Discovery of New Natural Products by Application of *X-hitting*, a Novel Algorithm for Automated Comparison of Full UV Spectra, Combined with Structural Determination by NMR Spectroscopy

Thomas Ostenfeld Larsen,*,† Bent O. Petersen,‡ Jens Ø. Duus,‡ Dan Sørensen,†,§ Jens C. Frisvad,† and Michael E. Hansen†

Center for Microbial Biotechnology, BioCentrum-DTU, Søltofts Plads, Technical University of Denmark, DK-2800 Kgs., Lyngby, Denmark, and Carlsberg Research Laboratory, Gamle Carlsberg Vej 10, DK-2500 Valby, Denmark

Received December 30, 2004

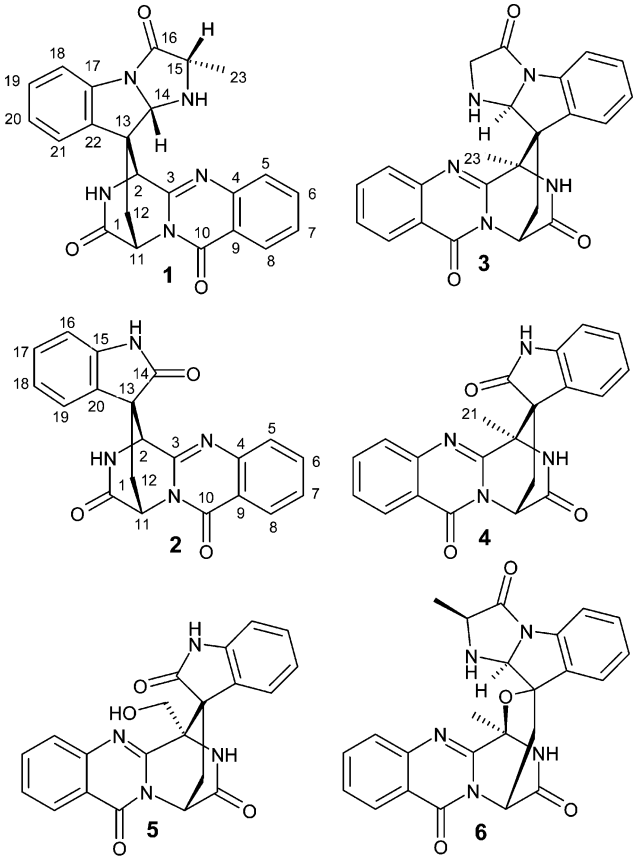
X-hitting, a newly developed algorithm for automated comparison of UV data, has been used for the tracking of two novel spiro-quinazoline metabolites, lapatins A (**1**) and B (**2**), in a screening study targeting quinazolines. The structures of **1** and **2** were elucidated by analysis of spectroscopic data, primarily 2D NMR.

The task of finding novel bioactive natural products is usually bioassay driven. One good example is the research by the National Cancer Institute (USA), which has been a major driving force for both industries and academia in the search for novel cancer agents for the past many years with the finding of taxol as one of the highlighted outcomes.[1] Often a certain type of compound turns out to be active in an assay. Two classic examples are the statins inhibiting cholesterol biosynthesis and the β-lactams inhibiting cell wall biosynthesis in Gram-positive bacteria.[1] More recently bioactive quinazoline compounds, reported from microfungi belonging to the genera *Penicillium* and *Aspergillus*, have attracted attention.[2,3] Spiroquinazoline (**3**) inhibits the function of substance P taking part in the inflammatory response,[2] whereas alantrypinone (**4**) and serantrypinone (**5**), originally described by Larsen and co-workers,[4,5] turned out to be anti-GABAergic insecticidal alkaloids.[3]

When having generated a promising hit in a bioassay, the normal procedure in the drug discovery process usually is to produce a large number of structurally analogous compounds either by traditional chemical synthesis or by combinatorial chemistry in order to study structure−activity relationships and to find even more active lead compounds. Alternatively to chemical synthesis of analogues, nature can be explored for structurally similar compounds by continued bioguided screening or by UV-spectroscopy guided screening.[6−9]

Quinazoline metabolites such as alantrypinone,[4] anacine and verrucines,[10] fiscalins,[11] fumiquinazolines,[12] and glyantrypine[13] all have very similar and characteristic UV spectra due to their anthranilic acid derived chromophore system and can therefore be tracked by UV-guided analysis for either analytical or preparative purposes.

The use of UV maximum and minimum absorption data for pure compounds is a well-known feature in connection with searching in databases such as *SciFinder*,[14] *Antibase*,[15] and *Marinlit*.[16] More recently a new method for the systematic and automated computer-assisted search of full UV spectra in a large number of datafiles for new natural products was introduced based on the new mathematical algorithm *X-hitting*.[17] In *X-hitting* known compounds are tracked by the feature *cross-hitting*, whereas likely new compounds are indicated by the *new-hitting* feature of *X-hitting*. The algorithm behind *X-hitting* compares the shape between two spectra and returns a similarity index describing how statistically similar the two spectra are. To capture information about the actual shape of each of the profiles at different scales (courser or finer details), a linear combination of the correlation between higher order derivatives is used. Two of the main criteria are retaining relations between "neighboring" values in spectra and that the algorithm has to be fast to compute. Using derivatives, these relationships can be incorporated by measuring the differences and the positions of topologies and extreme



* To whom correspondence should be addressed. Phone: +45 45 25 26 32. Fax: +45 45 88 49 22. E-mail: tol@biocentrum.dtu.dk.
† BioCentrum-DTU.
‡ Carlsberg Laboratory.
§ Present address: Department of Medicinal Chemistry, Merck Frosst Centre for Therapeutic Research, P.O. Box 1005, Pointe-Claire, Dorval H9R 4P8, Quebec, Canada.

10.1021/np040248s CCC: $30.25
Published on Web 05/28/2005

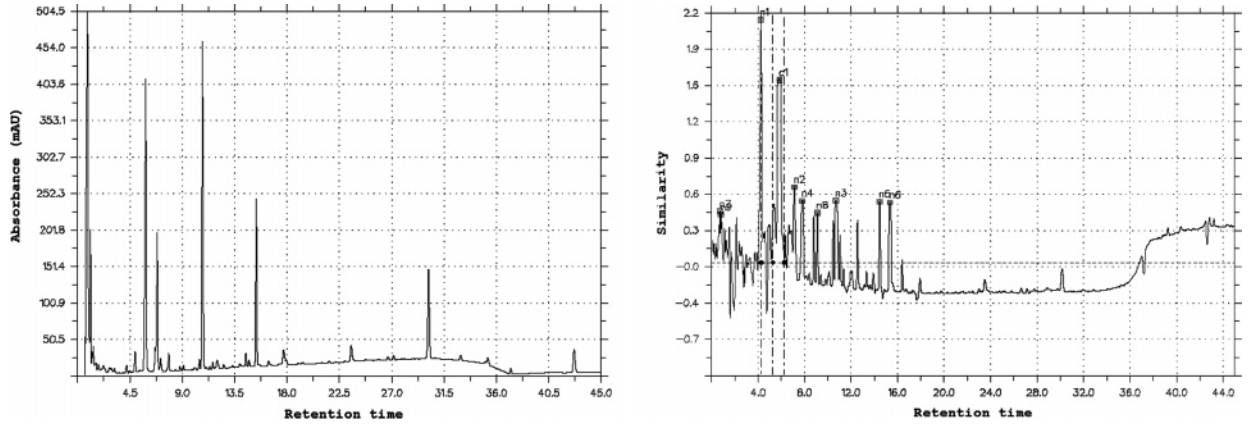


Figure 1. (Top) HPLC chromatogram showing the profile of metabolites present in a *Penicillium lapatayae* extract. (Bottom) Similarity profile for *P. lapatayae* Xhit# 97 versus alantrypinone (cross-hitting internal +,− 1 min). The peaks n1−8 indicate compounds with UV spectra very similar but not identical to that of alantrypinone, whereas peak c1 indicates the presence of alantrypinone (cross-hit) in the sample.

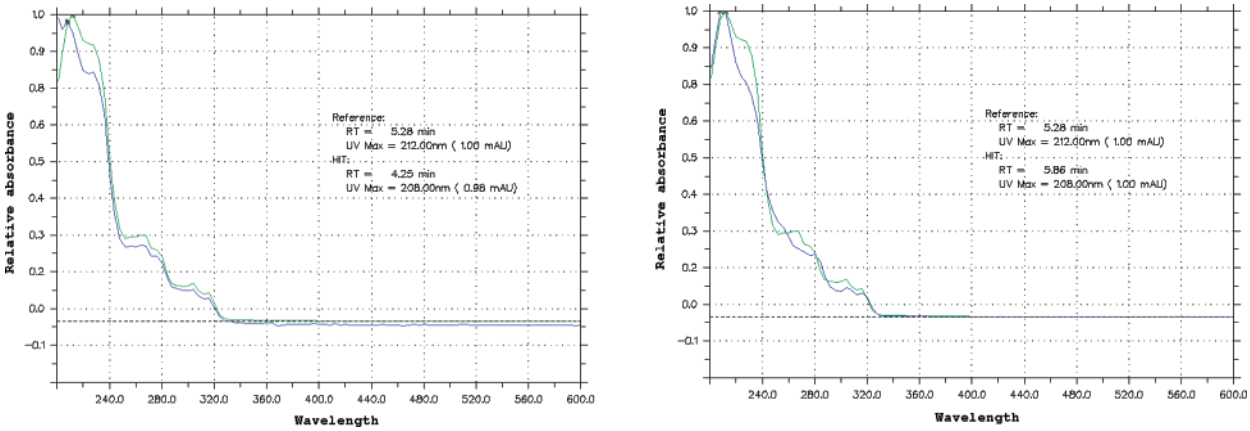


Figure 2. UV spectrum of the reference compound alantrypinone (in green), which is the target compound, and the blue UV spectra of the hits lapatin A (left) and lapatin B (right) seen as respectively peaks c1 and n1 in Figure 1.

points. Filtering techniques are applied to the derivatives in order to reduce the sensitivity to potential noise fragments in the spectra.[17]

The aim of this paper is to illustrate and prove the concept of the *new-hitting* feature of *X-hitting* by describing the finding of two new spiro-quinazoline metabolites, lapatins A and B, likely to be bioactive alkaloids due to their structural resemblance to known bioactive quinazoline metabolites.

Results and Discussion

When using the UV spectrum of alantrypinone as target spectrum for a cross search of all the UV data contained in the approximately 124 datafiles studied by Hansen et al.,[17] the similarity profile generated from a *P. lapatayae* extract showed several peaks (n1−8) with similarities significantly above the noise level, strongly indicating the presence of quinazoline-like compounds in the extract (Figure 1). In addition a compound (c1) eluting at 5.86 min fell into the (cross-hit) interval wherein the target compound alantrypinone was expected to appear if present in the extract.

However, since the similarity of the n1 spectrum toward the spectrum of alantrypinone was larger than that of c1 (Figure 2), the compound appearing at c1 could not be alantrypinone. On the basis of these findings it was decided to isolate and elucidate the structure of these two compounds, likely to have structures very similar to that of alantrypinone.

LC-MS of the major target c1 (**1**) indicated the formula $C_{23}H_{20}O_3N_5$ $(M + H)^+$, which upon search in *SciFinder* was found to match with the data for spiroquinazoline (**3**).[2] Thus the dereplication process supported the theory of the target being a quinazoline. By comparison of the chemical shift values of **1** with those of other quinazolines and in particular those of alantrypinone,[4] it was quite obvious that **1** indeed was a quinazoline metabolite. Thus the ^{1}H NMR data of compound **1** (Table 1) shows an amide proton, two 1,2 disubstituted benzene ring systems, and a $-CH-CH_2-$ sequence. Since both **3** and **4** also have a $-CH-CH_2-$ sequence originating from tryptophan, our data indicated that tryptophan is also a precursor for the biosynthesis of **1**. These fragments were confirmed by the ^{13}C data that showed three amide carbonyl groups and 13 aromatic carbons.

Comparison of the ^{13}C NMR data of **1** with those of spiroquinazoline (**3**) (Table 1) showed the two compounds to have very similar chemical shifts despite being recorded in different solvents. However, careful comparison of the ^{1}H NMR data of **1** (Table 1) with those of spiroquinazoline (**3**)[2] revealed that the two compounds had to be different since the C-23 methyl group appears as a singlet in **3**, whereas it appears as a doublet in **1**. The COSY data of **1** showed the protons of the C-23 methyl group to couple to the α-proton H-15 at 3.97 ppm, which further coupled to an NH resonating at 3.35 ppm. Both H-23 and H-15 showed HMBC correlations to the carbonyl C-16, establishing the fragment $-NH-CH(CH_3)-CO-$, strongly indicating the incorporation of alanine into **1** in a manner similar to that seen for fumiquinazoline C (**6**).[12]

As in alantrypinone (**4**), a NOE correlation could be seen between the CONH proton and the aromatic proton at C-21

Table 1. NMR Data for Lapatin A (**1**) (600.13 MHz (^{1}H), 150.92 MHz (^{13}C)) in DMSO-d_6 and Spiroquinazoline (**3**) in $CDCl_3$[2]

	spiroquinazoline (**3**)	lapatin A (**1**)				
position	^{13}C	^{13}C	^{1}H	J = Hz	NOE	HMBC to carbon #[a]
1	171.3	170.2				
2	61.8	56.6	4.156	3.9	23(m)	1,3,12,13,14,22
3	152.6	152.7				
4	146.7	147.5				
5	127.9	127.2	7.641	8.0	23(w)	7,9
6	134.7	134.5	7.843	7.6		4,8
7	127.7	126.7	7.561	7.6		5,9
8	127.2	126.3	8.186	8.0		4,6,10
9	120.6	120.2				
10	158.9	158.5				
11	52.1	53.1	5.483			1,3,10,12
12a	33.3	36.2	2.629	14.6, 2.4		1,2,13,14,22
12b			2.444	14.6, 3.2		1,11,13,14,22
13	56.4	51.4				
14	81.1	83.4	5.322	6.5, 1	12b(s), 12a(w)	2,12,13,16
15	52.8	58.8	3.972	7.2, 6.7, 1		13,16,23
16	170.1	171.0				
17	138.0	136.8				
18	116.2	114.3	7.461	7.6		20,22
19	129.4	129.1	7.401	7.8		17,21
20	126.4	124.7	7.232	7.5		18,22
21	124.0	125.9	7.163	7.5	12a(s)	13,17,19
22	135.7	138.0				
23	15.0	18.0	1.253	6.7	2	15,16
NH			3.348	6.5, 7.2	23(s), 12b(w)	2,3,11
CONH			9.182	3.9	21(m)	13,14,15,16

[a] The HMBC data are reported as from ^{1}H to ^{13}C.

Table 2. NMR Data for Lapatin B (**2**) (600.13 MHz (^{1}H), 150.92 MHz (^{13}C)) and Alantrypinone (**4**)[4] in DMSO-d_6

	alantrypinone (**4**)	lapatin B (**2**)				
position	^{13}C	^{13}C	^{1}H	J = Hz	NOE	HMBC to carbon #[a]
1	169.5	169.3				
CONH			9.629			
2	61.9	58.8	4.337	<1.0		1,3,12,13,22
3	152.7	151.0				
4	146.9	146.9				
5	127.2	127.1	7.677	7.9		7,9
6	134.0	134.5	7.875	7.9		4,8
7	126.5	126.9	7.606	7.9		5,9
8	126.0	126.2	8.220	8.0		4,6,10
9	120.2	120.1				
10	158.4	158.1				
11	51.8	52.4	5.557			1,3,10,13
12a	35.9	34.5	2.413	14.7, 3.1	21	1,2,11,13,14,22
12b			2.386	14.7, 2.5		1,2,11,13,14, 22
13	54.7	51.0				
14	176.6	176.4				
CONH			10.688	<1.0		
15	142.4	141.7				
16	109.5	109.7	6.944	7.7		20,22
17	128.8	128.9	7.314	7.5		17,21
18	122.0	121.8	7.091	7.5		18,22
19	123.6	123.8	7.206	7.2	12a(s)	13,17,19
20	129.8	130.5				
21	12.8					

[a] The HMBC data are reported as from ^{1}H to ^{13}C.

(Table 1), establishing the relative configuration around the spirocenter at C-13. The CD data established the absolute configuration of **1** since a comparison with the CD data of alantrypinone showed the two compounds to have practically opposite Cotton effects. Since **1** and **4** contain only two chromophore systems with conjugated double bonds, we therefore conclude that these two systems have to be oppositely positioned in the two compounds. Thus the tryptophan aromatic part of **1** has to be on the opposite side of the quinazoline ring system when compared to **4**, for which the structure was established by X-ray analysis.

The ^{13}C NMR data of **2** were similar to those of the target alantrypinone (**4**) (Table 2), in very good agreement with the UV data, also showing the spectroscopic features of the two compounds to be the most similar (n1, Figure 2). Twenty carbon signals almost overlapped for the two compounds, which differed by only one carbon (Table 2). This was confirmed by the MS data, which gave the structural formula $C_{20}H_{15}O_3N_4$ $(M + H)^+$, also indicating compound **2** to weigh 14 mass units less than **4**. Careful investigation of the NMR data revealed that **2** differs from **4** only at the C-2 position by having one methyl group less. Thus instead of a quaternary carbon at C-2 as in alantrypinone (**4**), the HSQC experiment of **2** showed the attach-

ment of a proton resonating at 4.34 ppm at the C-2 position (58.8 ppm) having a small coupling (ca. 1 Hz) to the amide proton at 9.18 ppm. This was confirmed by HMBC correlation from H-2 to C-1, C-3, C-12, C-13, and C-20. Thus lapatin B likely incorporates glycine instead of alanine in the formation of the condensed spiroquinazoline molecule. Since the CD data for **2** show the same Cotton effects as for **1**, we conclude that the two compounds have similar absolute configurations.

In conclusion the present work has proven the concept of the new algorithm *X-hitting*[17] for targetting new natural products with structural features very similar to already known compounds. The close structural relationship between the lapatins and other spiroquinazolines indicates the lapatins to have anti-GABAergic effects toward insects as shown for alantrypinone and serantrypinone;[3] however this still has to be demonstrated. Work is in progress in our laboratory to prove the scope of the *X-hitting* algorithm for targeting other types of biologically active alkaloids such as cyclic peptides, benzodiazepines, and statin polyketides.

Experimental Section

General Experimental Procedures. NMR spectra of the lapatins were recorded in 5 mm tubes at 600.13 MHz for ^{1}H and at 150.92 MHz for ^{13}C and at 300 K, using DMSO-d_6, on a Bruker DRX 600 as previously described.[18] The chemical shifts are given relative to DMSO (2.50 ppm for H^1 and 39.5 ppm for ^{13}C) (Tables 1 and 2). The circular dichroism (CD) spectra were measured on a JASCO J-710 spectropolarimeter and the UV spectra on a Hewlett-Packard 8452A diode array spectrophotometer. Analytical HPLC and mass spectrometric conditions were similar to those given by Nielsen and Smedsgaard,[19] and the retention index (RI) of the lapatins was calculated according to Frisvad and Thrane.[20]

Fungal Material and Fermentation. The *Penicillium lapatayae* isolate (IBT 10870) used for isolation of the lapatins was obtained from the IBT Culture Collection at the BioCentrum-DTU, Technical University of Denmark. The fungus was cultured for 14 days on 200 Petri dishes with ALK medium. *Penicillium* and *Aspergillus* strains used in the screening are described in Hansen et al.[17]

Extraction and Separation. The combined fungal mycelia were extracted twice with a total of 2.5 L of EtOAc to give approximately 3.0 g of crude extract. This extract was subjected to vacuum-liquid chromatography on a Waters C18 SPE column (10 g) using six times 100 mL of H_2O/MeOH to give the fractions A (90:10), B (75:25), C (50:50), D (25:75), E (0:100), and F (0:100 with 0.1% CF_3CO_2H). The C fraction (179 mg) enriched with lapatins A and B was subsequently separated on a Waters Prep Nova-Pak C-18 cartridge (25 × 100 mm, 6 μm, 60 Å) using 20 mL/min H_2O/MeCN (20:80) as mobile phase to give lapatin A (**1**) (5.3 mg) and a fraction enriched with lapatin B (5.7 mg). Finally this fraction was subjected to HPLC on a Waters Symmetry C-18 column (8 × 30 mm, 7 μm) using 6 mL/min H_2O/MeCN (20:80) as mobile phase to give lapatin B (**2**) (2.5 mg).

Lapatin A (1): [α] $^{22}_D$ 22° (*c* 1.1 EtOH); UV λ_{max} (EtOH) nm (log ϵ) 206 (4.43), 237 (4.05), 255 (3.96), 257 (3.97), 306 (3.50); CD (EtOH, *c* 0.011), $\Delta\epsilon$ (λ nm) 200 (+15.94), 219 (−13.66), 238 (+27.32), 247 (−7.11), 261 (+7.68), 289 (−1.94), 311 (+2.39); ^{1}H NMR, Table 1; HRESMS obsd $(M + H)^+$ at *m/z* 414.1658, calcd for $C_{23}H_{19}O_3N_5$ 414.1644; RI = 738.

Lapatin B (2): [α] $^{22}_D$ −20° (*c* 1.6, EtOH); UV λ_{max} (EtOH) nm (log ϵ) 255 (3.36), 265 (3.38), 277 (3.31), 302 (3.02), 314 (2.90); CD (EtOH, *c* 0.011), $\Delta\epsilon$ (λ nm) 213 (−8.14), 235 (+3.73), 247 (−1.66), 276 (+1.10), 296 (−1.01), 318 (+0.28); ^{1}H NMR, Table 1; HRESMS obsd $(M + H)^+$ at *m/z* 359.1150, calcd for $C_{23}H_{19}O_3N_5$ 359.1144; RI = 800.

Acknowledgment. We are indepted to Dr. S. E. Harnung for determination of the CD data measured on a modified JASCO instrument financed by the Danish Natural Science Research Council, Grant No. //-0373-1. The support of The Danish Research Councils (Grant No. 9901295) and the Center for Advanced Food Studies is acknowledged.

References and Notes

(1) Newman, D. J.; Cragg, G. M.; Snader, K. M. *J. Nat. Prod.* **2003**, *66*, 1022−1037.
(2) Barrow, C. J.; Sun, H. H. *J. Nat. Prod.* **1994**, *57*, 471−476.
(3) Kuriyama, T.; Kakemoto, E.; Takahashi, N.; Imamura, K.-I.; Oyama, K.; Suzuki, E.; Hariyama, K.; Yaguchi, T.; Ozoe, Y. *J. Agric. Food Chem.* **2004**, *52*, 3884−3887.
(4) Larsen, T. O.; Frydenvang, K.; Frisvad, J. C.; Christophersen, C. *J. Nat. Prod.* **1998**, *61*, 1154−1157.
(5) Ariza, M. R.; Larsen, T. O.; Petersen, B. O.; Duus, J.Ø.; Christophersen, C.; Barrero, A. F. *J. Nat.* Prod. **2001**, *64*, 1590−1592.
(6) Fiedler, H. P. *J. Chromatogr.* **1984**, *316*, 487−494.
(7) Fiedler, H. P. *Nat. Prod Lett.* **1993**, *2*, 119−128.
(8) VanMiddelsworth, F.; Cannell, R. C. P. In *Methods in Biotechnology, Vol. 4: Natural Products Isolation*; Cannel, R. J. P., Ed.; Humana Press Inc.: Totowa, NJ, 1998.
(9) Larsen, T. O.; Frydenvang, K.; Frisvad, J. C. *Biochem. Syst. Ecol.* **2000**, *28*, 881−886.
(10) Larsen, T. O.; Franzyk, H.; Jensen, S. R. *J. Nat. Prod.* **1999**, *62*, 1578−1580.
(11) Wong, S.-M.; Musza, L. L.; Kydd, G. C.; Kulling, R.; Gillum, A.; Cooper, R. *J. Antibiot.* **1993**, *46*, 545−553.
(12) Takahashi, C.; Matsushita, T.; Doi, M.; Minoura, K.; Shingu, T.; Kumeda, Y.; Numata, A. *J. Chem. Soc., Perkin Trans. 1* **1995**, 2345−2353.
(13) Penn, J.; Mantle, P. G.; Bilton, J. N.; Sheppard, R. N. *J. Chem. Soc., Perkin Trans. 1* **1992**, 1495−1496.
(14) SciFinder Scholar. www.cas.org/SCIFINDER/SCHOLAR/index.html.
(15) AntiBase.www.wileyeurope.com/WileyCDA/WileyTitle/productCd-0471468401.html.
(16) University of Canterbury Marinlit from the Department of Chemistry. www.chem.canterbury.ac.nz/research/marinlit.htm.
(17) Hansen, M. E.; Smedsgaard, J.; Larsen, T. O. *X-hitting*: An algorithm for de-replication and novelty detection of UV spectra in complex mixtures of natural products. *Anal. Chem.* **2005** (submitted).
(18) Larsen, T. O.; Petersen, B. O.; Duus, J. Ø. Lumpidin, a novel biomarker of some ochratoxin A producing Penicillia. *J. Agric. Food Chem.* **2001**, *49*, 5081−5084.
(19) Nielsen, K. F.; Smedsgaard, J. *J. Chromatogr. A* **2003**, *1002*, 111−136.
(20) Frisvad, J. C.; Thrane, U. *J. Chromatogr.* **1987**, *404*, 195−214.

NP040248S