Pergamon

PII: S0305-1978(96)00087-7

Biochemical Systematics and Ecology, Vol. 25, No. 1, pp. 65–71, 1997

0305-1978/97 $17.00+0.00

Terverticillate Penicillia Studied by Direct Electrospray Mass Spectrometric Profiling of Crude Extracts. II. Database and Identification

JØRN SMEDSGAARD

Department of Biotechnology, Build. 221, DTU-Technical University of Denmark, DK-2800 Lyngby, Denmark

Key Word Index—*Penicillium*; chemotaxonomy; electrospray mass spectrometry; ESMS; mass profile; database; identification.

Abstract—A mass spectral database was built using standard instrument software from 678 electrospray mass spectra (mass profiles) from crude fungal extracts of terverticillate taxa within the genus *Penicillium*. The match factors calculated from searching all the mass profiles stored in the database were used as a distance measure for grouping into taxa. Isolates from more than 75% of the taxa were placed correctly, that is with the highest match factors to other isolates of the their own species by library searches. The database was used as an expert system based on correct identification of the isolates stored according to classical taxonomic criteria. Mass profiles collected in previous studies could be identified by a search in the database.

Introduction

Two problems are often encountered in fungal taxonomy: has this culture or metabolite profile been seen before and to which taxon does it belong? The latter question might be straightforward, but rather laborious for the trained taxonomist, but it can be very difficult for a novice. The question of a novel metabolite profile on the other hand is often a matter of memory and notes. The determination of secondary metabolite profiles using either TLC, HPLC or electrospray mass spectrometry (ESMS) (Filtenborg *et al.*, 1983; Frisvad and Thrane, 1987; Smedsgaard, 1997) have proved to be of considerable help in the identification of cultures within the genus *Penicillium* (Frisvad and Filtenborg, 1989; Svendsen and Frisvad, 1994; Smedsgaard and Frisvad, 1996). In cases where a culture does not fit into one of the described taxa, the cultural information including the metabolite profile have to be stored until other similar isolates are found and a decision made as to whether a new taxon can be described. While physiological data can be easily stored and searched in databases, TLC profiles or HPLC chromatograms are not easily stored in and retrieved from databases. Searching a database with full chromatograms is very difficult as no commercial software is available for that purpose. Mass spectra determined from crude plug extracts made directly from identification cultures by ESMS can very easily be used as mass profiles to create a database as described by Smedsgaard and Frisvad (1996) and Smedsgaard (1997). It has furthermore been demonstrated that substantial taxonomic information is contained within the mass profiles, as approximately 75% of over 50 taxa described within the terverticillate penicillia could be segregated into taxa by cluster analysis based on complete mass profiles (Smedsgaard and Frisvad, 1997).

(*Received* 20 *June* 1996; *accepted* 9 *September* 1996)

Almost all software packages for instrument control and data processing supplied with modern mass spectrometers include options that allow the user to create and search libraries constructed from stick (centroid) mass spectra (Warr, 1991). Although the library is designed for storage of classical electron impact (EI) spectra from pure compounds, it can be used for spectra obtained from highly complex mixtures using other ionisation techniques. The most common search method uses probability based matching (PBM) developed by McLafferty and coworkers (McLafferty *et al.*, 1974; Pesyna *et al.*, 1976; Atwater *et al.*, 1985; Warr, 1991), although the precise algorithm is not always available from the different manufacturers. A library is normally constructed from reduced spectra, e.g. containing the 50 most intense nominal mass peaks and an index library with the eight most intense peaks from each spectrum (Warr, 1991). Search routines will also include weighting giving higher weight to the higher mass when calculating the quality of the fit. Search results are ranked on the basis of matching factors calculated forward (how likely it is that the unknown spectrum is a pure sample of the library spectrum; thus all ions in the unknown spectrum are considered) or reversed (how likely it is that the unknown spectrum contains the library spectrum; thus only ions in the library spectrum are considered). The reversed fit is the most usable ranking for mass profiles because there will always be more peaks in an unknown spectrum than in a library spectrum (with a maximum of 50 peaks). The reversed fit will only consider mass peaks in the library spectrum in match factors and rankings. A match factor of 1000 is maximum corresponding to a perfect match.

A database system constructed from mass profiles of crude extract can be considered either as an expert system or as a memory aid to find a previously encountered similar isolate (mass profile). Expert knowledge added to the database is in this case the correct identification of the cultures according to an accepted taxonomy. As described by Smedsgaard and Frisvad (1996b), the analyst can determine a mass profile from crude fungal extracts of unknown cultures and thus obtain a preliminary identification by a simple search in the database. The quality of the search results will, however, be dependent on the quality of the information stored in the database.

The aim of this study was to evaluate the usability of the mass spectral library facilities, included with a standard mass spectrometer, as an aid to identification within a large group of closely related taxa.

Material and Methods

Mass spectra (mass profiles) were determined as previously described in part I (Smedsgaard and Frisvad, 1997) of this paper using the method described in Smedsgaard and Frisvad (1996)Smedsgaard (1997). The procedure is in short: three plugs cut from seven day-old cultures grown on either Czapek Yeast Autolysate agar (CYA) or Yeast Extract Sucrose agar (YES) were extracted ultrasonically in 0.5 ml solvent (dichlormethane:methanol:ethylacetate, 2:1:3, with 0.5% formic acid). The solvent was evaporated under a gentle stream of nitrogen and the residues were redissolved in 0.4 ml 75% methanol with 0.6% formic acid and 0.02% HCl ultrasonically and filtered. The samples were analysed by direct injection of 5 µl extract into the electrospray ion source of a single stage quadropole mass spectrometer (Fisons/VG TRIO 2000) using 90% methanol as carrier at a flowrate of 6 µl/min. All instrumental parameters were optimized to reduce source fragmentation and reactions and maximize sensitivity. Twenty-four continuum scans (3 s pr. scan) were summarized to a continuum spectrum to reduce noise, followed by a background subtraction, smoothing (Savitzky-Golay, three times with a peak width of 0.75 amu) and calculating the centroid spectrum. The centroid spectra were added to a database created using the facilities in the instruments operating software (MassLynx 2.00). Three hundred and thirty-nine isolates of the genus *Penicillium* were selected from the IBT culture collection at the Department of Biotechnology, DTU, representing the taxa described in subgenus *Penicillium* with a predominantly terverticillate conidiophore and identified according to morphology, physiology and secondary metabolites. A total of 678 mass profiles were added to the database.

Results and Discussion

A mass profile from *P. scabrosum* grown on CYA is shown in Fig. 1 as a continuum spectrum (top) and the centroid (stick) spectrum (bottom). Distinct ions are seen in the profile corresponding to the protonated mass of the following metabolites produced by *P. scabrosum*: 238 amu to viridicatin; 252 amu to 3-methoxy-viridicatin and 254 amu to viridicatol. Other ions seen in the spectrum are from either unknown metabolites or background ions (noise in the low mass part of the spectrum). As seen in the centroid spectrum, there are large numbers of ions, typically in the range between 100 and 300 ions in a mass profile. The database will only contain the largest 50 ions, thus small mostly noise ions are eliminated when the spectrum is added to the database. Searching this *P. scabrosum* mass profile (IBT 6637 from CYA), already stored in the database, will

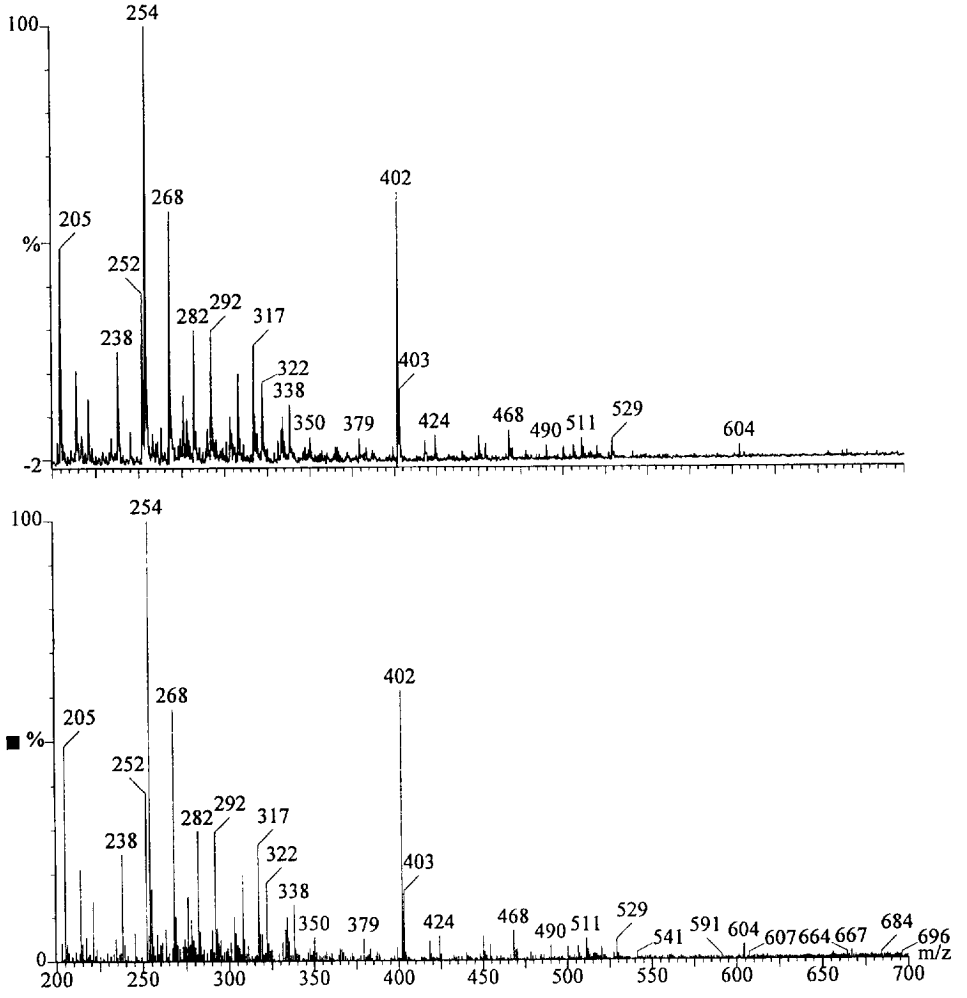
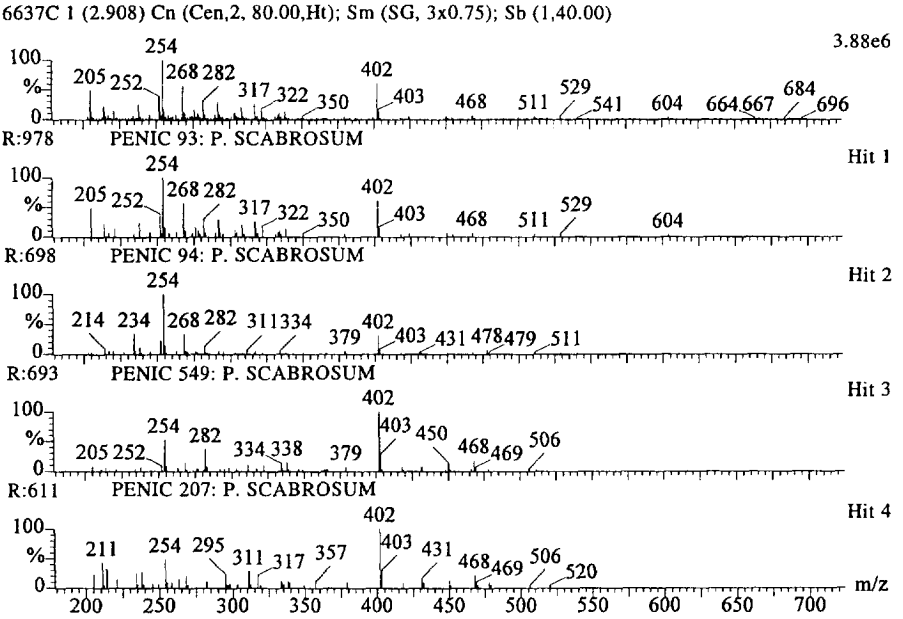


FIG. 1. MASS PROFILE FOR *P. SCABROSUM* (IBT 6637) GROWN ON YES. Ions corresponding to the protonated molecular mass ($M+H^+$) of the following metabolites can be found in the profile: viridicatin (237 amu), 3-methoxy-viridicatol (252 amu), viridicatol (254 amu), cyclopeptin (281 amu), cyclopenin (295 amu), cyclopenol (311 amu), 8-*O*-demethyl-pseurotin A (418 amu) and pseurotin A (432 amu).

result in a typical database search report from the MassLynx software as shown in Fig. 2. The report shows the searched mass profile at the top followed by the three best fitting mass profiles (shown with fit number 2 at the top, as the best fit is the mass profile searched). The results are listed at the bottom of the report where the standard compound name has been used for isolate name. The first figures in CAS (Chemical Abstract Service) number system, which is a part of the database software, is used for collection reference numbers, the middle figures for substrate information (10 = CYA, 11 = YES) and the last figure for identity status (0 = identity ok, 9 = new taxa). The results are ranked after reversed match, but the forward match factor is included in the list. The difference between the reversed and forward match factor for hit number 1 is due to ions omitted when the spectrum was stored in the library. This difference can be used as a



Hit	Compound Name	CAS	REV	for
1	P. SCABROSUM	6637-10-0	978	961
2	P. SCABROSUM	6637-11-0	698	579
3	P. SCABROSUM	16088-11-0	693	561
4	P. SCABROSUM	12258-10-0	611	421
5	P. CLAVIGERUM	5523-10-0	588	441
6	P. NEOECHINULATUM	5600-10-0	573	336
7	P. NEOECHINULATUM	5595-10-0	564	374
8	P. SCABROSUM	17219-10-0	545	332
9	P. CYCLOPIUM	12841-10-0	529	313
10	P. SP 7.	18283-10-9	524	116
11	P. SOLITUM	11993-10-0	504	128
12	P. SCABROSUM	12258-11-0	500	355
13	P. FREII	15374-11-0	499	239
14	P. SP	14074-10-9	487	139
15	P. FREII	15374-10-0	485	278

FIG. 2. DATABASE SEARCH REPORT OF *P. SCABROSUM* IBT 6637. The report shows the unknown spectra at the top followed by the best hits. Since IBT 6637 is already included in the database, the best fit is the isolate itself (match 978) followed by other *P. scabrosum* isolates. A list of the search results is shown at the bottom.

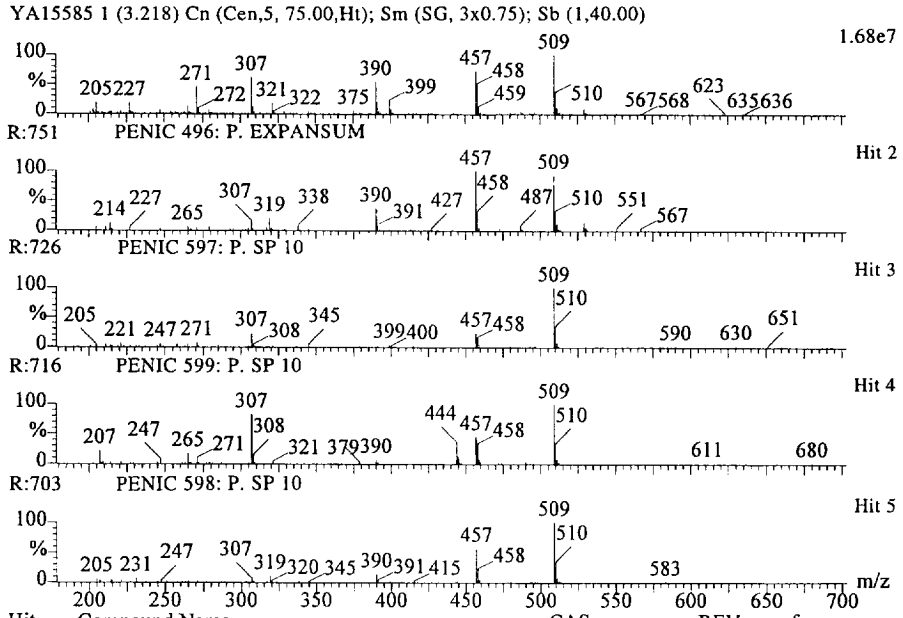
TABLE 1. SUMMARY OF DATABASE SEARCHES OF ALL *P. ITALICUM* ISOLATES GROWN ON CYA. THE MASS PROFILE USED IN THE SEARCH IS ALWAYS FOUND WITH THE HIGHEST RANK. Only four *P. italicum* isolates are included in the database; those other isolates will be found between the six first hits. Reversed ranking is used

		Isolates searched			
	Rank	12955 CYA	13005 CYA	13067 CYA	15661 CYA
1	IBT no.	12955 CYA	13005 CYA	13067 CYA	15661 CYA
	name	*P. italicum*	*P. italicum*	*P. italicum*	*P. italicum*
	score	956	980	973	970
2	IBT no.	13005 CYA	15661 CYA	13005 CYA	13005 CYA
	name	*P. italicum*	*P. italicum*	*P. italicum*	*P. italicum*
	score	822	848	753	853
3	IBT no.	15661 CYA	12955 CYA	13067 YES	12955 CYA
	name	*P. italicum*	*P. italicum*	*P. italicum*	*P. italicum*
	score	814	825	733	797
4	IBT no.	18387 CYA	13067 YES	15661 CYA	15661 YES
	name	*P. sp 28*	*P. italicum*	*P. italicum*	*P. italicum*
	score	677	745	669	752
5	IBT no.	13067 CYA	15661 YES	12955 CYA	13067 YES
	name	*P. italicum*	*P. italicum*	*P. italicum*	*P. italicum*
	score	611	720	641	700
6	IBT no.	13592 CYA	13067 CYA	15661 YES	18387 CYA
	name	*P. camemberti*	*P. italicum*	*P. italicum*	*P. sp 28*
	score	581	667	638	651

purity measure. A library search with a mass profile stored in the database was found to give the mass profile itself as the best result with a match over 900 followed by other mass profiles. A search without limits on match factors will always give a search result if just one ion in the unknown mass profile is found in a library mass profile, even though no other similarity can be seen.

TABLE 2. SUMMARY OF DATABASE SEARCHES OF *P. CAMEMBERTI* SHOWING THE WORST RESULTS OBTAINED. Only three out of five searched isolate gave the right identity in line two, not counting the isolate itself. Reversed ranking is used

		Isolates searched				
Rank		3505 CYA	11570 CYA	13592 CYA	14856 CYA	15441 CYA
1	IBT no.	3505 CYA	11570 CYA	13592 CYA	14856 CYA	15441 CYA
	name	*P. camemberti*	*P. camemberti*	*P. camemberti*	*P. camemberti*	*P. camemberti*
	score	909	972	976	955	979
2	IBT no.	6774 CYA	13592 CYA	3505 CYA	13592 CYA	10924 CYA
	name	*P. nalgiovense*	*P. camemberti*	*P. camemberti*	*P. camemberti*	*P. commune*
	score	716	745	727	744	863
3	IBT no.	10837 CYA	14745 CYA	10837 CYA	14856 YES	15441 YES
	name	*P. commune*	*P. verrucosum*	*P. commune*	*P. camemberti*	*P. camemberti*
	score	711	731	719	742	812
4	IBT no.	13592 CYA	15170 CYA	14856 CYA	10837 CYA	6774 CYA
	name	*P. camemberti*	*P. solitum*	*P. camemberti*	*P. commune*	*P. nalgiovense*
	score	685	716	716	734	781
5	IBT no.	3468 CYA	12803 CYA	6071 CYA	13042 CYA	16848 CYA
	name	*P. commune*	*P. verrucosum*	*P. monomematosum*	*P. nalgiovense*	*P. sp. 17*
	score	681	712	689	727	775
6	IBT no.	10924 CYA	15685 CYA	13042 CYA	13592 YES	16710 CYA
	name	*P. commune*	*P. solitum*	*P. nalgiovense*	*P. camemberti*	*P. commune*
	score	658	680	676	659	761



Hit	Compound Name	CAS	REV	for
3	P. SP 10	16712-11-9	726	647
4	P. SP 10	16715-11-9	716	617
5	P. SP 10	16715-10-9	703	599
6	P. SP 10	16712-10-9	698	565
7	P. EXPANSUM	15717-10-0	694	581
8	P. SP 10	16716-10-9	685	588
9	P. EXPANSUM	16705-10-0	552	388
10	P. EXPANSUM	16943-10-0	551	388
11	P. SP 10	16716-11-9	548	475
12	P. EXPANSUM	15685-11-0	524	485
13	P. CRUSTOSUM	10513-10-0	504	84
14	P. VENETUM	16215-10-0	458	71
15	P. CRUSTOSUM	16510-10-0	457	83
16	P. CRUSTOSUM	14747-10-0	429	82
17	P. CLAVIGERUM	4899-10-0	418	105

FIG. 3. MASS PROFILE SEARCH REPORT FOR *P. EXPANSUM* COLLECTED IN ANOTHER STUDY. *P. sp 10* is a taxa not yet described but closely related to *P. expansum*.

Performing systematic searches of all mass profiles used to create the database allowed a grouping of isolates where the match factors can be used as a reversed distance. Thus the higher the match factor, the closer are the isolates. Table 1 illustrates a typical set of results obtain from searches of mass profiles for *P. italicum* grown on CYA. The first hits are the isolate itself followed by other correct matches. Using the second line it can be seen that isolate IBT 13005 is forming a centre point as it is nearest to all other isolates. Looking down the IBT 13005 column the nearest neighbour is 15661 followed by 12955, 13067 on YES, 15661 on YES and 13067, all identified as *P. italicum*. Typically some differences are seen in mass profiles of *P. italicum* grown on CYA and YES; thus the results are dominated by mass profiles from CYA. Only four isolates of *P. italicum* on CYA and YES are included in the database, and therefore other taxa are expected to appear in the search reports. Results in Table 1 are for more than 75% of the

taxa in this study. These findings include all taxa that were considered grouped in the first part of the study (Smedsgaard and Frisvad, 1996a). The poorest result was obtained from searches with isolates of *P. camemberti* as shown in Table 2. However these results are better than those obtained from the cluster analysis described in part I of this paper (Smedsgaard and Frisvad, 1996a). Perfect search results were obtained for *P. olsonii*, as only *P. olsonii* isolates, where there was a maximum of 10 correct fits found (five isolates grown on two substrates).

Search results from a mass profile of *P. expansum* cultivated on YES collected in a previous study and not included in the database are shown in Fig. 3. Two *P. expansum* mass profiles are found as best hits followed by several mass profiles from a taxon not yet described, but closely related to *P. expansum*. These profiles from the undescribed taxon were originally added to the database as a memory aid to see if more isolates were encountered.

Using the MS database system as an expert system demands that correctly identified isolates are entered to the system. The database system included with mass spectrometers is normally designed to use only one parameter: the mass spectrum. There has been very little development in the PBM routine used as the search routine in mass spectral databases. An improved routine that would allow other parameters to be included in the search such as simple physiological characters could form a strong basis for an expert system.

References

Atwater, B. L., Stauffer, D. B. and McLafferty, F. W. (1985) Reliability ranking and scaling improvements to the probability based matching system for unknown mass spectra. *Anal. Chem.* **57**, 899–903.

Filtenborg, O., Frisvad, J. C. and Svendsen, J. A. (1983) Simple screening method for moulds producing intracellular mycotoxins in pure cultures. *Appl. Environ. Microbiol.* **45**, 581–585.

Frisvad, J. C. and Filtenborg, O. (1989) Terverticillate Penicillia: Chemotaxonomy and mycotoxin production. *Mycologia* **81**, 837–861.

Frisvad, J. C. and Thrane, U. (1987) Standardized high-performance liquid chromatography of 182 mycotoxins and other fungal metabolites based on alkylphenone retention indices and UV-VIS spectra (diode array detection). *J. Chromatogr.* **404**, 195–214.

McLafferty, F. W., Hertel, R. H. and Villwock, R. D. (1974) Probability based matching of mass spectra. *Org. Mass Spectrom.* **9**, 690–702.

Pesyna, G. M., Venkataraghavan, R., Dayringer, H. E. and McLafferty, F. W. (1976) Probability based matching system using a large collection of reference mass spectra. *Anal. Chem.* **48**, 1362–1368.

Smedsgaard, J. and Frisvad, J. C. (1996) Using direct electrospray mass spectrometry in taxonomy and secondary metabolite profiling of crude fungal extracts. *J. Microbiol. Meth.* **25**, 5–17.

Smedsgaard, J. (1997) Micro-scale extraction procedure for standardized screening of fungal metabolite production in cultures. *J. Chromatogr. A* (in press).

Smedsgaard, J. and Frisvad, J. C. (1997) Terverticillate Penicillia studied by direct electrospray mass spectrometric profiling of crude extracts: I. Chemosystematics. *Biochem. Syst. Ecol.* (in press).

Svendsen, A. and Frisvad, J. C. (1994) A chemotaxonomic study of the terverticillate penicillia based on high performance liquid chromatography of secondary metabolites. *Mycol. Res.* **98**, 1317–1328.

Warr, W. A. (1991) Spectral Databases. *Chemometric. Intell. Lab. Syst.* **10**, 279–292.