ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
16 (1998) 813–824

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

Active drug substance impurity profiling Part I. LC/UV diode array spectral matching

Edgar C. Nicolas *, Thomas H. Scholz

The DuPont Merck Pharmaceutical Company, Deepwater, NJ 08023, USA

Received 10 December 1996; accepted 17 April 1997

Abstract

Monitoring of drug substance impurities is routinely accomplished using HPLC. However, HPLC retention times can vary, resulting in uncertainty as to whether a peak at a new retention time is a new impurity. Because standards of the minor impurities (less than 0.1% by area) are not usually available, some method is needed to characterize each of these peaks without isolating them. This on-line characterization might be accomplished using UV diode array spectral matching. This work sought to assess the sensitivity and selectivity of UV spectral matching for monitoring the impurity profile of drugs, using as an illustration DuP 941, an anti-cancer drug under development. An ultraviolet spectral data library was generated for a number of the DuP 941 impurities in the earliest safety lot. Impurities in several subsequent lots of DuP 941 were then examined to see how well their spectral characteristics matched those of the spectra contained in the library. We found LC/UV spectral matching to be a powerful method to monitor Dup 941 impurities even down to levels well below 0.1% by area. Critical factors that were shown to influence the utility of the technique include detector sensitivity, lamp intensity, and the presence of other impurities with very similar UV spectra.
© 1998 Elsevier Science B.V.

Keywords: Impurity profiling; HPLC; Diode array spectra; Spectral matching

1. Introduction

The safety of a drug product is dependent not only on the toxicological properties of the active drug substance itself, but on the impurities that it contains. These impurities are often synthetic process impurities, hence changes in the drug's synthetic process could result in a change in the impurity profile and in questions about safety. Process changes are common as a drug candidate is carried through development and the process is scaled up and optimized. Also, when generic versions of a currently marketed drug are produced, the synthetic processes are likely to be different from the one used by the original manufacturer.

Analytical monitoring of impurities in new drug substances is a key component of the recent guideline issued by the International Conference on Harmonization (ICH) [1]. When the drug substance impurities which were present in early safety and clinical lots are not present in substantially higher amounts in subsequent lots, and no

* Corresponding author.

0731-7085/98/$19.00 © 1998 Elsevier Science B.V. All rights reserved.
PII S0731-7085(97)00131-3

new impurities are seen, then clinical studies may proceed on schedule. Even when a higher level of a previously-seen impurity is present, new safety studies could still be avoided if it can be shown by scientific rationale (daily dose, clinical experience) that a higher threshold is justifiable. But when a new impurity is observed in a lot of drug substance, and it cannot be removed, then additional safety studies may be required. This could delay the development timeline. It is therefore vital, when monitoring the impurity profile of a drug substance, to be able to determine whether impurities in one lot are the same impurities that were present in earlier lots.

Gradient HPLC with UV detection is the technique commonly used to monitor the impurities in a drug substance, although many other techniques are also used [2]. Usually, few of the impurity peaks are identified, and even fewer have available reference standards. Impurities are commonly coded and tracked using a descriptor such as retention time. This can lead to uncertainty, especially if peaks overlap, if they shift in retention time (as can happen in gradient HPLC), or if there are numerous closely spaced peaks.

The development and commercialization of diode array detectors more than 10 years ago provided the analyst with a means to assign a tentative identity to chromatographic peaks [3–8]. The approach has been to generate a UV absorbance spectrum for the unknown peak and to 'match' this, using some algorithm, to a spectrum in a spectral library. The technique has been used to identify drugs in mixtures and usually involves the use of reference standards for generation of the spectral library [4–6]. Several investigators have applied the technique to the identification of impurities in a drug [8,9]. The major impurities have usually been the focus in these investigations, and the many minor impurity peaks, usually at less than 0.1 UV area percent, have not been characterized.

In this paper we report our investigation into using UV diode array spectral matching to track not only known major drug substance impurities, but also minor impurity peaks which had never been isolated and identified. We wanted to determine if LC/UV diode array spectral matching was sensitive and reliable enough to effectively monitor and differentiate spectrally similar impurity peaks present in different drug lots at levels less than 0.1% by area, for which no reference standards exist.

We evaluated the approach using DuP 941 (losoxanthrone), an anticancer drug under development at DuPont Merck. Over time, improvements have been made in its synthetic process, and the impurity profile of the various drug substance lots used for animal safety, clinical safety and efficacy studies has changed. The structures of DuP 941 and three known impurities are shown in Fig. 1.

2. Experimental

2.1. Equipment

Experiments were performed on two different Hewlett Packard liquid chromatographs, an HP1090 and an HP1100. The diode array detector of the HP1090 uses 205 diodes and has a maximum diode resolution of 2 nm. The HP1090 we used was equipped with an 8-nm slit assembly and a micro-flow cell with a pathlength of 6 mm. The HP1100 uses 1024 diodes and has a maximum diode resolution of 0.9 nm. It has a programmable slit width which was set to 4 nm and the pathlength of its flowcell was 10 mm. UV spectral matching was done using Hewlett Packard software revision A.01.00.

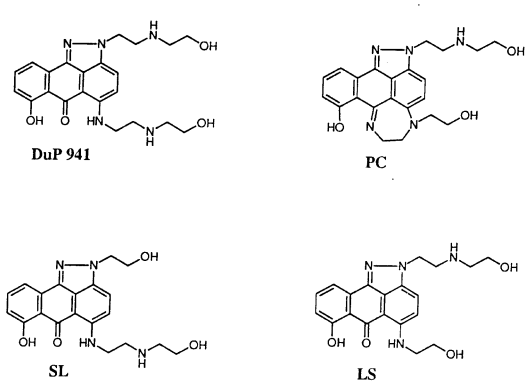

Fig. 1. Structures of DuP 941 and three known impurities.

2.2. Method

Solutions of the different drug substance lots were prepared at a concentration of 5 mg/ml using the initial mobile phase of the gradient. The injection volume was 10 μl, resulting in 50 μg being injected on column. The column used was a 4.6 mm × 15 cm Zorbax SB-C8 column with 3.5-μm particles. The initial mobile phase was acetonitrile–water–trifluoroacetic acid (10:90:0.1, v/v/v). A 20-min linear gradient was used, with a final mobile phase composition of acetonitrile–water–trifluoroacetic acid (40:60:0.1, v/v/v). The flow rate was 1.5 ml/min.

A wavelength of 240 nm was used for the chromatographic signal with diode array spectral data collected over a range of 210–600 nm. Each UV spectrum that was used for the UV spectral matching analysis was a peak apex spectrum from which a baseline reference spectrum was subtracted. The baseline reference spectrum was generated by the software by linear interpolation between two baseline spectra chosen by the software from each side of the peak. The baseline for each of the peaks was drawn manually.

The methods used in this work were similar but not identical to the methods used in our QC laboratories. We observed that when the samples are stored as solutions under conditions that are not strictly controlled, impurity B seems to decompose to impurity A (see Fig. 2). Hence the relative amounts of impurity A and B shown in this study may not be representative of these lots. This conversion however does not effect the validity of the results and conclusions of this work.

3. Results and discussion

In this study, we characterized the impurity profiles of four early lots of DuP 941 drug substance. Lots 3 and 5 had been prepared using one synthetic route, and lots 7 and 8 using a second route. Fig. 2 shows LC/UV gradient chromatograms for these lots generated using the HP1100 liquid chromatograph. Note the numerous impurity peaks. The shift in retention times between lot 3 and the subsequent lots was due to incomplete column equilibration, and was deliberately left uncorrected in this study.

In order to compare peaks from lot to lot, each peak being examined was given a designation consisting of the lot number and a letter. Letter designations were first assigned to certain impurities in lot 3. For example, the lot 3 peak at 5.8 min is designated '3A'. If an impurity in a subsequent lot was thought to be the same impurity as in lot 3, then the letter used for its designation in lot 3 was also used in the subsequent lot (for example, 5A, 7A and 8A are thought to be the same impurity as 3A). Any peaks thought to be new impurities in subsequent lots were given new sequential letter designations (for example, 7H).

The structures of the three isolated and identified impurities, which have been given the names 'PC', 'SL', and 'LS' are shown in Fig. 1. Reference standards of these impurities were available for this study. In lot 3 the PC and LS impurities are present and were labeled as the 'B' peak and the 'C' peak, respectively. Impurity SL was a new impurity first seen in lot 7, and it was labelled with the letter H (see Fig. 2). Because reference standards were not available for most of the impurities, the amounts of the impurities present in each lot are reported only in terms of peak area percent relative to the DuP 941 peak. The impurity peaks ranged in size from 0.23% by area down to 0.001% by area. If one assumes that the extinction coefficient for each of the impurities is the same as that for DuP 941, and that the detector response is linear, then the impurities being examined ranged from 500 pg to 115 ng injected on column.

3.1. The LC/UV diode array spectral library

UV spectra were acquired and processed as described in Section 2, and stored in a UV spectral library. Two libraries were built: one with spectra acquired on the HP1100 and another with spectra acquired on the HP1090. Library spectra for DuP 941, PC, SL, and LS, for which there were standards, were obtained during an HPLC separation of a mixture of these standards (see Fig. 3). Library spectra of the impurities in lot 3 (the earliest safety lot) and the new impurity first

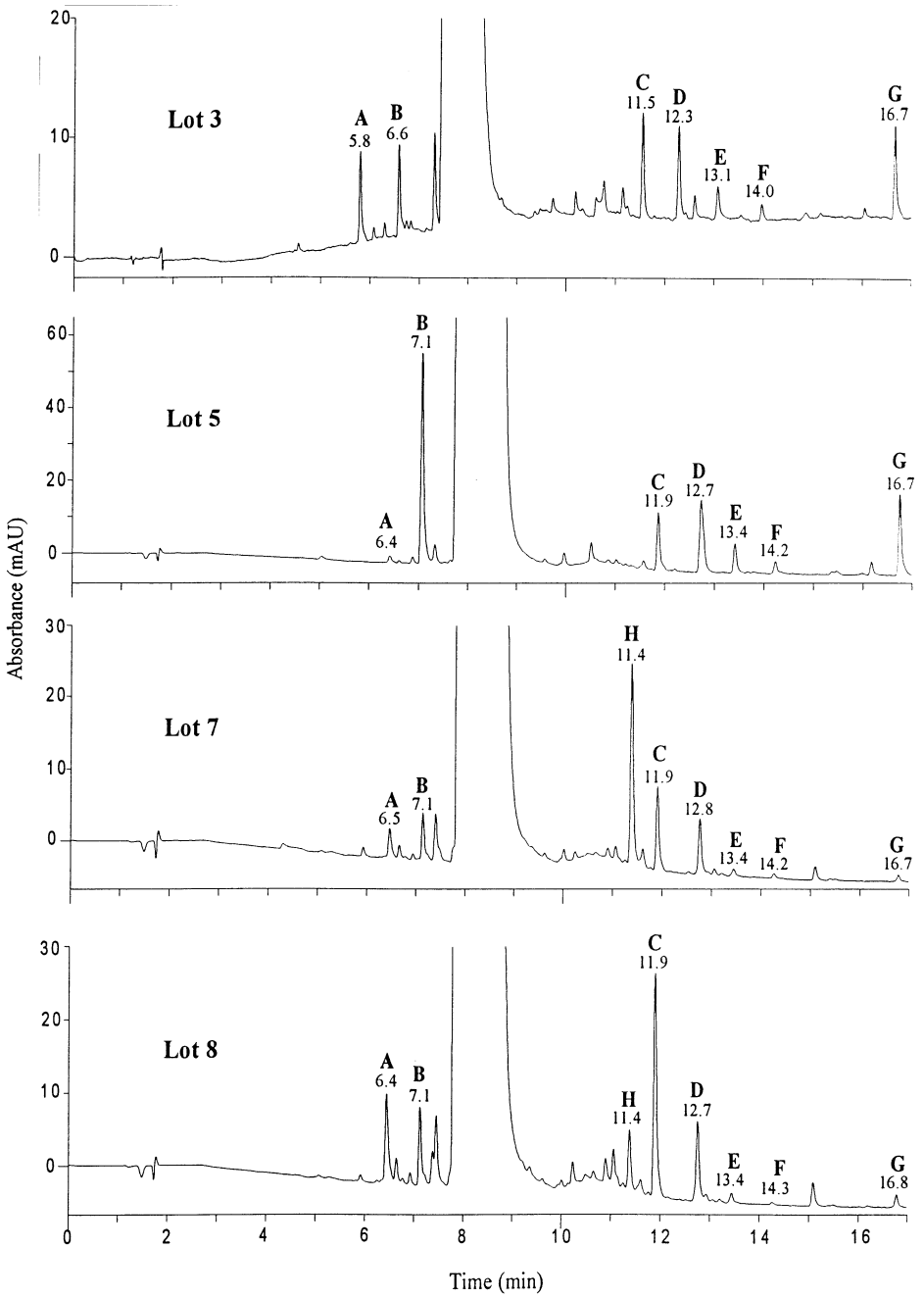

Fig. 2. LC/UV chromatograms of DuP 941 lots acquired on an HP1100.

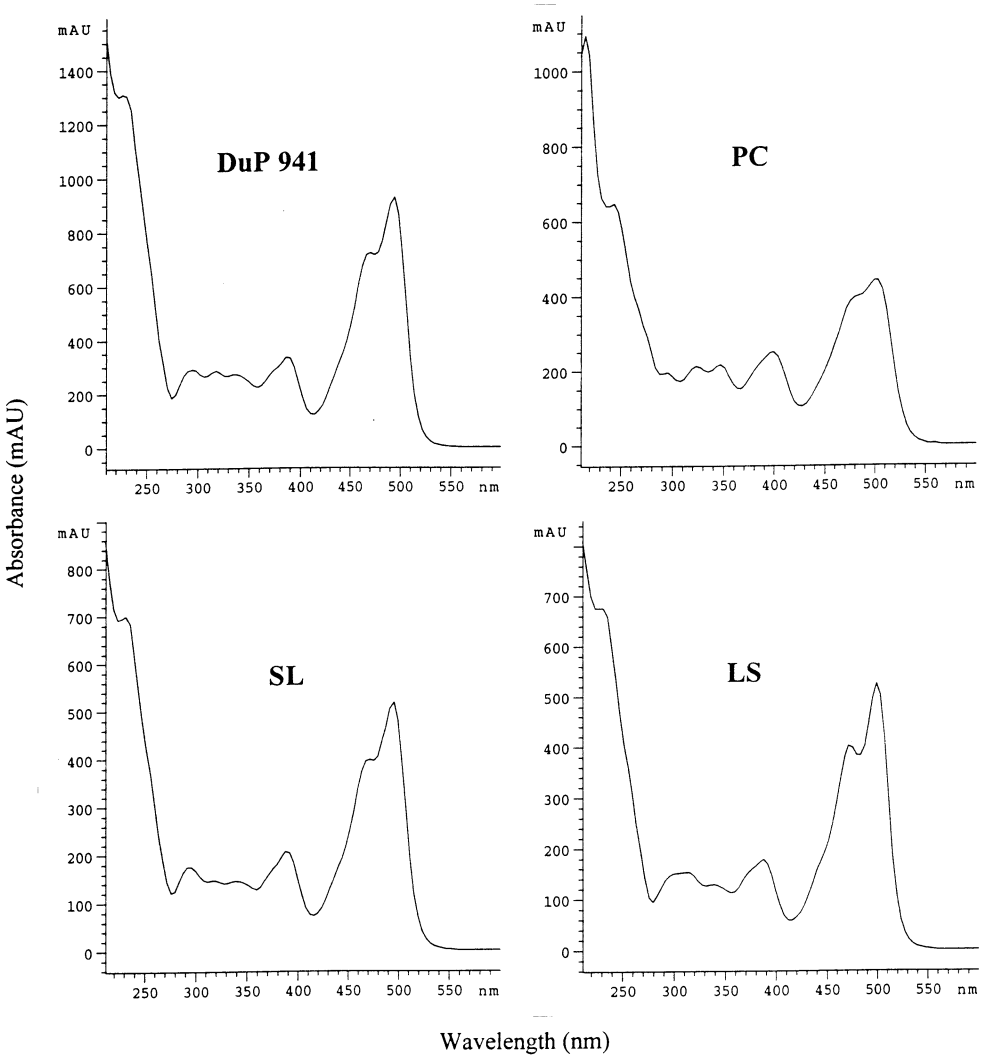

Fig. 3. LC/UV spectra of standards of DuP 941 and known impurities PC, SL and LS.

observed in lot 7 (impurity 7H) were acquired from chromatographic runs of the lot 3 and lot 7 samples (see Fig. 4). As is common with synthetic process impurities, all of the spectra have a general appearance that is very similar to the spectrum of DuP 941.

3.2. Lot-to-lot monitoring of impurities using UV spectral matching

Spectra of each of the chosen impurity peaks in the chromatograms of lots 5, 7, and 8 (see Fig. 2) were generated and compared to all the spectra in the library using Hewlett Packard's spectral match factor analysis software. This software plots the absorbances of two spectra against each other for each wavelength. A linear regression is performed for these data and the square of the regression coefficient is determined. This value is multiplied by 1000 to give the match factor. According to Hewlett Packard literature, a match factor greater than 990 is considered a 'very probable' match. Match factors from 950 to 990 indicate 'possible' matches and numbers below 950 indicate that there is no match.

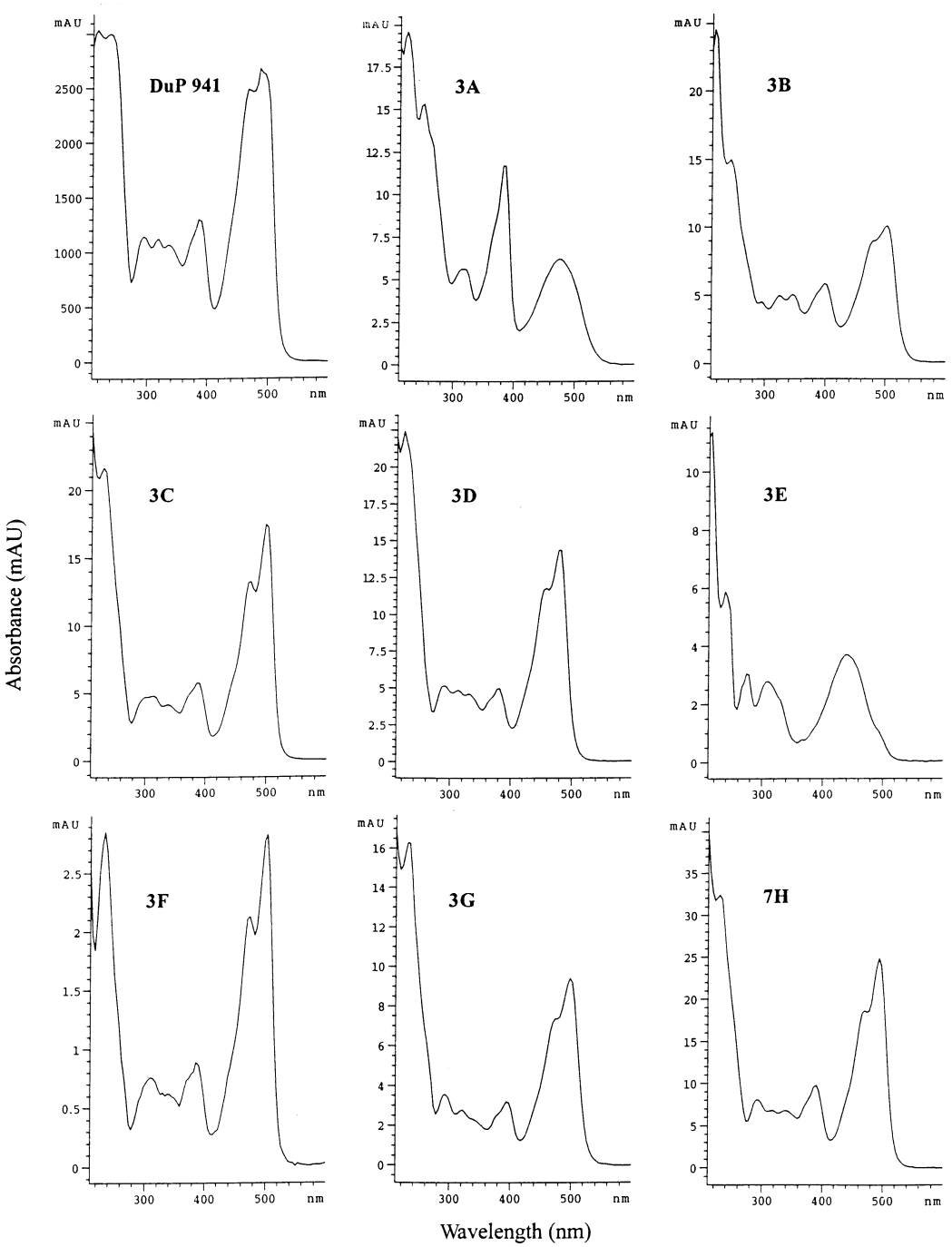

Fig. 4. HP1100 UV spectral library of DuP 941 and impurities.

For every impurity peak examined, based on its retention time and place in the elution order, there was a spectrum in the library that was expected to match the impurity's spectrum. Table 1 lists the match factors (M.F.) to the expected matching spectrum in the UV spectral library for impurities in lots 5, 7, and 8. Results for both the HP1100 and the HP1090 are given. Table 1 also lists the match factors for three of the library spectra (3B, 3C, and 7H) with the spectrum of the reference standard (PC, LS, or SL) that each was expected to match. The desired result is that the match factor to the expected impurity spectrum in the spectral library would be greater than 990, while obtaining match factors of less than 950 to spectra of other impurities in the library.

Table 1
UV match factor analysis of DuP 941 lots using the HP1100 and the HP1090

Peak[a]	Area (%)[b]	Expected match[c]	Match factor[d]	
			HP1100[e]	HP1090[f]
5A	0.01	3A	997	914
7A	0.02	3A	996	992
8A	0.06	3A	995	997
3B	0.08	PC	999	999
5B	0.23	3B	997	999
7B	0.03	3B	995	996
8B	0.04	3B	997	995
3C	0.12	LS	998	999
5C	0.07	3C	1000	994
7C	0.05	3C	1000	994
8C	0.12	3C	1000	998
5D	0.11	3D	1000	999
7D	0.04	3D	997	996
8D	0.04	3D	994	991
5E	0.04	3E	1000	994
7E	0.006	3E	997	g
8E	0.007	3E	998	g
5F	0.02	3F	999	938
7F	0.004	3F	995	g
8F	0.001	3F	977	g
5G	0.10	3G	1000	998
7G	0.005	3G	996	g
8G	0.008	3G	1000	g
7H	0.12	SL	998	999
8H	0.004	7H	1000	993

[a] Peak designation codes: first character is lot number; second character is presumed identity. Presumed identity of 7F and 8F is probably incorrect. See Table 2.
[b] These levels are from HP1100 chromatograms.
[c] The HP1100 area percent levels for the lot 3 peaks were as follows: 3A, 0.09%; 3B, 0.08%; 3C, 0.12%; 3D, 0.12%;; 3E, 0.05%; 3F, 0.02%; 3G, 0.09%.
[d] A match factor of greater than 990 indicates a very probable match. Match factors to spectra other than the expected match were 950 or less (no match), except as noted in Table 2.
[e] The HP1100 had a 10-mm pathlength flowcell and a 4-nm slit width.
[f] The HP1090 had a 6-mm pathlength flowcell and a 8-nm slit width.
[g] Below limit of detection (LOD = 0.01%).

The eight impurities chosen for spectral matching were present in varying levels in the four lots. Most were lower than 0.10% by area (50 ng on column based on the assumptions mentioned earlier). Many were below 0.05% by area. Despite these low levels, very probable matches (M.F. > 990) to the expected spectra in the library were obtained in almost all cases. For example, impurity B was present at levels of 0.23, 0.03 and 0.04% by area in lots 5, 7, and 8, respectively. Impurities 5B, 7B and 8B each gave match factors of 995 or greater to the 3B impurity in the spectral library, for both the HP1100 and the HP1090. The 3B impurity itself (present at 0.08% by area) was confirmed to be compound PC based on its match factor of 999 when compared to the reference material spectrum. Although not included in Table 1, the match factors for 5B, 7B, and 8B to the PC standard were also greater than 990.

A very important observation to note is that match factors to other spectra in the library, to which no match was expected, were usually below 950 and are not included in Table 1. The exceptions (when an unexpected match gave a match factor greater than 950) are discussed later in Section 3.4.

3.3. Sensitivity

For spectral matching, 'sensitivity' might be defined as the ability to produce a very probable match between two spectra of the same compound when the compound is present at low concentrations. Sievert and Drouen have suggested using the term 'match threshold' for this [7]. The match threshold is expected to vary from compound to compound. For a particular compound, it could be improved by injection of a

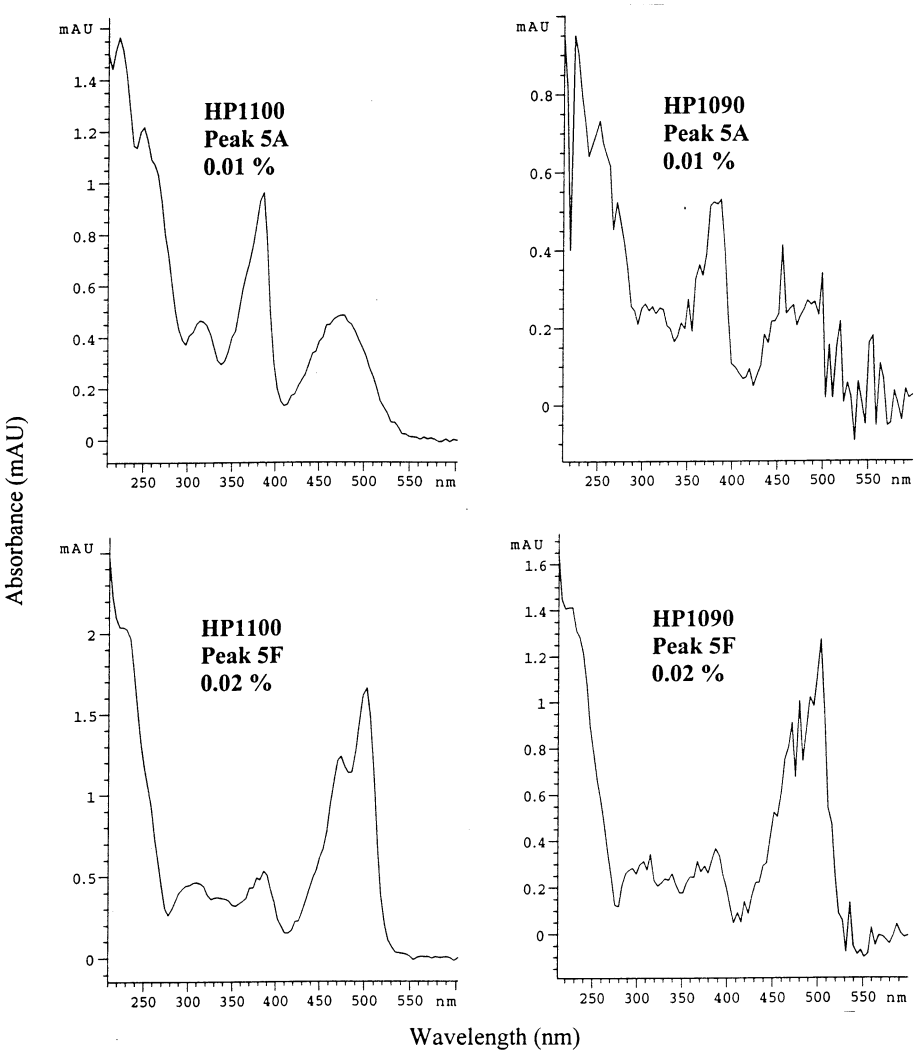

Fig. 5. LC/UV spectra of peaks 5A and 5F from two different diode array detectors. See Section 2 for a description of the differences between the detectors.

more concentrated solution, if solubility is favorable; or by injecting a larger volume, if the chromatography is not affected.

The data in Table 1 show that, using the HP1100, the spectral matching technique can be used for these impurities even at levels below 0.05% by area. It was only with peak 8F at 0.001% by area (500 pg on column based on the assumptions mentioned previously) that a less than very probable match was obtained. The match thresholds for the HP1090 were not as low. The low level impurities E, F, and G in lots 7 and 8, which were present at less than 0.01% by area on the HP1100, were not even detected on the HP1090 and so no spectra were generated. And the spectra of two other impurities, 5F and 5A, detected on the HP1090 at levels of 0.02 and 0.01% by area respectively, failed to match the expected library spectra (M.F. < 950). An examination of the HP1090 spectra for 5A and 5F (see Fig. 5) did reveal significantly poorer signal-to-noise ratios when compared to the HP1100 spectra.

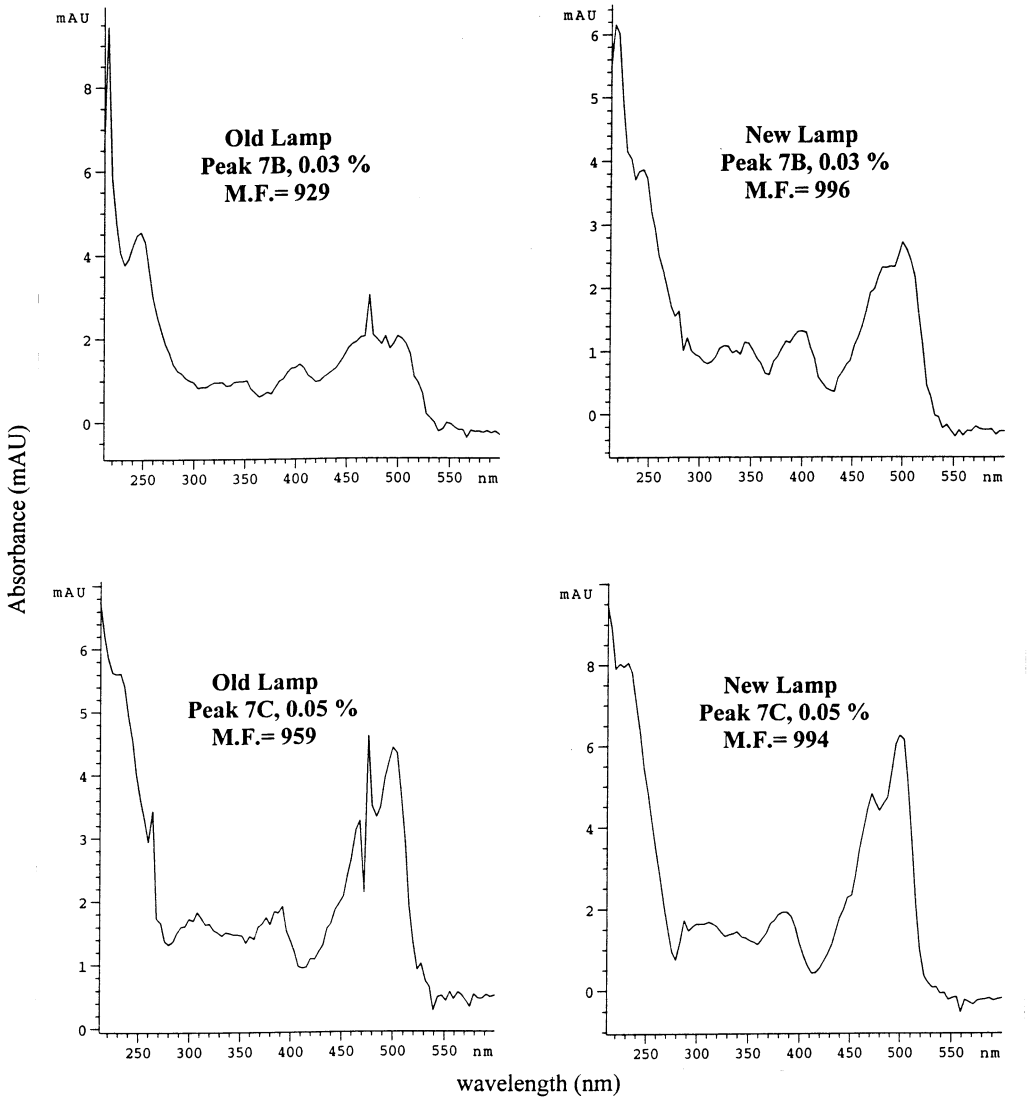

Fig. 6. LC/UV spectra peaks 7B and 7C on an HP1090 with and old versus new detector lamp.

The greater sensitivity observed when using the HP1100 was not a surprise because, according to the manufacturer, the HP1100 detector was designed to be more sensitive than the HP1090. In addition, the particular HP1090 used in this study was equipped with a 6-nm pathlength micro-flow cell, unlike the HP1100 which had the standard 10-nm pathlength flow cell.

Another factor that has been shown by others to influence match factor sensitivity is lamp intensity [7]. The influence of lamp intensity was also observed in this study. Fig. 6 shows spectra for impurities 7B and 7C generated on the HP1090 equipped with a new lamp compared to the spectra obtained when the lamp was an old one (about 50% of the intensity of a new lamp). The spectra obtained with the old lamp show increased noise (relatively large, sharp spikes and dips) which would be expected to affect spectral matching. Note in Fig. 6 that with the old lamp, 7B's spectrum did not match (M.F. = 929) the 3B library spectrum and 7C's spectrum gave only a

Table 2
Unexpected UV spectral matches of DuP 941 impurities

Peak[a]	Area %	Expected match	Match factor	Unexpected match	Match factor
5C	0.07	3C	1000	7H[b]	977
7C	0.05	3C	1000	7H	978
8H	0.04	7H	1000	3C	979
5C	0.07	3C	1000	3F[c]	999
7C	0.05	3C	1000	3F	999
8C	0.12	3C	1000	3F	998
5F	0.02	3F	999	3C	998
7F[d]	0.004	3F	995	3C	994
8F[d]	0.001	3F	977	3C	979
5F	0.02	3F	999	7H	966
7F[d]	0.004	3F	995	7H	964
8H	0.04	7H	1000	3F	967

[a] Peak designation codes: first character is lot number; second character is presumed identity.
[b] Peaks C and H are the positional isomers LS and SL, respectively.
[c] Impurity F is possibly an N-BOC compound very similar in structure to impurity C.
[d] The assignment of F to these peaks in lots 7 and 8 is tentative and probably incorrect. See text.

possible match (M.F. = 959) to the 3C library spectrum. However, with a new lamp both 7B and 7C gave very probable match factors (996 and 994, respectively) to the expected library spectra. Additional data analogous to that shown in Table 1 were obtained with the old lamp but not reported here. The conclusion reached from that data was that the spectral matching approach was unreliable for impurities below 0.1% by area. Therefore, using a good detector lamp in the most sensitive detector available can be critical to the success of the spectral matching technique when examining impurities at these low levels.

3.4. Selectivity

For match factor analysis to be of utility in monitoring drug substance impurities, not only must the technique be sensitive, but it must also be able to discriminate between very similar spectra. For most of the impurities examined in this study, which were all spectrally similar, the technique did prove to be very discriminating. Most impurities examined gave very probable matches only to the expected library spectrum. Also of significance was the lack of even possible matches (match factors between 950 and 989) to other library spectra. It was anticipated, though, that spectral matching might have difficulty distinguishing impurities that were positional isomers.

In this study two impurities, C and H, were known to be positional isomers. These impurities had been isolated and standards of them had been synthesized. See structures of LS (peak C) and SL (peak H) in Fig. 1. As the HP1100 data in Table 2 show, spectral matching was able to distinguish between these two isomers. Even though with this pair of impurities possible matches (M.F. = 977 to 979) were obtained to the wrong isomer (5C, 7C, 8C to 7H; and 8H to 3C), very probable matches (M.F. = 1000) were obtained to the correct isomer.

However, there was one pair of impurities that spectral matching was not able to differentiate. As Table 2 shows, spectral matching was unable to distinguish between impurities C and F. It is interesting to note that LC/MS/MS analysis of impurity F indicated that it may be an N-BOC compound, very similar in structure to reference standard LS (impurity C) [10]. Impurities 5C, 7C, and 8C each gave very probable matches to the library spectra of 3F as well as to 3C. Similarly, impurities 5F and 7F also gave very probable matches not only to 3F, but also to 3C. The ambiguity in this case can be resolved by a simple check of the retention times of the impurities. In

all of the chromatograms of the various drug lots (see Fig. 2) impurities C and F are separated by a minimum of 2.3 min. So by comparing retention times as well as spectral match factors, impurities C and F should be able to be distinguished from one another.

Another pair of impurities that is compared in Table 2 is impurities H and F. Impurity H is an isomer of impurity C, but whereas spectral matching was unable to distinguish between C and F, spectral matching did distinguish between H and F. Impurities 5F and 7F gave very probable matches to 3F but only possible matches to 7H. Likewise, 8H gave a very probable match to 7H but only a possible match to 3F.

Lastly, in the case of impurity 8F, which was present at only 0.001% by area, no very probable matches were obtained. Only possible matches were obtained, and to three different library spectra, 3F, 3C, and 7H. As was noted above, LC/MS/MS analysis of these same lots of DuP 941 provided evidence that impurity F is an N-BOC compound. If that is true then lots 7 and 8 should not contain any of impurity F because N-BOC protecting group chemistry was not used subsequent to the synthesis of lot 5. In fact, LC/MS/MS did not detect impurity F in either lot 7 or 8 [10]. That evidence suggests that although there was a very small peak in the UV chromatograms of lots 7 and 8 which had a similar retention time to impurity 3F, UV diode array spectral matching incorrectly indicated very probable or possible matches to the UV spectrum of impurity 3F. Fortunately, the tracking of impurities at such low levels as these is not a necessity.

4. Conclusions

The technique investigated, UV spectral matching, was found to be extremely powerful for monitoring the impurity profile of a drug substance. In attempting to confirm that trace impurities (less than 0.1% by area) in a drug substance lot were the same impurities as those found in an earlier lot, we have demonstrated that UV spectral matching has excellent sensitivity, even down to 0.05% by area. The ability of spectral matching to discriminate between structurally similar compounds and even positional isomers was also shown to be very good. In the one case where discrimination was not achieved, the use of retention times and elution sequence was able to resolve the ambiguity. However, we have found that at levels much lower than 0.05% by area, using spectral matching and retention time only can lead to incorrect identity conclusions. If one would desire to monitor impurities present at much less than 0.05% by area it may be advisable to obtain additional evidence, such as LC/MS/MS, to confirm that an impurity is the same as in an earlier lot. We suggest the reader refer to Part II of our work on impurity profiling [10].

The particular diode array optical bench and flowcell being used, along with the age of the detector lamp, are critical factors in determining to what level UV spectral matching can be used to monitor trace impurities. An HP1100 DAD equipped with a standard flowcell was shown to be much more sensitive than an HP1090 DAD equipped with a micro flowcell. It was also shown that use of an old detector lamp, having reduced intensity, can have a serious adverse effect on the success of spectral matching. Therefore, validation of a spectral matching method for ruggedness and sensitivity is required before using the technique to monitor trace impurities.

In the absence of reference materials for the impurities in a drug substance, UV diode array spectral matching can be a very useful tool, when coupled with retention time and elution sequence information, to more accurately determine whether a chromatographic peak is a new or recurring impurity.

Acknowledgements

The authors wish to thank Dr. Jim Shea for providing samples of the lots of DuP 941 and the known impurities which were used in this study.

References

[1] International Conference on Harmonization Quality Guideline III/5442/94-EN, Impurities in New Drug Substances, in: The Federal Register 61 (1996) 372–374.
[2] D. Giron, Boll. Chim. Farm. 133 (1994) 201–220.
[3] T. Alfredson, T. Sheehan, J. Chromatogr. Sci. 24 (1986) 473–482.
[4] D.W. Hill, T.R. Kelley, K.J. Langner, Anal. Chem. 59 (1987) 350–353.
[5] D.M. Demorest, J.C. Fetzer, I.S. Lurie, S.M. Carr, K.B. Chatson, LC-GC 5 (1987) 128–142.
[6] R.W. Giuffre, in: L. Huber, S. George (Eds.), Diode Array Detectors in HPLC, Marcel Dekker, New York, 1993, pp. 163–197.
[7] H.J.P. Sievert, A.C.J.H. Drouen, in; L. Huber, S. George (Eds.), Diode Array Detectors in HPLC, Marcel Dekker, New York 1993, pp. 51–126.
[8] G. Cavina, L. Valvo, B. Galinella, R. Porra, A.L. Savella, J. Pharm. Biomed. Anal. 10 (1992) 437–446.
[9] G. Cavina, R. Alimenti, B. Galinella, L. Valvo, J. Pharm. Biomed. Anal. 10 (1992) 685–692.
[10] E.C. Nicolas, T.H. Scholz, J. Pharm. Biomed. Anal. 16 (1998) 825–836.