ORGANIC LETTERS

2000
Vol. 2, No. 3
289−292

New Diffusion-Edited NMR Experiments To Expedite the Dereplication of Known Compounds from Natural Product Mixtures

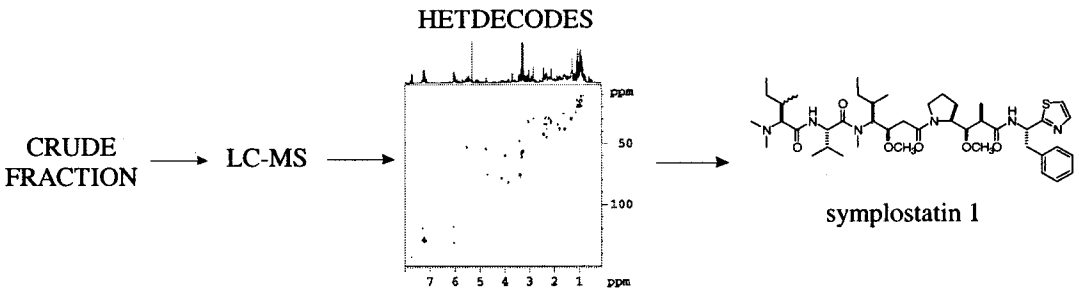
R. Thomas Williamson,[†] Eleanor L. Chapin,[‡] Andy W. Carr,[‡] Jeffrey R. Gilbert,[‡] Paul R. Graupner,[‡] Paul Lewer,[‡] Pat McKamey,[‡] John R. Carney,[§] and William H. Gerwick*,[†]

College of Pharmacy, Oregon State University, Corvallis, Oregon 97331, Dow AgroSciences, Indianapolis, Indiana 46268, and Kosan Biosciences, Inc., 3832 Bay Center Place, Hayward, California 94545

bill.gerwick@orst.edu

Received November 9, 1999

ABSTRACT



We present two new diffusion-edited NMR experiments, improved DECODES and HETDECODES, that sort the constituents in a mixture by their individual diffusion coefficients. These experiments should allow the partial NMR spectral assignment and cursory structure elucidation of compounds in a complex mixture as an aid in the dereplication of known or nuisance compounds.

Natural products continue to provide a rich diversity of useful compounds. In our laboratories, the focus of this search has been on the isolation of anticancer and insecticidal compounds from marine cyanobacteria.[1]

One of the major problems encountered in natural product drug discovery programs is the dereplication of known compounds. The ability to identify known or nuisance compounds at an early stage of fractionation represents huge savings in both time and effort. Methods currently used for dereplication include chromatography-based approaches such as LC MS, LC NMR, and LC NMR-MS. Herein we present a complementary approach that takes advantage of the variance in translational diffusion of organic compounds as a function of their molecular size.

During the course of our collaborative search for compounds with insecticidal activity, a mixed assemblage of the cyanobacterium *Lyngbya*/*Schizothrix* sp. from the Fiji Islands yielded an extract that was potently active against tobacco bud worm (*Heliothis virescens*).

Preliminary bioassay-guided fractionation of the extract led to concentration of insecticidal activity in two fractions, A and B, after initial chromatography over LH-20 with methanol as the eluting solvent.[2] Fraction A was further fractionated by ODS Bond Elut and ODS-HPLC, yielding a

[†] Oregon State University.
[‡] Dow AgroSciences.
[§] Kosan Biosciences, Inc.

(1) (a) Sitachitta, N.; Gerwick, W. H. *J. Nat. Prod.* **1998**, *61*, 681−684. (b) Hooper, G. J.; Orjala, J.; Schatzman, R. C.; Gerwick, W. H. *J. Nat. Prod.* **1998**, *61*, 529−533. (c) Shimizu, Y. *Chem. Rev.* **1993**, *93*, 1685−1698.

10.1021/ol991239r CCC: $19.00
Published on Web 01/12/2000

single active compound.[2] This compound was determined to be a highly *N*-methylated cyclic depsipeptide containing 11 aliphatic amino acid residues. The structure of this potently cytotoxic and insecticidal compound will be the focus of a separate communication. On the other hand, chromatography of fraction B was not so straightforward.

Analysis of fraction B by low-resolution LC-APCI-MS (APCI = atmospheric pressure chemical ionization) showed it to contain a series of compounds with molecular weights of 785, 845, and 799. A routine molecular weight search, utilizing the MarineLit database, hinted that these compounds could be related to the dolastatin class of cytotoxic peptides.[3] Further fractionation by ODS Bond-Elut SPE cartridges and ODS-HPLC distributed the activity to a number of fractions with fraction B2B being the most potent.[2]

To confirm whether the major compound in fraction B2B was in fact a known compound (based on a molecular weight of 798.5), a diffusion-edited NMR approach was taken. This approach was employed because of the small amount of the fraction available (16.5 mg containing many components) and the documented unfavorable chromatographic properties of dolastatin 10 and the closely related compound symplostatin 1.[3]

The first experiment used to probe the ^{1}H NMR spectral features of the mixture was based on the previously reported diffusion-encoded spectroscopy (DECODES) experiment developed by Shapiro and co-workers.[5] This approach was originally designed for the analysis of relatively well defined mixtures from combinatorial synthesis libraries. The DECODES experiment takes advantage of the differing diffusion coefficients of unlike compounds in a mixture to "tag" the various spin systems in a ^{1}H−^{1}H TOCSY spectrum. By grouping all of the spin systems with the same diffusion coefficient as being part of the same molecule, the proton spectra of a given molecule can be assigned. This approach has the advantage over the HR-DOSY experiments in that it can be applied to mixtures with highly overlapped spin systems.[6,7] The only requirement for resolving a spin system is that one correlation in the TOCSY spectrum be resolved. However, one disadvantage of the DECODES experiment is that it relies on conventional phase cycling for the suppression of T_1 noise which requires that a minimum number of eight scans be completed for each increment in the 2D spectrum. In addition, imperfect phase cycling can often introduce spectral artifacts and phase distortions. This is especially true if there is a large dynamic range present in the sample as would normally be the case in the analysis of crude fractions from a natural product isolation scheme. To minimize these problems so that the DECODES spectrum could be applied to even more complex mixtures, we have introduced several improvements over the original experiment (Figure 1a).

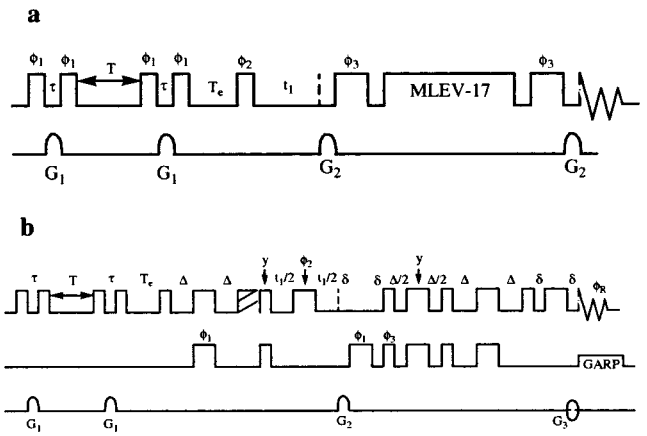


Figure 1. (a) Pulse sequence for the improved DECODES experiment. A diffusion time of 100 ms, an eddy current delay of 50 ms, and a homonuclear isotropic mixing period of 60 ms were used. Phase cycling was carried out as follows: $\phi_1 = x, -x, -x, x, y, -y, -y, y$; $\phi_2 = x, -x, -x, x, y, -y -y, y$; $\phi_3 = x, -x, x, -x, y, -y, y, -y$; $\phi_R = x, -x, -x, x, y, -y, -y, y$. For phase sensitive detection, ϕ_2 is incremented according to the TPPI method.[4] (b) Pulse sequence for the HETDECODES experiment. A diffusion time of 50 ms and an eddy current delay of 50 ms were used. Phase cycling was carried out as follows: $\phi_1 = x, x, -x, -x$; $\phi_2 = x, x, -x, -x$; $\phi_3 = y, y, -y, -y$; $\phi_R = x, -x, -x, x$. For phase sensitive detection, ϕ_3 is incremented according to the TPPI method and G_3 is inverted for each successive FID according to the echo−antiecho protocol.[4] For both pulse sequences, narrow bars represent 90° pulses, wide bars represent 180° pulses, and hash-marked bars represent high power spin lock pulses of 2 ms duration. If no phase cycling is indicated, pulses were applied along the *x* axis.

The first modification is the application of sinusoidal shaped gradients during the longitudinal eddy current diffusion (LED) portion of the sequence. It was recently shown that sinusoidal shaped gradients provide a marked improve-

(2) Collection WG1170 was an inseparable mixture of *Lyngbya* sp. and *Schizothrix* sp. that was collected from the Fiji Islands in February 1997. The cyanobacteria were extracted with CH_2Cl_2−MeOH to afford 1.0 g organic extract. This extract was fractionated over LH-20 with MeOH to afford 11 × 50 mL fractions (A−K). Fraction A was further purified by C_{18} flash chromatography (YMC-ODS-A) using a gradient of 60−100% MeOH. The fraction eluting with 90% MeOH was concentrated and subjected to ODS-HPLC (YMC ODS-AQ 250 × 10 mm column) using a solvent system of 85% MeOH−H_2O to provide two fractions. Fraction 2 contained a new cytotoxic and insecticidal *N*-methylated cyclic undecadepsipeptide. Fraction B from the LH-20 fractionation was further fractionated on a 12 cm^3 C_{18} Mega-Bond Elut SPE cartridge with a gradient consisting of 80−100% MeOH−H_2O. Fraction B2 was further fractionated by an additional 12 cm^3 C_{18} Mega-Bond Elut ODS column with 80, 90, and 100% MeOH−H_2O to provide fraction B2B, the fraction used for the diffusion-edited experiments presented in this Letter.

(3) (a) Pettit, G. R.; Kamano, Y.; Herald, C. H.; Tuinman, A. A.; Boettner, F. E.; Kizu, H.; Schmidt, J. M.; Baczynskyj, L.; Tomer, K. B.; Bontems, R. J. *J. Am. Chem. Soc*. **1987**, *109*, 6883−6885. (b) Haarigan, G. C.; Luesch, H.; Yoshida, W. Y.; Moore, R. E.; Nagle, D. G.; Paul, V. J.; Mooberry, S. L.; Corbett, T. H.; Valeriote, F. A. *J. Nat. Prod*. **1998**, *61*, 1075−1077.

(4) (a) Davis, A. L.; Keeler, J.; Moskau, J. *J. Magn. Reson.* **1992**, *98*, 207−216. (b) Boyd, J.; Soffe, N.; John, B. K.; Plant, D.; Hurd, R. *J. Magn. Reson*. **1992**, *98*, 660−664.

(5) Lin, M.; Shapiro, M. *J. Org. Chem*. **1996**, *61*, 7617−7619.

(6) Four excellent reviews of DOSY NMR have been published: (a) Gounarides, J. S.; Chen, A.; Shapiro *J. Chromatogr*. **1999**, *725*, 79−90. (b) Morris, G. A.; Barjat, H. In *Methods for Structure Elucidation by High-Resolution NMR*; Elsevier Science: New York, 1997; pp 209−226. (c) Johnson, C. S., Jr. *Prog. Nucl. Magn. Reson*. **1999**, *34*, 203−256. (d) Pelta, M. D.; Barjat, H.; Morris, G. A.; Davis, A. L.; Hammond, S. J. *Magn. Reson. Chem.* **1998**, *36*, 706−714.

(7) An alternative approach for measuring diffusion coefficients utilizing Accordion Spectroscopy has been published: Millet, O.; Pons, P. *J. Magn. Reson.* **1998**, *131*, 166−169.

ment in reproducibility and reduced eddy current effects when compared to the customary use of rectangular gradients for diffusion-based experiments.[8] The second modification is the inclusion of a gradient selected, pure absorption, phase sensitive TOCSY experiment for the homonuclear isotropic mixing period.[9] As can be seen in Figure 2, this new

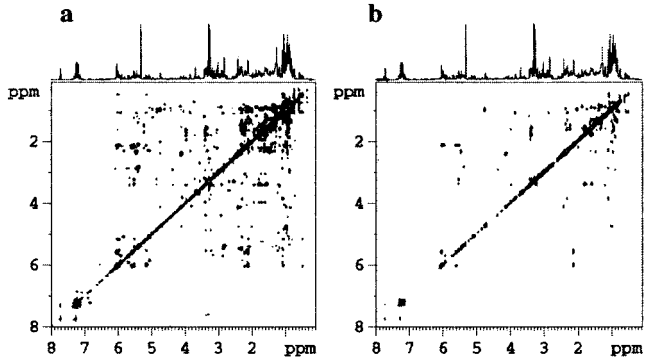


Figure 2. Spectra from improved DECODES experiments acquired with gradient strengths of (a) 15% and (b) 45% of maximum. Note the simplified spectra present in (b**)** due to suppression of faster diffusing components.

sequence provides exceptional phase characteristics and artifact suppression. Since the majority of biologically active compounds of interest from natural product extracts are over 400 amu, a diffusion time and gradient strength can often be chosen so that most of the lower molecular weight impurities are suppressed. If the impurities are of similar molecular size to that of the compound or compounds of interest, a series of DECODES experiments can be acquired and from them an accurate relative diffusion coefficient can be measured (Figure 3). Once the diffusion coefficients for

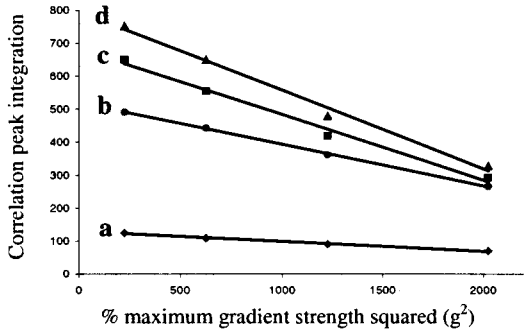


Figure 3. Line (a**)** represents relative rate of diffusion for symplostatin 1. Lines (b), (c), and (d) represent relative rates of diffusion for some of the major impurities in fraction B2B. R^2 values for all slopes were >0.99.

the different spin systems are known, it is a simple task to identify those correlations that belong to the compound of interest. With the spins systems identified, the ^{1}H chemical shift values can then be compared with literature values to make a confident identification of the known compound. If the molecule of interest is unknown, a great deal of insight can be gleaned as to its structure class based on ^{1}H NMR chemical shifts and partial structures derived from the TOCSY spectrum. In turn, these data can provide valuable information for subsequent chromatographic protocols.

If the ^{1}H NMR shifts are somewhat ambiguous, or if more information is required for positive identification of the compound, diffusion-edited heteronuclear shift correlation experiments can be used. The diffusion-edited PEP-HSQC[10] (preservation of equivalent pathways heteronuclear single quantum coherence) experiment begins with an LED preparation period similar to that used for the homonuclear DECODES (Figure 1b). The LED preparation period attenuates the one-bond heteronuclear correlations from the well-known PEP-HSQC experiment as a function of the diffusion coefficients for the differing constituents in a mixture. A diffusion coefficient can be easily calculated for each correlation in the spectrum by varying the diffusion times or gradient strengths in a number of experiments (Figure 4). As in the homonuclear DECODES experiment,

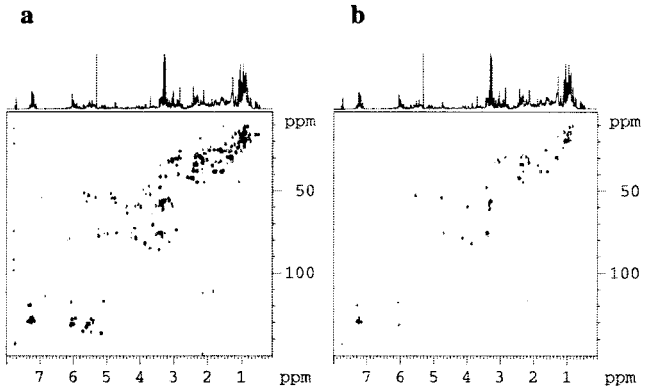


Figure 4. Spectra from HETDECODES experiments taken with gradient strengths of (a) 15% and (b) 65% of maximum. Note the simplified spectra present in (b) due to suppression of faster diffusing components.

identical diffusion coefficients can be related as being part of the same molecule and HSQC correlations for each compound of interest can be assigned. This experiment, designated the heteronuclear DECODES or HETDECODES pulse sequence, allows the assignment of a ^{13}C NMR chemical shift for each protonated carbon in the molecule.

(8) Price, W. S.; Hayamizu, K.; Ide, H.; Arata, Y. *J. Magn. Reson.* **1999**, *139*, 205−212.
(9) Parella, T.; Sanchez-Ferrando, F.; Virgil, A. *J. Magn. Reson.* **1997**, *125*, 145−148.
(10) (a) Kay, L. E.; Keifer, P.; Saarinen, T. *J. Am. Chem. Soc.* **1992**, *114*, 10663−10665. (b) Palmer, A. G., III; Cavanagh, J.; Wright, P. E.; Rance, M. *J. Magn. Reson.* **1991**, *93*, 151−170. (c) Kontaxis, G.; Stonehouse, J.; Laue, E. D.; Keeler, J. *J. Magn. Reson. Ser. A* **1994**, *111*, 70−76. (d) Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletsky, O.; Glaser, S. J.; Sorenson, O. W.; Griesenger, C. *J. Biomol. NMR* **1994**, *4*, 301−306.

This experiment provides a much more sensitive and time saving alternative to the previously published 3D DOSY-HMQC experiment.[11] As for the homonuclear DECODES experiment, these chemical shifts can be used to positively assign a known compound, or to identify partial structures in an unknown molecule to guide its subsequent purification.

High resolution LC-QTOF-MS-MS (QTOF = quadrupole time of flight) confirmed that the compound identified from the crude fraction B2B was symplostatin 1 on the basis of a comparison of high-resolution fragmentation patterns with those reported for the closely related dolastatin 10.[3] To compare the quality of the NMR data obtained with these new experiments to data acquired for the pure compound, the crude fraction containing symplostatin 1 was subjected to two additional tiers of fractionation. The sample was applied to a Bond Elut Phenyl column, washed with CH_2Cl_2, and eluted with MeOH. The MeOH eluent was applied to a Si Bond Elut cartridge and eluted with 5% MeOH−EtOAc, yielding pure symplostatin 1. As can be seen in Figure 5, the TOCSY and PEP-HSQC for pure symplostatin 1 appear highly comparable to those shown in Figures 2 and 4 for the diffusion-edited experiments of impure fraction B2B (symplostatin 1 is ∼20% of the fraction).

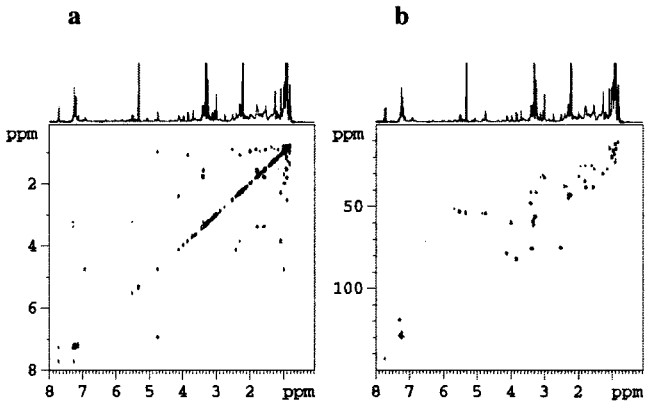


Figure 5. (a) Pure absorption $^1H-^1H$ TOCSY and (b) PEP-HSQC spectra for pure symplostatin 1 after two additional tiers of fractionation. These spectra are comparable to those shown in Figures 2a and 4a.

These experiments add new tools to the "bag of tricks" available for dereplication of known or nuisance compounds in natural product drug discovery programs. Although not intended to replace LC-based MS methods, they should provide a very useful compliment to LC-MS in dereplication efforts. The time savings provided by these experiments will allow more efficient recognition of known compounds in biologically active fractions and should therefore increase the throughput of these endeavors. In addition, the HET-DECODES sequence presented herein allows for accurate diffusion measurements to be made for molecules with very highly overlapped ^1H NMR spectra. Because correlations are sorted by ^{13}C NMR chemical shift, up to a 15-fold increase in spectral dispersion compared with those of homonuclear TOCSY or COSY experiments is realized. Not only should these experiments find applications in natural product chemistry but they should also prove useful in combinatorial synthesis and other branches of organic chemistry.

Acknowledgment. We acknowledge NIH Grant CA52955 and Dow AgroSciences for financial support, Melodie Graber and Greg Hooper for making these algal collections, the Fijian government for permission to make these algal collections, and Victor Hsu for helpful discussions. The Bruker DRX 600 spectrometer was purchased in part through grants from the National Science Foundation (BIR-9413692) and the W. M. Keck Foundation.

Supporting Information Available: Example data from both the improved DECODES and the HETDECODES experiments and annotated Bruker pulse programs. This material is available free of charge via the Internet at http://pubs.acs.org and from the authors at http://web.orst.edu/∼williaro/pulseprograms.html.

OL991239R

(11) Barjat, H.; Morris, G. A.; Swanson, A. G. *J. Magn. Reson.* **1998**, *131*, 131−138.