectra. Match values ranged from 1·65 to 2·26 for correct identification and were greater than 112·1 for other plants. It was also possible to discriminate between different plant parts of *D. purpurea*, with match values ranging from 1·52 to 2·26 for leaves and greater than 29 for other parts of the same plant. The use of correlation coefficients and the Correlation in Wavelength Space methods proved less conclusive, with resulting values for leaves from different plants being very high, and in all but one case, above 0·9. A two-wavelength, nearest neighbours analysis was carried out for de-trended (baseline corrected), standard normal variate-corrected spectra at 1150 and 2160 nm. This resulted in the successful identification of unknown samples.

NIR spectroscopy has the potential for the rapid identification of *D. purpurea*, and possibly for other natural products of pharmaceutical interest.

The use of near-infrared (NIR) spectroscopy for the characterization of herbal medicines such as *Digitalis* is not widespread. Most interest has thus far been focused on agricultural products of commercial interest (Osborne et al 1993). Natural product characterization is necessary to check for adulteration, confirm purity, and achieve identification (Corti et al 1990). Recently, there has been renewed interest in plants used in traditional medicine as potential sources of novel therapeutic agents. *Digitalis*, for example, has been used to treat cardiac patients since 1785. Indeed two of its components, the glycosides digoxin and digitoxin, are the most popular treatments for rapid atrial fibrillation (Cox & Balick 1994). Current methods for analysing *Digitalis*, which include chromatographic techniques and microscopy (Evans 1989a; British Pharmacopoeia 1999), are time consuming and destructive. Thus, NIR spectroscopy may be a useful method for the identification of *Digitalis*, as it is rapid, non-destructive, and can provide simultaneous information about the chemical composition and physical state, including moisture content and particle-size data (Moffat et al 1997).

Correspondence: Mariko Kudo, Centre for Pharmaceutical Analysis, The School of Pharmacy, University of London, 29/39 Brunswick Square, London WC1N 1AX, UK.
E-Mail: mariko@chemb.ulsop.ac.uk

Corti et al (1990) demonstrated the potential of NIR spectroscopy for the differentiation of ginseng and grape seeds of various ages and geographical origins, as well as adulterated or false specimens. In another study, herbal materials such as *Cassia*, *Ganoderma*, and *Smilacis rhizoma* were accurately discriminated according to geographical origins by NIR spectroscopy using partial least squares regression (Woo et al 1998). More recently, the use of NIR spectroscopy for the analysis of *Umbelliferae* seeds was shown for samples of different ages and varieties (Kudo et al 1998). However, such studies are few in number and much development in the field of natural product characterization by NIR spectroscopy is necessary.

In this study, we used NIR spectroscopy for the identification of *Digitalis purpurea*, the most widely used species of *Digitalis*. It was hoped that

NIR spectroscopy could be used to achieve rapid identification and to confirm purity of this medicinally important natural product.

Materials and Methods

Instrumentation

NIR spectroscopy measurements were made using a FOSS NIRSystems (Silver Spring, MD) 6500 Rapid Content Analyser in diffuse reflectance mode over the wavelength range 1100–2500 nm.

Samples

Twelve dried leaf samples of five species of *Digitalis* (*D. purpurea*, *D. lanata*, *D. mertonensis*, *D. ambigua*, and *D. orientalis*), stem samples of *D. purpurea*, one leaf sample each of some other pharmaceutically important natural products and a larger collection of 130 natural products were obtained from the pharmacognosy archives at The School of Pharmacy, University of London.

Method

All samples were finely powdered using a mortar and pestle and placed in 10-mm glass vials (Waters). Each vial was then placed in the Rapid Content Analyser and scanned. Each recorded spectrum, taking approximately 40 s, was an average of 32 scans. Twelve spectra were obtained for each sample, and to verify reproducibility, the vial was shaken and tapped between each measurement.

Data analysis

Spectral data were imported into FOSS Vision software. To remove the effects of scatter and differences in particle size, and therefore to get reproducible spectra, several spectral pretreatments were used. These included first, second, third and fourth derivatives (Osborne et al 1993), de-trending (Barnes et al 1989), and the use of standard normal variates (SNV) (Barnes et al 1989).

Identification of samples was then attempted using four methods: Maximum Distance in Wavelength Space, Correlation in Wavelength Space, correlation coefficient and a two-wavelength nearest neighbours analysis.

The Maximum Distance in Wavelength Space (FOSS Vision Software manual 1998) statistical comparison method used the wavelength range 1100–2500 nm and a second-derivative gap size of 10 nm. A mean product spectrum and a standard deviation spectrum was calculated for each set of spectra. During analysis, each unknown spectrum was subtracted from the mean spectrum and divided by the standard deviation at each wavelength. This method was applied to the five *Digitalis* species and leaf samples of other pharmaceutically important natural products.

In a second type of mathematical analysis, spectral data were again imported into FOSS Vision software and subjected to various data pretreatments. This time, identification was attempted using the Correlation in Wavelength Space method. This method calculated the dot product between the spectrum of interest and the mean spectrum of each library product using equation 1 (Blanco et al 2000):

$$\rho_{jk} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{p}} (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{\sqrt{\sum_{i=1}^{\mathrm{p}} (x_{ij} - \bar{x}_j)^2} \sqrt{\sum_{i=1}^{\mathrm{p}} (x_{ik} - \bar{x}_k)^2}} \qquad (1)$$

where p represents the number of wavelengths, k and j denote the sample and reference products, respectively, and x_i is the measured value at wavelength i. $\bar{x}_j$ is the average spectrum for the reference product j and $\bar{x}_k$ is the average spectrum for the sample. The sample was qualified as the product when the resulting value was higher than a pre-set threshold (FOSS Vision Software manual 1998). This method was applied to the five *Digitalis* species and a variety of leaf samples of other natural products.

In a further mathematical analysis involving correlations, spectral data from the five *Digitalis* species were imported into Microsoft Excel 97 and the correlation coefficient was calculated between each species and between various other natural leaf products. Identification of an unknown *D. purpurea* sample was also attempted by calculating the correlation coefficient between each of the five *Digitalis* species and the unknown. The unknown was classified as the species which gave rise to the highest correlation. The correlation coefficient, r, is denoted by equation 2 (Graham 1999):

$$r = \frac{S_{xy}}{S_x S_y} \qquad (2)$$

where x represents values for one species, and y represents those for another species. S_{xy} is the covariance, given by equation 3:

$$S_{xy} = \frac{\sum xy}{\mathrm{n}} - \overline{xy} \qquad (3)$$

where n is number of variables.

S_x is the standard deviation of x, given in equation 4:

$$S_x = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} \quad (4)$$

and S_y is the standard deviation of y, given in equation 5:

$$S_y = \sqrt{\frac{\sum y^2}{n} - \bar{y}^2} \quad (5)$$

In a final analysis, the spectra of the five *Digitalis* species were imported into Microsoft Excel 97, and the data corresponding to absorbance at two wavelengths were plotted against each other. Various data pretreatments and wavelengths were tried to find the most successful combinations.

This method was then used to carry out a nearest neighbours analysis (Massart et al 1988). Different samples of *D. purpurea*, *D. lanata*, *D. mertonensis* and *D. orientalis* were treated as unknowns and scanned once. Their absorbance at the two relevant wavelengths were then plotted against each. The Euclidean distances, D, between the data points of the unknowns and every remaining point were calculated using equation 6:

$$D = \sqrt{(U_x - K_x)^2 + (U_y - K_y)^2} \quad (6)$$

where U is the unknown, K is the known, and x and y are their x and y co-ordinates, respectively.

The resulting distances were subsequently ranked, and the 12 smallest values (i.e. nearest data points) were selected. The unknown was then classified as the species which gave rise to the majority of the nearest neighbours.

Results and Discussion

Spectral characteristics

Figure 1 shows typical spectra obtained for the five *Digitalis* species. As leaves may contain a variety of substances including cellulose, lignin, chlorophyll, mucilage, tannin, cutin, volatile oil, calcium oxalate and calcium carbonate as well as active components, for example digoxin (Evans 1989b), it was hoped that the differences observed were due to differences in leaf chemistry between species. Each spectrum was characterized by overlapping overtones and combinations originating from the mid-infrared region. Mathematical pretreatments are usually carried out on sample spectra to reduce or eliminate particle size dependence. A common mathematical technique is to take first or second-derivative spectra. Other commonly used mathematical treatments are SNV and de-trending (Barnes et al 1989). The SNV approach effectively removes the multiplicative interferences of scatter and particle size, and de-trending accounts for the variations in baseline shift, which is generally the case in the reflectance spectra of powdered or densely packed samples, with the use of a second-degree polynomial regression (Osborne et al 1993).

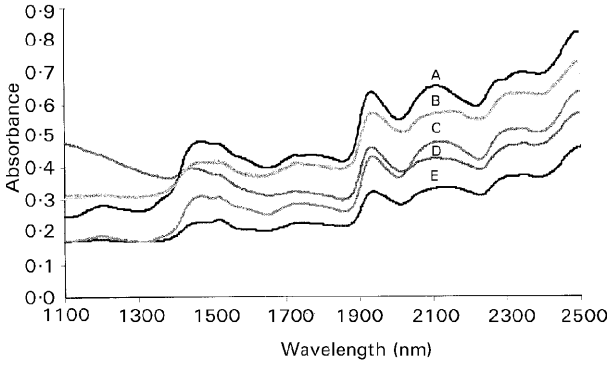

Figure 1. Typical spectra of *Digitalis* species. *D. purpurea* (A), *D. lanata* (B), *D. mertonensis* (C), *D. orientalis* (D) and *D. ambigua* (E).

Maximum Distance in Wavelength Space

For the Maximum Distance in Wavelength Space identification method and the two correlation methods, SNV-corrected, second-derivative spectra were used as they were the most characteristic of the material being analysed. Figure 2 shows SNV-corrected, second-derivative-transformed spectra of the five *Digitalis* species.

Using the maximum distance in wavelength space method on SNV-corrected, second-derivative-transformed spectra in FOSS Vision software

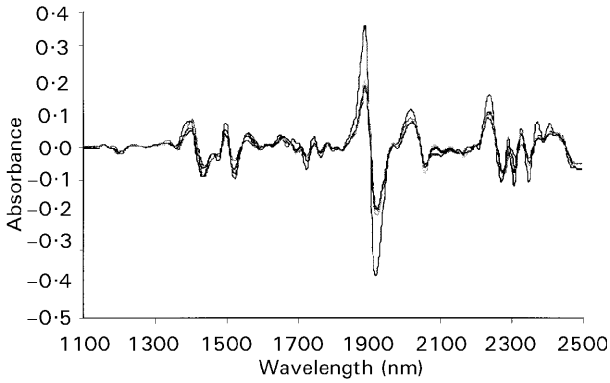

Figure 2. Standard normal variate-corrected second-derivative transformed spectra of five *Digitalis* species (*D. purpurea*, *D. lanata*, *D. mertonensis*, *D. orientalis* and *D. ambigua*).

Table 1. Maximum Distance in Wavelength Space match values between leaves of five *Digitalis* species and other leaf samples for standard normal variate-transformed, second-derivative spectra.

Leaf sample	*D. purpurea*	*D. lanata*	*D. mertonensis*	*D. orientalis*	*D. ambigua*
Belladonna	28·8–34·0–45·4	14·8–15·7–16·2	39·7–41·3–44·3	18·1–19·0–19·9	14·7–15·5–18·9
Hyoscyamus	39·3–46·5–53·7	16·5–19·0–24·9	47·8–52·5–56·6	18·1–21·9–24·2	26·3–30·3–31·8
Stramonium	38·9–41·9–48·1	17·3–17·9–20·5	35·5–46·0–49·2	22·7–26·7–27·2	25·2–27·5–29·6
Buchu	48·9–56·5–65·6	42·7–48·0–53·7	25·7–27·1–28·8	23·4–25·9–28·4	50·9–59·5–69·8
Eyebright	22·1–27·0–34·4	24·6–26·6–27·8	28·8–31·2–34·6	30·7–32·3–38·9	28·8–31·2–34·6
Senna	25·4–32·7–36·2	24·0–27·3–28·6	29·5–33·6–35·3	16·4–17·7–20·1	16·4–17·6–19·0
Peppermint	37·1–39·2–41·0	15·6–16·6–17·4	50·5–52·5–54·4	19·0–21·2–22·3	19·3–19·7–20·2
Hop	23·1–24·3–35·1	11·2–12·1–16·4	29·4–35·7–39·7	16·2–16·8–17·7	11·4–12·2–13·9
Coltsfoot	48·4–55·8–62·9	31·7–35·2–38·5	42·7–47·0–53·2	37·2–40·0–42·6	30·8–35·3–40·8

Results are presented as smallest value-median value (underlined)-largest value.

Table 2. Maximum Distance in Wavelength Space match values between leaf samples of five *Digitalis* species for standard normal variate-corrected, second-derivative spectra.

Leaf sample	*D. purpurea*	*D. lanata*	*D. mertonensis*	*D. orientalis*	*D. ambigua*
D. purpurea	1·65–1·90–2·26				
D. lanata	7·41–9·04–11·0	1·57–1·96–2·20			
D. mertonensis	27·9–31·4–34·2	12·4–14·4–16·1	1·52–1·91–2·25		
D. orientalis	14·8–15·2–16·6	14·2–16·7–18·3	16·7–19·7–21·2	1·32–1·91–2·31	
D. ambigua	10·0–11·3–12·7	10·6–19·0–21·7	14·1–18·5–21·0	60·7–72·4–78·4	1·62–2·03–2·37

Results are presented as smallest value-median value (underlined)-largest value.

it was possible to completely differentiate *Digitalis* from the other leaf samples. Table 1 shows resulting match values between the five *Digitalis* species and leaf samples of other natural products of pharmaceutical interest. Numbers ranged from 11·2 to 69·8, well above the value 3 which is a commonly used threshold value for identification (Blanco et al 2000). Often, a more practical approach is for users to decide upon the most suitable threshold depending on their specific problems and methods (Blanco et al 2000). Large match values were also observed when comparing *D. purpurea* to the other plant materials in the larger collection of 130 natural products. The five lowest values were 14·7, 16·6, 18·5, 19·2 and 21·3 for lemon verbena, Thai, Turkish, and South African cannabis leaves, belladonna flowers and camomile flowers, respectively.

When samples of *Digitalis* were matched against the correct species and the other four species, they could be easily identified and differentiated (Table 2). The correct identities were observed for values of the maximum distance between 1·32 and 2·37, whilst the maximum distance observed between the species was greater than 7·41. There was therefore a substantial gap of 5·03 between the highest correctly matched value and the lowest mismatched value. A reasonable threshold value for identification of *Digitalis* would therefore be 4·0. Based on this value, the method was able to discriminate between the five *Digitalis* species and between *Digitalis* species and other genuses. Among the five *Digitalis* species, *D. purpurea* appears to be most similar to *D. lanata*, then to *D. ambigua*, then to *D. orientalis*, and finally to *D. mertonensis*.

When the method was applied to leaves and stems of *D. purpurea*, the values between the same plant parts ranged from 1·52 to 2·26, and between different plant parts from 29·04 to 67·41. The Maximum Distance in Wavelength Space method was therefore able to distinguish between different plant parts of the same species. Figure 3 shows the

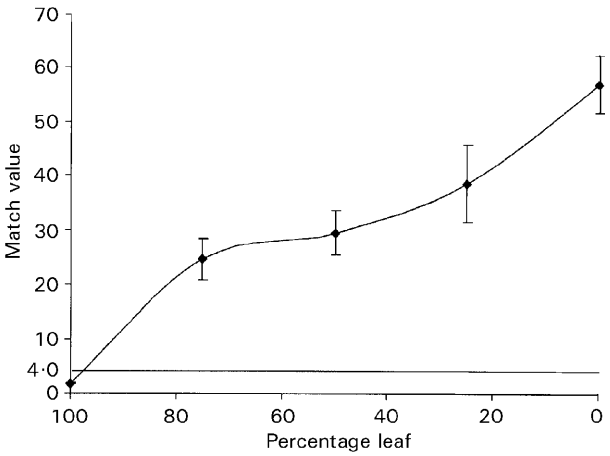

Figure 3. Maximum distance in wavelength space match values obtained from attempted identification of successive dilutions of *D. purpurea* leaf. Error bars indicate ± s.d.

sensitivity of the technique for the discrimination of plant parts. It is clear that a sample with even small amounts of stem (2%) could be detected and identified as distinct from a leaf-only sample using a threshold value 4·0. Thus, adulteration of leaf material by just a small percentage of stem (or any other material) could be detected.

Correlation in Wavelength Space

When the Correlation in Wavelength Space method was applied between the five *Digitalis* species and other leaf products, relatively high values were observed, the lowest being 0·890 for *D. mertonensis* against Stramonium, and the highest being 0·991 for *D. ambigua* against hop leaves. Values between the five *Digitalis* species ranged from virtually complete correlation (0·999) to 0·953. When *D. purpurea* was compared with the larger database of 130 natural products, some low correlation values were observed, the lowest five values being 0·651, 0·656, 0·670, 0·685, and 0·732. However, as the materials that produced the values were all non-leaf in origin (3 types of poppy seeds, rhubarb root, and cloves), dissimilarities would be anticipated. As resulting values were generally so high, setting a threshold limit appears to be problematic, and this method was not successful in discriminating between the samples used here, although it is widely used as a sensitive identification and qualification method for raw materials (Blanco et al 2000).

Correlation coefficients

The *r* values between *Digitalis* and other leaf samples, were rather high, ranging from 0·891 to 0·990. Results between the five species were also quite variable, values being for the most part above 0·96, except between *D. mertonensis* and *D. orientalis* (0·936). For different plant parts, the *r* value between *D. purpurea* leaves and stems was 0·903. Using the method on *D. purpurea* and the larger database of 130 plant materials produced high results overall, although some low values were observed, with the five lowest being 0·654, 0·696, 0·722, 0·738, and 0·771 for 2 types of poppy seed, rhubarb root, cloves, and ergot, respectively. However, as these samples were of non-leaf origin, low values may be expected. Again, similar to the Correlation in Wavelength Space method, because the resulting values were generally so high, it appears that the use of *r* values was not successful for the discrimination of leaf samples.

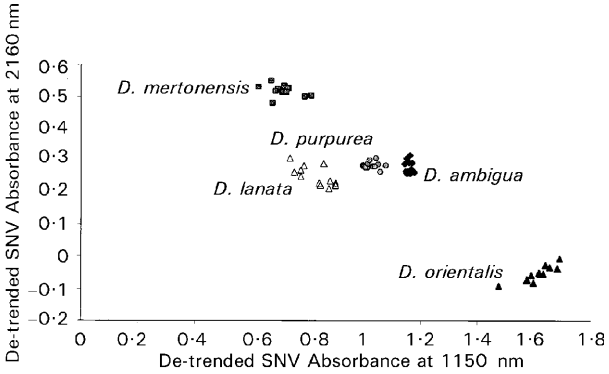

Figure 4. A two-wavelength plot of five *Digitalis* species (*D. purpurea*, *D. lanata*, *D. mertonensis*, *D. orientalis* and *D. ambigua*).

Two-wavelength plot

The NIR absorption spectrum of water includes 5 bands with maxima at 1940, 1450, 1190, 970, and 760 nm at room temperature (Osborne et al 1993). These peaks could be used to quantify water content of materials and thus were avoided as various wavelength combinations and different data pretreatments were tested on the five *Digitalis* species. The final combination used was the two wavelengths 1150 and 2160 nm, on baseline-corrected (de-trended), SNV-transformed spectra. These wavelengths were chosen empirically as the resulting plot (Figure 4) showed the best visual separation of the various *Digitalis* species compared with the other combinations tried.

Identification using nearest neighbours

When the nearest neighbours method of analysis was carried out on the resulting plot, it was clear that each unknown species could be correctly identified. That is, it was possible to correctly assign all the species which were treated as unknowns to the groups that gave rise to the smallest Euclidean distances (i.e. nearest neighbours). It is clear from Table 3 that the closest neighbours for each unknown group came from the species to which it theoretically belonged.

Method comparison

Table 4 shows a summary of the results for identifying *D. purpurea* using NIR spectroscopy and four different methods. Based on these results, it appears that Maximum Distance in Wavelength Space is the most successful method for the identification of *D. purpurea*, with very distinctly low

Table 3. Average Euclidean distances between five *Digitalis* species and four unknown samples between points on a two-wavelength plot for de-trended standard normal variate spectra, wavelengths 1150–2160 nm.

	D. purpurea	*D. lanata*	*D. mertonensis*	*D. orientalis*	*D. ambigua*
D. purpurea (unknown)	0·071	0·182	0·416	0·667	0·163
D. lanata (unknown)	0·213	0·119	0·203	0·884	0·336
D. mertonensis (unknown)	0·443	0·265	0·184	1·104	0·567
D. orientalis (unknown)	0·686	0·853	1·077	0·052	0·580

The smallest value for each row is underlined.

Table 4. Various values between *D. purpurea* leaves and other leaf samples including other *Digitalis* species.

Species	Match value	Correlation	*r*	Euclidean Distance
D. purpurea	1·90	1·000	1·000	0·024
D. lanata	9·04	0·994	0·996	0·205
D. mertonensis	31·4	0·964	0·965	0·392
D. orientalis	15·2	0·968	0·969	0·673
D. ambigua	11·3	0·987	0·987	0·127
Belladonna	34·0	0·982	0·983	0·524
Hyoscyamus	46·5	0·978	0·979	0·177
Stramonium	41·9	0·964	0·964	0·277
Buchu	56·5	0·936	0·937	0·352
Eyebright	27·0	0·964	0·965	0·398
Senna	32·7	0·984	0·984	0·149
Peppermint	39·2	0·990	0·990	0·045
Hop	24·3	0·987	0·987	0·222
Coltsfoot	55·8	0·951	0·952	0·323

Median match values are for standard normal variate (SNV) corrected second-derivative spectra using Maximum Distance in Wavelength Space in FOSS Vision software; correlation is the average correlation values using Correlation in Wavelength Space in FOSS Vision software on SNV-corrected second-derivative spectra; the average correlation coefficient, *r*, for SNV-corrected second-derivative spectra; and the average Euclidean distances are between points on a two-wavelength plot for de-trended SNV spectra, wavelengths 1150 and 2160 nm.

or high match values being observed depending on matched or mismatched samples. It was also sensitive enough to successfully distinguish between different plant parts. As resulting values were rather high, the use of correlation values did not appear to be very successful for discrimination of the samples and more work is needed to make these methods more effective and robust. The use of two-wavelength plots is useful in pattern recognition and gives a good visual idea of the differences between the species. However, this method is still in its preliminary stages, and it is still necessary to devise a method to select the most ideal wavelength combinations and data pretreatments to use, as there may exist other combinations that are potentially more successful than the ones used here.

Conclusion

NIR spectroscopy has the potential for rapid characterization and identification of *D. purpurea*, and may be useful in identifying other plant leaves. NIR spectroscopy is faster than commonly used traditional techniques. While current methods may take hours, analysis by NIR spectroscopy can be performed in a matter of minutes. This study is still in the preliminary stages, and the development of new chemometric techniques is desirable, as well as the improvement of the techniques used here.

Acknowledgements

We thank FOSS NIRSystems for the loan of a spectrophotometer, and Sheelagh Halsey for advice with Vision software. We also thank the Centre for Pharmacognosy, The School of Pharmacy, University of London for supplying plant samples.

References

Barnes, R. J., Dhanoa, M. S., Lister, S. J. (1989) Standard normal variate transformation and de-trending of near-infrared reflectance spectra. Appl. Spectroscopy 43: 772–777

Blanco, M., Eustaquio, A., Gonzalez, J. M., Serrano, D. (2000) Identification and quantitation assays for intact tablets of two related pharmaceutical preparations by reflectance near-infrared spectroscopy: validation of the procedure. J. Pharm. Biomed. Anal. 22: 139–148

British Pharmacopoeia (1999) Vol 1. H. M Stationery Office, London, pp 505–506

Corti, P., Dreassi, E., Franchi, G. G., Corbini, G., Moggi, A., Gravina, S. (1990) Application of NIR reflectance spectroscopy for the identification of drugs derived from plants. Int. J. Crude Drug. Res. 28: 185–192

Cox, P. A., Balick, M. J. (1994) The ethnobotanical approach to drug discovery. Scientific American 270: 82–87

Evans, W. C. (1989a) Trease and Evans' Pharmacognosy. 13th edn, Ballière Tindall, London, pp 504–512

Evans, W. C. (1989b) Trease and Evans' Pharmacognosy. 13th edn, Ballière Tindall, London, pp 23–24

FOSS Vision Software manual (1998) FOSS NIRSystems, Silver Spring, MD

Graham, A. (1999) Teach Yourself Statistics. 2nd edn, Hodder & Stoughton, London, pp 191–197

Kudo, M., Moffat, A. C., Watt, R. A. (1998) The rapid characterization of natural products and herbal medicines by near-infrared spectroscopy. J. Pharm. Pharmacol. 50 (Suppl.): 258

Massart, D. L., Vandeginste, B. G. M., Deming, S. N., Michotte, Y., Kaufman, L. (1988) Chemometrics: a Textbook. Elsevier Science Publishers B.V., Amsterdam, pp 395–397

Moffat, A. C., Jee, R. D., Watt, R. A. (1997) New near infrared centre of excellence for Europe. Eur. Pharm. Rev. 2: 37–40

Osborne, B. G., Fearn, T., Hindle, P. H. (1993) Practical NIR Spectroscopy with Applications in Food and Beverage Analysis. 2nd edn, Longman, Harlow

Woo, Y. A., Cho, C. H., Kim, H. J., Cho, J. H., Cho, K. K., Chung, S. S., Kim, S. J., Kim, J. H. (1998) Discrimination of herbal medicines according to geographical origin (Korea, China) using near-infrared reflectance spectroscopy. Yakhak Hoeji. 42: 359–363