# Structural sequencing of oligopeptides aided by 1H iterative full-spin analysis

Gao, Wei; Mcalpine, James B.; Choules, Mary P.; Napolitano, José G.; Lankin, David C.; Simmler, Charlotte; Ho, Ngoc Anh; Lee, Hanki; Suh, Joo-Won; Burton, Ian W.; Cho, Sanghyun; Franzblau, Scott G.; Chen, Shao-Nong; Pauli, Guido F.



**Publisher's version / Version de l'éditeur:**
https://doi.org/10.1021/acs.jnatprod.7b00207
*Journal of Natural Products, 80, 10, pp. 2630-2643, 2017-10-16*

JOURNAL OF NATURAL PRODUCTS

Article

Cite This: J. Nat. Prod. 2017, 80, 2630-2643

pubs.acs.org/jnp

# Structural Sequencing of Oligopeptides Aided by $^1$H Iterative Full-Spin Analysis

Wei Gao,[†,‡] James B. McAlpine,[†,‡] Mary P. Choules,[†,‡] José G. Napolitano,[†,§] David C. Lankin,[†,§] Charlotte Simmler,[†,§] Ngoc Anh Ho,[⊥] Hanki Lee,[⊥] Joo-Won Suh,[⊥,∥] Ian W. Burton,[▽] Sanghyun Cho,[‡] Scott G. Franzblau,[‡] Shao-Nong Chen,[†,‡,§] and Guido F. Pauli*[,†,‡,§]

[†]Department of Medicinal Chemistry and Pharmacognosy, College of Pharmacy, University of Illinois at Chicago, Chicago, Illinois 60612, United States

[‡]Institute for Tuberculosis Research, College of Pharmacy, University of Illinois at Chicago, Chicago, Illinois 60612, United States

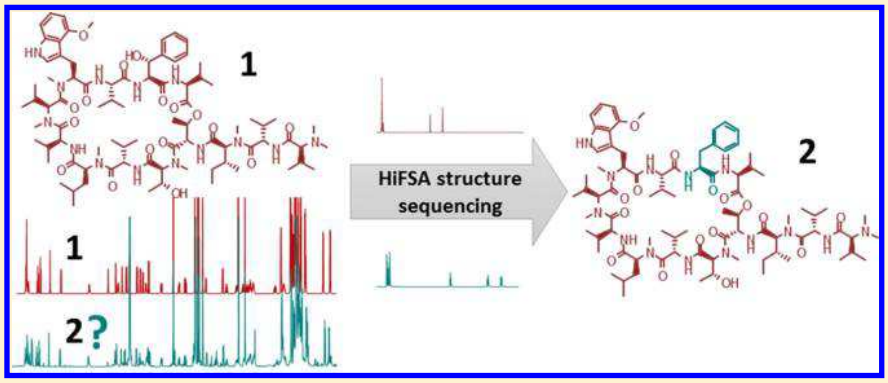
[§]Center for Natural Product Technologies, College of Pharmacy, University of Illinois at Chicago, Chicago, Illinois 60612, United States

[⊥]Center for Nutraceutical and Pharmaceutical Materials, Myongji University, Yongin, Gyeonggi-do 17458, South Korea

[∥]Division of Biosciences and Bioinformatics, Myongji University, Yongin, Gyeonggi-do 17458, South Korea

[▽]Institute for Aquatic and Crop Resource Development, National Research Council, Halifax, Nova Scotia B3H 3Z1, Canada

*Supporting Information*




**ABSTRACT:** This report describes an approach using $^1$H NMR iterative full-spin analysis (HiFSA) to extract definitive structural information on unknown peptides from 1D $^1$H NMR data. By comparing the experimental data and HiFSA fingerprint of a known analogue, it is possible to isolate the characteristic $^1$H subspectrum of the different amino acids and, thus, elucidate the structure of the peptide. To illustrate this methodology, a comprehensive analysis of five new anti-*Mycobacterium tuberculosis* peptides (**2**−**6**), all analogues of ecumicin (**1**), was carried out. The method was validated by demonstrating congruence of the HiFSA-based structures with all available data, including MS and 2D NMR. The highly reproducible HiFSA fingerprints of the new ∼1600 amu peptides were generated in this process. Besides oligo-peptides, the HiFSA sequencing approach could be extended to all oligomeric compounds consisting of chains of monomers lacking H−H spin−spin coupling across the moieties. HiFSA sequencing is capable of differentiating complex oligomers that exhibit minor structural differences such as shifted hydoxyl or methyl groups. Because it employs the basic and most sensitive 1D $^1$H NMR experiment, HiFSA sequencing enables the exploration of peptide analogues up to at least 2000 amu, even with basic contemporary spectrometers and when only sub-milligram amounts of isolates are available.

Peptides are a class of biologically active molecules and play many essential physiological roles. As natural peptides have evolved to bind certain targets specifically, they are generally recognized as highly selective and efficacious and, therefore, are attractive starting points for new drug design. More than 60 FDA approved peptide drugs are marketed and hundreds are under development.[1] "Small" cyclic peptides appear to be particularly interesting, as their cyclic backbone may protect against protease degradation and eventually lead to oral bioavailability, as seen in the cases of the orally delivered cyclic peptide drugs cyclosporin A, desmopressin, and linaclotide.[1,2]

However, the structure elucidation of cyclic peptides is generally challenging, because LC-MS-based proteomic approaches are not very effective without proper prior linearization. Current NMR-based methods derive mostly

**Received:** March 9, 2017
**Published:** October 16, 2017





DOI: 10.1021/acs.jnatprod.7b00207
J. Nat. Prod. 2017, 80, 2630−2643

from the concept of sequential assignment developed by Wüthrich and co-workers in the early 1980s.[3] Briefly, as each amino acid contains an individual $^1$H spin system isolated by amide groups ("spin cages"), the amino acid side-chain spin-systems are first identified by, for example, 2D homonuclear COSY and TOCSY and/or heteronuclear experiments, such as HSQC−COSY and HSQC−TOCSY.[4−7] Subsequently, the neighboring residues are identified via NOEs or long-range couplings, using experiments such as 2D-NOESY or 2D-HMBC, respectively. In addition to $^1$H−$^{13}$C experiments, a number of $^1$H−$^{15}$N experiments, including HSQC, HMBC, HSQMBC-TOCSY, CIGAR-HMBC, and LR-HSQMBC, have been reported as being useful tools.[7−10] However, compared to 1D $^1$H NMR, greater quantities of sample are usually necessary for recording high-quality 2D NMR spectra, but sufficient isolation yields are not always achievable. On the other hand, $^1$H NMR-based analyses are able to accommodate even sub-milligram samples, rendering a wealth of structural information encoded in rather complex coupling patterns of the 1D $^1$H NMR spectra.[11] However, this is seldom used in the early stage of structural analysis.

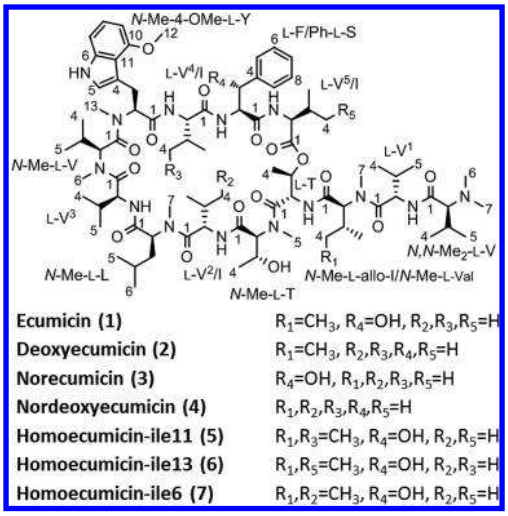



**Figure 1.** Structures of ecumicin and analogues.

The present report describes a new structural sequencing methodology for the structure elucidation of cyclic oligopeptides of ca. 1600 amu and subsequent development of their digital $^1$H NMR profiles, mainly arising from interpretation of their 1D $^1$H NMR spectra. This approach enables structure elucidation with even sub-milligram samples using standard RT probes on entry-level cryomagnets and even lower sample quantities when using cryoprobe NMR instrumentation. The method relies on the $^1$H iterative full-spin analysis (HiFSA),[12] an iterative quantum mechanical-based postacquisition processing approach for deriving digital, all-inclusive $^1$H NMR profiles that reproduce the experimental $^1$H NMR spectra of organic molecules in their entirety.[13] As demonstrated recently in the study of steviol glycosides,[14] complex $^1$H NMR profiles can potentially be taken apart to yield the HiFSA fingerprints of individual sugar moieties of these bisdesmosides. By analogy, as each amino acid in a given peptide sequence has a distinctive $^1$H NMR spin pattern, a $^1$H NMR subspectrum could provide sufficient information for structural identification.

The applicability and effectiveness of this method was assessed by the structure elucidation of six analogues (**2**−**7**) of ecumicin (**1**).[15,16] Ecumicin is a 1599 amu macrocyclic tridecapeptide, produced by an actinomycete strain, *Nonomuraea* sp. MJM5123,[15] that exerts potent bactericidal activity by targeting the hexameric chaperon protein ClpC1 of *Mycobacterium tuberculosis* (*M. tb*). This constitutes a new anti-TB drug target.[16] The HiFSA profile of ecumicin[17] was used as the foundation to the generation of the HiFSA fingerprints of individual amino acids and the comprehensive analysis of the $^1$H NMR spectra of **2**−**7**. The yields of the seven peptide congeners (**1**−**7**) varied significantly in quantities ranging from picograms to tens of milligrams per liter of the fermentation at harvest. Thus, structure elucidation of **2**−**7** has been very challenging due to the limited quantities of some compounds and the fact that these analogues are very similar but also relatively large molecules (all around 1600 amu), differing only in the presence of a hydroxy or a methylene group or, in some cases, the location of a particular methylene group.

## ■ RESULTS AND DISCUSSION

**Isolation.** Deoxyecumicin (**2**), norecumicin (**3**), and nordeoxyecumicin (**4**) were isolated from the methanolic mycelial extract of *Nonomuraea* sp. MJM5123 by a three-step, bioassay-guided fractionation scheme as reported previously.[15] Samples of **3** and **4** were further purified by RP-HPLC. The final samples of **2**, **3**, and **4** were pale yellow powders with purities of 60.1%, 75.6%, and 77.4% as determined by the 100% qNMR method,[18] respectively. All isolates gave rise to characteristic peptide/protein UV spectra, with absorptions corresponding to the peptide bonds and the aromatic amino acids observed (see Supporting Information). High-resolution MS analyses revealed that they have exact masses of 1583.0055, 1584.9980, and 1568.9934 amu, respectively. Whereas the large molecular weights precluded the initial inference of their molecular formulas, these values were approximately 16, 14, and 30 amu, respectively, less than the exact mass of **1** and thus suggested the absence of an oxygen atom in **2**, of a methylene group in **3**, and of both an oxygen atom and a methylene group in **4**, when compared to **1**. The corresponding molecular formulas derived for **2**, **3**, and **4** were $C_{83}H_{134}N_{14}O_{16}$, $C_{82}H_{132}N_{14}O_{17}$, and $C_{82}H_{132}N_{14}O_{16}$, respectively.

Three homoecumicins [-ile11 (**5**), -ile13 (**6**), and -ile6 (**7**)] were similarly isolated as pale yellow powders, with 100% qNMR purities of 41.9%, 38.4%, and 25.9%. The three peptides also showed the characteristic peptide UV and IR spectra corresponding to the amide groups (see Supporting Information). They exhibited an identical exact mass of 1613.0140 amu, which is approximately 14 amu more than the mass of **1**, suggesting the presence of an additional methylene group in each of **5**, **6**, and **7** when compared to **1** and indicated a common molecular formula of $C_{84}H_{136}N_{14}O_{17}$.

The close similarity in the structures makes the isolation process a highly challenging one. With a carefully selected solvent system, high-speed counter current chromatography (HSCCC) could successfully separate the analogues of **1**−**4** from an enriched fraction, but the separation of the homoecumicin isomers (**5**−**7**) has been very difficult even with semipreparative HPLC. Due to overlapped HPLC peaks and the low natural abundance of these congeners, the homoecumicin isomers were concentrated as relatively impure samples at sub-milligram quantities. The isolation yields were 756, 483, 55, 3.6, 1.9, and 3.4 $\mu$g per liter fermentation culture for **2**, **3**, **4**, **5**, **6**, and **7**, respectively.


DOI: 10.1021/acs.jnatprod.7b00207
J. Nat. Prod. 2017, 80, 2630−2643

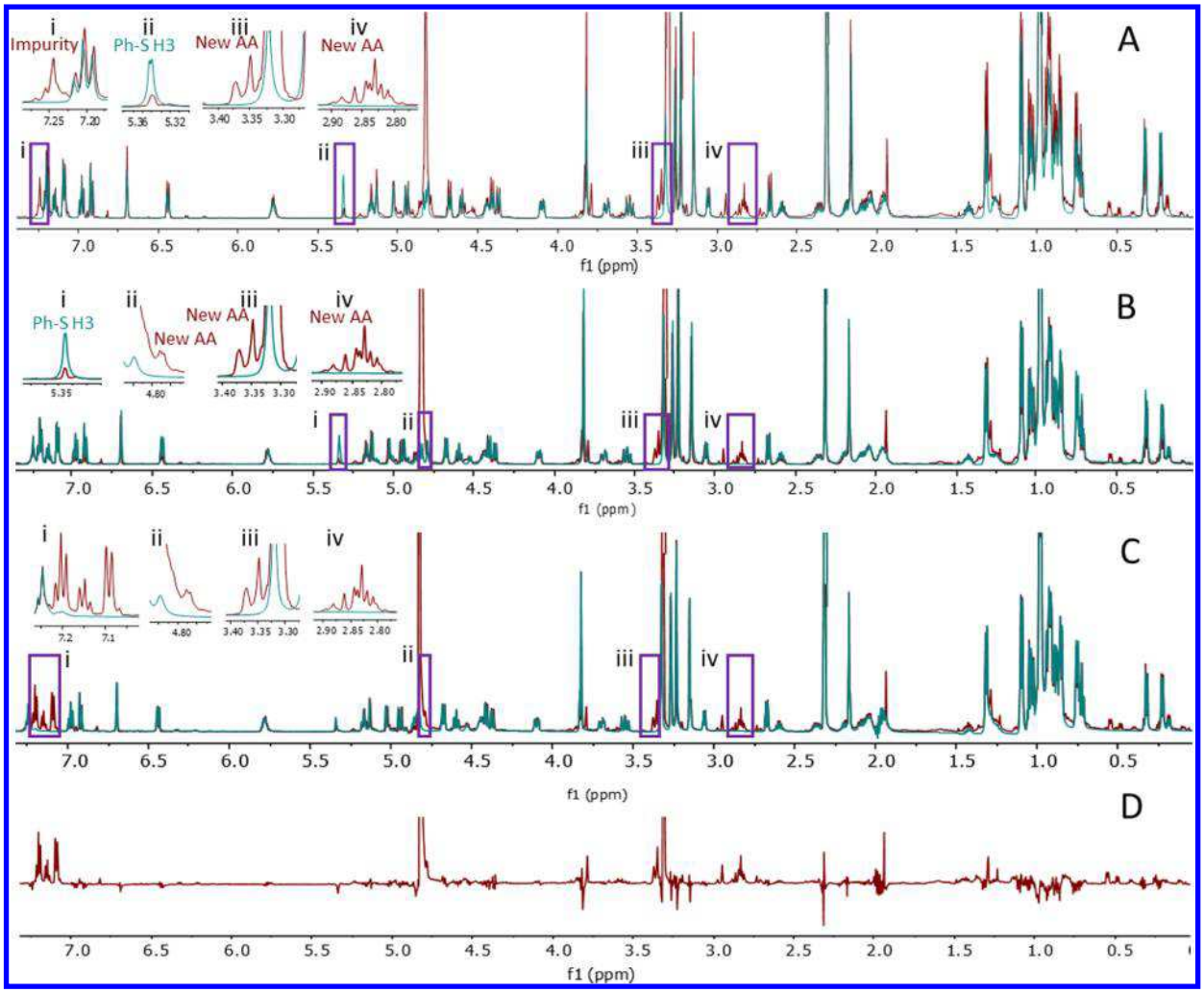



**Figure 2.** HiFSA structural sequencing of deoxyecumicin (**2**). (A) Comparison of the pseudo HiFSA profile with the $^1$H NMR spectrum (600 MHz, $CD_3OD$). (B) Comparison of the pseudo HiFSA profile including the major impurity with the $^1$H NMR spectrum. (C) Comparison of the HiFSA profiles of 12 mutual amino acids with the $^1$H NMR spectrum revealed the structure difference between **1** and **2**. (D) The HiFSA profile of 12 mutual amino acids was subtracted from the $^1$H NMR spectrum of deoxyecumicin. Red, experimental $^1$H NMR spectrum; blue, pseudo HiFSA profile.

**Anti-*M. tb* Activity.** These six new analogues of **1** have shown similar anti-*M. tb* activity to that of **1**. MICs against the *M. tb* H37Rv strain in the microplate Alamar Blue assay are 0.55, 0.15, and 0.13 $\mu$g/mL for **2**, **3**, and **4**, whereas the MIC of ecumicin is 0.26 $\mu$g/mL.[16] Due to the limited sample available, anti-*M. tb* activities of **5**−**7** were not measured individually, but the parent fraction containing mainly these three peptides was evaluated and an MIC of 0.26 $\mu$g/mL observed.

**The HiFSA Structural Sequencing Workflow.** The workflow of the NMR structural sequencing consisted of the following four main steps. Step 1: Finding commonalities and differences: preliminary identification of resonance(s) that reflect the structural difference(s) between the reference molecule (**1**) and an unknown analogue was done by comparing prospective "pseudo" HiFSA fingerprints to the $^1$H NMR spectrum of the unknown analogue. A pseudo HiFSA fingerprint was then generated by refining the HiFSA profile of the reference molecule, **1**, against the experimental spectrum of the analogue to be elucidated, adjusting for minor spectral differences. Step 2: Identification of the common amino acids: a $^1$H NMR fingerprint that represents the amino acids common to both molecules was generated by extracting all spectral parameters of the amino acids that contain the resonance(s) identified in the previous step from the HiFSA fingerprint of the reference molecule. Step 3: Identification of the altered amino acid(s): the resonances that constitute the spin system of the altered amino acid in the analogue were identified by comparing the HiFSA profile generated in the previous step to the experimental spectrum of the altered analogue. Step 4: Elucidation of the structure of the amino acid based on the diagnostic $^1$H NMR $\delta$ and $J$ patterns.

Instead of using a set of general digital (HiFSA) representations of amino acid $^1$H NMR data for the analyses of **2**−**7**, the well-studied congener peptide ecumicin (**1**) was used as the reference molecule in the present study. The relatively large number of amino acids and the rigid ring structure of the ecumicin family of peptides create unique microenvironments for each amino acid, affecting the proton chemical shifts significantly. For example, although having very similar coupling patterns, the five valine residues in **1** have distinctive and unique proton chemical shifts. The terminal methyl protons of the *N*-methyl-leucine residue are shielded by two spatially proximate aromatic rings and thus give rise to unusually high field signals. Because of highly likely variations in


DOI: 10.1021/acs.jnatprod.7b00207
J. Nat. Prod. 2017, 80, 2630−2643

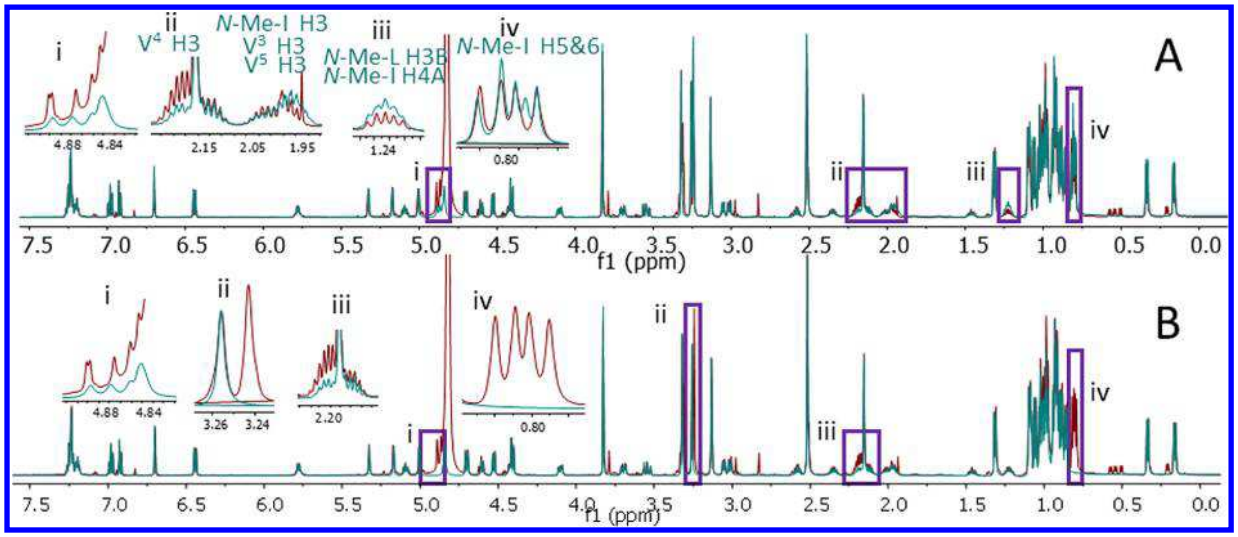



**Figure 3.** HiFSA structural sequencing of norecumicin (**3**). (A) Comparison of the pseudo HiFSA profile with the experimental spectrum (900 MHz, $CD_3OD$). (B) Comparison of the HiFSA profile of 12 mutual amino acids with the experimental spectrum revealed the structure difference between ecumicin and norecumicin. Red, experimental $^1$H NMR spectrum; blue, pseudo HiFSA profile.

the $^1$H chemical shifts of the individual amino acids in complex peptides with at least secondary structure, and due to the lack of models that can predict these variations with sufficient accuracy, relying on a set of general digital representations of amino acid $^1$H $\delta/J$ data for the analysis remains problematic. To overcome this difficulty, the HiFSA profile of **1** was used in the present study, as the foundation for the generation of the HiFSA fingerprints of individual amino acids for **2**−**7**. As all seven peptides are close analogues, the microenvironments for corresponding amino acids were initially assumed to be similar. In addition, PERCH iteration (see Experimental Section) was applied to the fingerprint of **1** against the spectrum of the unknown in the first step of the HiFSA sequencing method. This allowed adjustment for any subtle spectral differences that would be introduced by the small structural difference(s) at a remote location and/or different experimental conditions. Meanwhile, the major spectral differences that reflected any corresponding structural variations would remain for the most part unchanged during the iteration.

**Structure Elucidation of Deoxyecumicin (2).** Comparison of the pseudo HiFSA profile and $^1$H NMR spectrum of **2** (Figure 2A) revealed that the structural differences of **1** and **2** were within the phenyl-serine unit: the $\beta$-proton signal of the phenyl-serine unit ($\delta$ 5.34 ppm, doublet) present in **1** was absent in **2**. Additionally, three resonances ($\delta$ 7.25, 3.35, and 2.84 ppm) that integrated for approximately one proton each were only observed in the experimental spectrum, but not in the pseudo HiFSA profile. Whereas resonances at $\delta$ 3.35 and 2.84 ppm may represent protons belonging to **2**, the signal at $\delta$ 7.25 ppm is likely an impurity. As the broad, asymmetric nature of the singlet at $\delta$ 7.25 ppm resembles an overlapped phenylalanine proton resonance in **1**, it might be attributed to a peptide impurity in the sample.

To remove the impact of the major impurity on analysis of the spectrum and to confirm the origin of the $\delta$ 7.25 ppm signal, a second pseudo HiFSA profile was generated, again using the HiFSA profile of **1** as the basis for both **2** and the major impurity that contributes to the broad singlet at $\delta$ 7.25 ppm. As shown in Figure 2B, the second pseudo HiFSA better explains the $^1$H NMR spectrum of **2**, justifying further calculations to be made on this basis.

After removing the signals of phenyl-serine from the second pseudo HiFSA profile, $^1$H NMR spectra of common amino acids were added for consideration. Comparison of the resulting spectra with the experimental spectrum by visual examination resulted in isolation of the $^1$H NMR subspectrum of the different amino acid that was only present in **2** (Figure 2C). This amino acid was identified as phenylalanine due to the signals for five aromatic protons ($\delta$ 7.05 to 7.25 ppm), an $\alpha$-proton ($\delta$ 4.80 ppm, overlapped with the HDO peak), and two presumably $\beta$-protons ($\delta$ 2.84 and 3.35 ppm).

Theoretically, comparison of HiFSA profiles and experimental $^1$H NMR spectra could be done by a spectral subtraction approach. However, due to the relatively high abundance of other signals (impurities and conformers), subtracting the HiFSA profile of the common amino acids from the experimental spectrum of **2** led to a noisy subspectrum (Figure 2D), which complicated further analysis. The 100% qHNMR method[18] determined that the deoxyecumicin sample contained **2** and two closely related species in the ratio 100:35:18. Even though one major impurity was considered during the HiFSA process, signals of the minor impurity could still easily mask signals of the main compound. A one-proton double doublet of the main compound has four lines, so each line has the integration of approximately 0.25 H equivalents; one line of a methyl doublet from the 15% impurity would integrate for 0.23 H equivalents. As the $\beta$-proton signals mostly have dqq multiplicity in **1**−**7**, each line would have an even smaller integration than 0.25H. At similar line width, the terminal methyl proton singlet from the minor impurity would create background noise, which would readily mask the signal of a high-field $\beta$-proton signal. Additionally, a very subtle shift of the spectral window that corresponds to only a few data points is possible when generating the 256k $^1$H spectrum from the PMS text file in a different software (e.g., Mnova). While such a subtle change could remain undetected visually, it may create sufficient noise to be potentially confused with a $\beta$-proton signal from the main peptide when performing a point-by-point subtraction of two 256k $^1$H NMR spectra. Comparison of spectra by subtraction requires rigorous alignment of the two spectra and a reliable mechanism to distinguish signals of analytes from noise. Due to limitations of


DOI: 10.1021/acs.jnatprod.7b00207
J. Nat. Prod. 2017, 80, 2630−2643

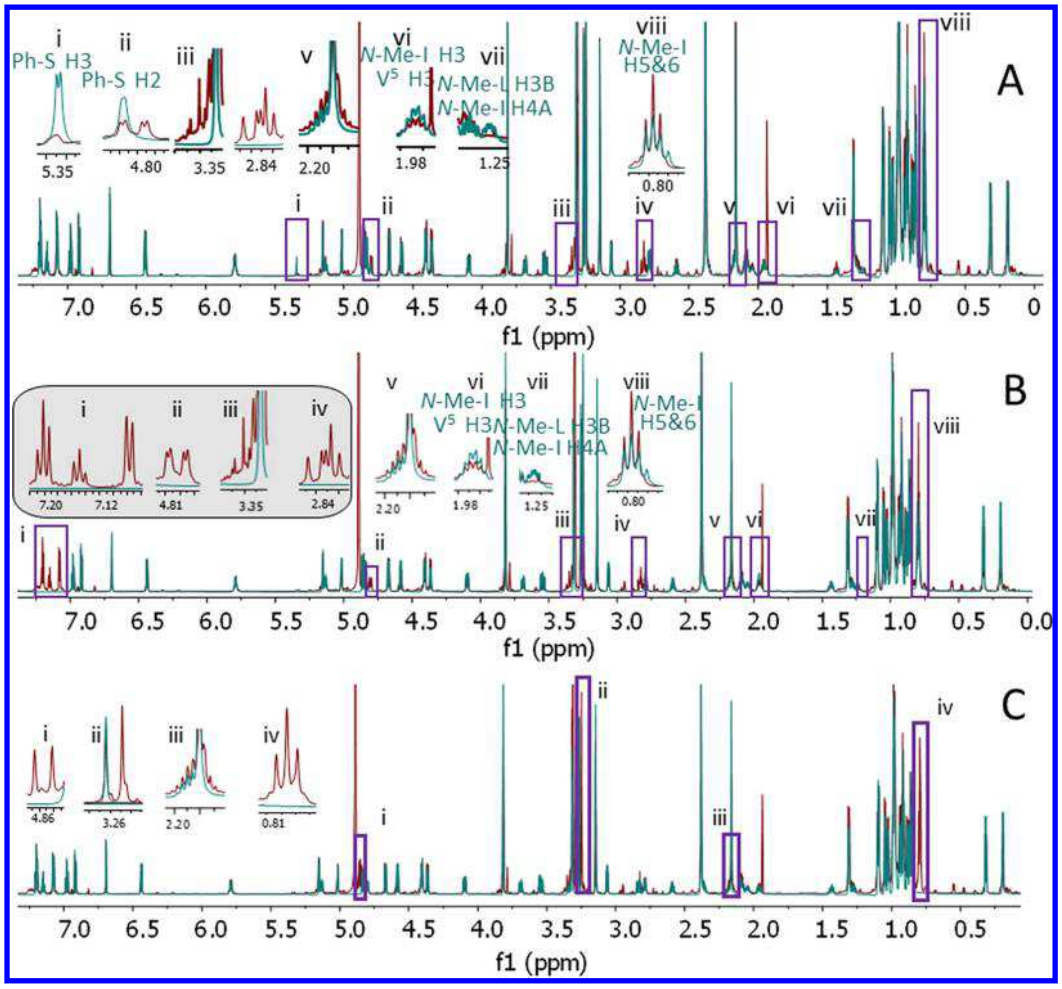



**Figure 4.** HiFSA structural sequencing of nordeoxyecumicin (**4**). (A) Comparison of the pseudo HiFSA profile with the experimental spectrum (900 MHz, $CD_3OD$). (B) Comparison of the HiFSA profile of 11 mutual amino acids plus the *N*-methyl-isoleucine (*N*-Me-I) residue with the experimental spectrum and (C) HiFSA profile of 11 mutual amino acids plus the *N*-methyl-valine residue with the experimental spectrum revealed the structure difference between ecumicin and nordeoxyecumicin. Red, experimental $^1$H NMR spectrum; blue, pseudo HiFSA profile.

available algorithms and/or tools at the time of this study, this step of the spectral analysis was, therefore, performed by visual comparison.

**Structure Elucidation of Norecumicin (3).** Comparison of the pseudo HiFSA profile and the experimental spectrum revealed four different signals (Figure 3A). The signals at $\delta$ 2.19 ppm integrated for two protons in the experimental spectrum, but only one of them could be explained in the pseudo HiFSA profile; resonances at $\delta$ 1.96 ppm integrated for two protons in the experimental spectrum, but in the pseudo HiFSA profile, the signals integrated for three protons that belong to the *N*-methyl-isoleucine and two valine units. A signal at $\delta$ 1.23 ppm integrated for only one proton in the experimental spectrum, but in the pseudo HiFSA profile, it integrated for two protons of *N*-methyl-isoleucine and *N*-methyl-leucine residues. Furthermore, one methyl signal ($\delta$ 0.81 ppm) belonging to the *N*-methyl-isoleucine residue gave rise to a doublet in the experimental spectrum, but to a triplet in the calculated spectrum. Collectively, this information led to the assumption that the structural difference between **1** and **3** is related to the *N*-methyl-isoleucine residue.

The $^1$H NMR subspectrum of the different amino acid in **3** was isolated by comparison of the spectra of the common amino acids with the experimental one (Figure 3B). As the distinguishing amino acid unit has one $\alpha$-proton ($\delta$ 4.88 ppm, half masked by the broad HDO resonance), one *N*-methyl singlet ($\delta$ 3.24 ppm), one $\beta$-proton ($\delta$ 2.18 ppm), and two methyl groups ($\delta$ 0.81 and 0.79 ppm), it was concluded that this observation is consistent with the presence of an *N*-methyl-valine residue.

**Structure Elucidation of Nordeoxyecumicin (4).** The pseudo HiFSA profile of **4** was compared to its experimental spectrum. This time, spectral differences were observed involving both the phenyl-serine and the *N*-methyl-isoleucine residues (Figure 4A). The $\beta$-proton signal of the phenyl-serine moiety of **1** ($\delta$ 5.34 ppm, doublet) was absent from the experimental spectrum of **4**, whereas the $\alpha$-proton signal of the same residue ($\delta$ 4.80 ppm, doublet in **1**) showed up as a dd in the experimental spectrum. In addition, two resonances at $\delta$ 3.35 and 2.84 ppm that integrated for a little more than one proton equivalent each were only observed in the experimental spectrum. The signal(s) at $\delta$ 2.19 ppm integrated for two protons in the experimental spectrum, but only accounted for one proton in the pseudo HiFSA profile. In addition, the signal at $\delta$ 1.97 ppm integrated for only one proton in the experimental spectrum, but in the pseudo HiFSA profile it integrated for two protons that were associated with the *N*-methyl-isoleucine and a valine residue. Whereas the signal at $\delta$


DOI: 10.1021/acs.jnatprod.7b00207
J. Nat. Prod. 2017, 80, 2630−2643

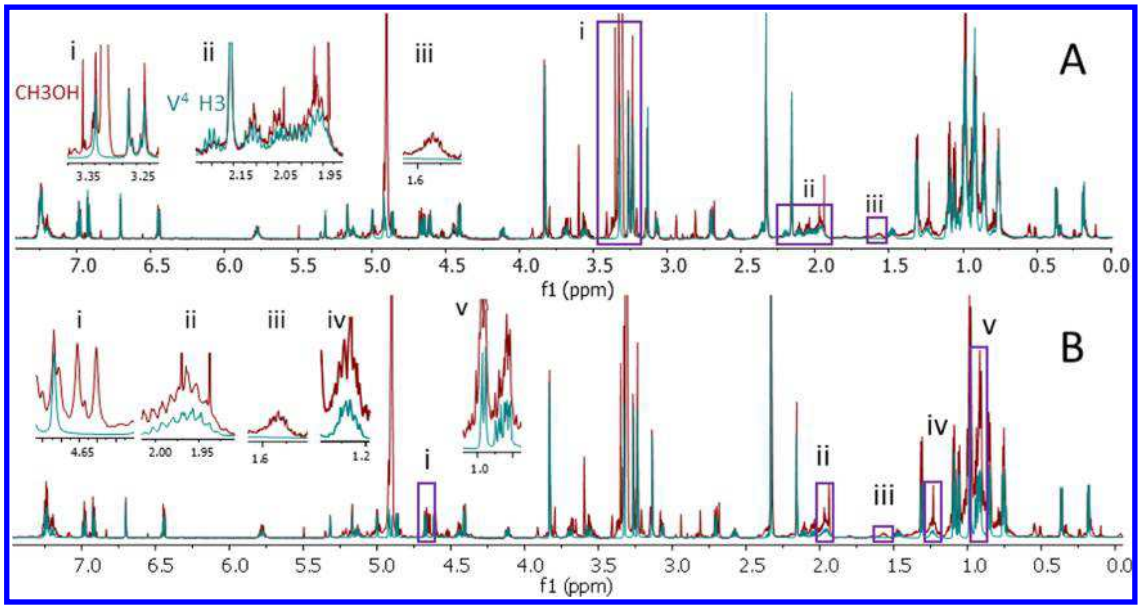



**Figure 5.** HiFSA structural sequencing of homoecumicin-ile11 (**5**). (A) Comparison of the pseudo HiFSA profile with the $^1$H experimental spectrum (700 MHz, $CD_3OD$). (B) Comparison of the HiFSA profile of 12 mutual amino acids with the experimental spectrum only suggested the possible identity of the different amino acid. Red, experimental $^1$H NMR spectrum; blue, pseudo HiFSA profile.

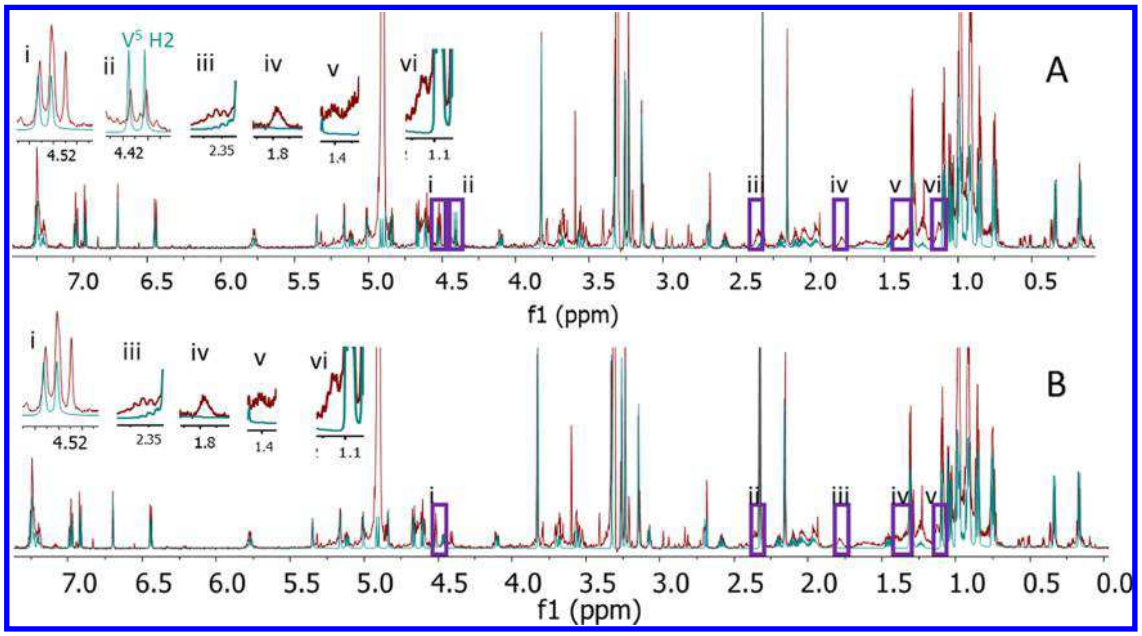



**Figure 6.** HiFSA structural sequencing of homoecumicin-ile13 (**6**). (A) Comparison of the pseudo HiFSA profile with the experimental spectrum (700 MHz, $CD_3OD$). (B) Comparison of the HiFSA profile of 12 mutual amino acids with the experimental spectrum suggested the identity of the different amino acid. Red, experimental $^1$H NMR spectrum; blue, pseudo HiFSA profile.

1.23 ppm integrated for only one proton in the experimental spectrum, the pseudo HiFSA profile showed that it represented two protons belonging to *N*-methyl-isoleucine and *N*-methyl-leucine residues. The methyl signal at $\delta$ 0.87 ppm showed up as a doublet in the experimental spectrum, but in the pseudo HiFSA profile it was a triplet belonging to the *N*-methyl-isoleucine unit.

Initially, spectral parameters corresponding to the phenyl-serine residue were removed from the pseudo HiFSA profile, leading to a new HiFSA profile of the common amino acids plus the *N*-methyl-isoleucine residue. A comparison revealed seven resonances in the experimental spectrum that did not match with any signals from the modified HiFSA profile. Six of the signals (marked by the gray box in Figure 4B) represented the $^1$H NMR pattern of a phenylalanine.

Thus, after manual addition of the spectral parameters of phenylalanine and reiteration, the spectral parameters on the *N*-methyl-isoleucine residue were removed from the HiFSA profile. This time, the comparison led to the identification of the $^1$H NMR subspectrum of an *N*-methyl-valine residue in **4**. As shown in Figure 4C, the spectral features of this *N*-methyl-valine residue include signals tentatively assigned to one $\alpha$-proton ($\delta$ 4.86 ppm, doublet), one *N*-methyl group ($\delta$ 3.25 ppm, 3H, singlet), one $\beta$-proton ($\delta$ 2.19 ppm, multiplet), and two methyl groups ($\delta$ 0.88 and 0.87 ppm, doublets). This


DOI: 10.1021/acs.jnatprod.7b00207
J. Nat. Prod. 2017, 80, 2630−2643

completed the assignment of the spectrum for nordeoxyecumicin (**4**).

**Structure Elucidation of Homoecumicin-ile11 (5).** Despite the fact that the HR-MS data indicated the homoecumicin-ile11 sample to be very pure, the upfield signals ($\delta$ 0.0 to 0.4 ppm) in the $^1$H NMR spectrum suggested that it contained at least three different analogues of **1** in the ratio of 100:29:17. Because of this, the pseudo HiFSA of **5** was generated from the ecumicin HiFSA profile, which was used to simulate both the major peptide and one major impurity (Figure 5A). One sharp singlet was observed at $\delta$ 3.35 ppm, which integrated for approximately three protons. This was considered an indication for an additional *N*-methylation or *O*-methylation in homoecumicin-ile11, which explained the 14 amu difference. However, as the signal line width was much smaller than that of any neighboring *N*-methyl and *O*-methyl signals, this signal more likely arose from the residual $CH_3OD$ in the NMR solvent.[19] Additional multiplet signals at $\delta$ 1.57 and 1.22 ppm, each integrating for one proton, indicated an additional methylene group in **5**, explaining the 14 amu mass difference. The observation that the $\beta$-signal of one of the valines in **1** [the fourth valine ($V^4$) or 11th amino acid from the *N*-terminal] had shifted from $\delta$ 2.20 to 1.95 ppm was an indication that the additional methylene group might have been inserted next to this $\beta$-proton of **1**.

Comparison of the HiFSA profiles of the common amino acids with the experimental spectrum (Figure 5B) revealed that the different amino acid in **5** is a leucine or isoleucine residue. However, heavy overlap in the terminal methyl signal region by impurity or conformer signals hampered the deduction of the coupling patterns of the methyl proton signals for final identification of this residue. For the same reason, extraction of the coupling pattern of the $\beta$-proton signal was problematic. However, the unambiguous doublet multiplicity of the $\alpha$-proton allowed the conclusion that the altered amino acid could only be an isoleucine.

**Structure Elucidation of Homoecumicin-ile13 (6).** Close similarities between the upfield signal patterns in the $^1$H NMR spectra of the samples containing **5** and **6** as major constituents suggested that **6** was accompanied by three ecumicin analogues in the ratio 100:42:18:8. The HRMS data indicated that all four were isomeric compounds. Therefore, the HiFSA profile of **1** was again used to build both the major component and the major impurity when generating the pseudo HiFSA profile of **6** (Figure 6A). Comparing the pseudo HiFSA profile of **6** to the experimental spectrum, some differences in the terminal methyl region were observed. However, heavy overlap in the region and the existence of multiple peptide species in the sample containing **6** made it impossible to identify the true difference. Proton multiplets at $\delta$ 2.36, 1.79, 1.40, and 1.12 ppm, each integrating for approximately one proton, were observed only in the experimental spectrum. According to their coupling patterns, two of them could belong to an additional methylene group, resulting in the 14 amu molecular weight increase in **6** compared to **1**. The other two signals were considered to result from the overlapped impurity signals. The pseudo HiFSA profile contained a strong doublet at $\delta$ 4.40 ppm belonging to the $\alpha$-proton of the C-terminal amino acid, a valine residue in **1**, whereas the experimental spectrum has only some impurity signal at $\delta$ 4.40 ppm. Although almost completely masked by the broad HDO peak, an additional doublet was observed at $\delta$ 4.51 ppm in the experimental spectrum. This observation suggested that the C-terminal amino acid in **6** may be a leucine or isoleucine residue.

Comparing the HiFSA profile of the common amino acids with the experimental spectrum (Figure 6B) confirmed that the C-terminal amino acid in **6** is indeed an isoleucine residue. Although the heavily overlapped $\beta$-signal and terminal methyl signal regions made the elucidation difficult, the doublet coupling pattern of the $\alpha$-proton finalized the assignment of the isoleucine.

**Structure Elucidation of Homoecumicin-ile6 (7).** In a similar manner to that for **5** and **6**, multiple isomers with very near identical chromatographic behavior invalidated the MS-based purity assignments. The impure character of the isolate containing **7** was evident from the upfield signals in the $^1$H NMR spectrum, suggesting the existence of at least six ecumicin-related species in the approximate ratio of 100:82:60:54:30:19. The high residual complexity and the relatively high proportions of analogues contained in the sample of **7** pushed the limits of the computational iteration involved in HiFSA iteration. Therefore, iteration of the spin parameters was done manually, followed by the HiFSA sequencing method as described above, both aided by additional 2D COSY information. The detailed elucidation procedure can be found in the Supporting Information.

**Verification of the Proposed Structures.** The slightly higher isolation yield of **2**−**4** made the acquisition of 1D-$^{13}$C, 2D-COSY, 2D-HSQC, 2D-HMBC, and 2D-semiselective HMBC experiments possible. These data (see Supporting Information) fully supported the structures proposed by the HiFSA sequencing approach, which was completed prior to the acquisition of the 2D NMR spectra. Following the HiFSA structural sequencing method, the connectivity of the amino acids was deduced from the assertion that the ecumicin family of peptides is biosynthesized by the same nonribosomal peptide synthetase (NRPS). In addition, substituent chemical shift (s.c.s.) effects are highly sensitive indicators of structural similarity. The topology and magnitude of s.c.s effected in out-of-sequence and other major topological analogues affect the entire molecule and are at least an order of magnitude larger than what was observed for **2**−**7**. Notably, the observed intramolecular aromatic induced s.c.s. effects led to small but diagnostic s.c.s. patterns in **2**−**7**, similar to aromatic solvent induced chemical shifts (ASICS) in aromatic NMR solvents, but highly selective for certain parts of the molecules. These observations are in line with the presence of simple "point mutations", with shifted Me and/or added OH groups. Therefore, the observed single or dual alteration(s) of the amino acid had to be located in place of the original amino acid(s) in **1**. Unlike in **1**, which has two isochronic carbonyl carbons making mass fragmentation patterns a requirement for the final structural elucidation, all carbonyl carbons in **2**−**4** showed distinct resonances. Therefore, the 2D semiselective HMBC experiments performed with **2**−**4** were able to determine the sequences unequivocally, which mutually confirmed the structures proposed by the HiFSA sequencing method.

Samples **5**−**7** were considered pure by LC-HRMS analysis. However, $^1$H NMR spectroscopy clearly revealed their impure nature and the differences among these samples, which were otherwise not discernible even by high-resolution MS. In addition, information embedded in the $^1$H NMR data was able to support identification of the different amino acids and eventual structure elucidation. For **5**−**7**, due to very limited


DOI: 10.1021/acs.jnatprod.7b00207
J. Nat. Prod. 2017, 80, 2630−2643

Table 1. $^{1}H$ and $^{13}C$ NMR Data of Deoxyecumicin (2; $CD_3OD$, 600 MHz for $^{1}H$, 100 MHz for $^{13}C$)

| moiety and position | | $\delta_C$, type | $\delta_H$, $J$ in Hz[a] |
|---|---|---|---|
| *N,N*-Me$_2$-V | 1 | 173.35, C | |
| | 2 | 75.78, CH | 2.6681, d (9.03) |
| | 3 | 28.49, CH | 2.0725, dqq (9.03, 7.33, 6.66) |
| | 4, 5 | 42.40, CH$_3$ | 2.3140, s |
| | 6 | 20.16, CH$_3$ | 0.9807, d (7.33) |
| | 7 | 19.40, CH$_3$ | 0.8510, d(6.66) |
| L-V[1] | 1 | 174.96, C | |
| | 2 | 55.76, CH | 4.6735, d (8.37) |
| | 3 | 31.57, CH | 2.0725, dqq (8.37, 6.49, 6.58) |
| | 4 | 19.48, CH$_3$ | 1.0450, d (6.49) |
| | 5 | 19.69, CH$_3$ | 1.0268, d (6.58) |
| *N*-Me-L-*allo*-L | 1 | 172.86, C | |
| | 2 | 61.60, CH | 4.9446, d (11.32) |
| | 3 | 34.42, CH | 1.9491, dddq (11.32, 2.21, 7.00, 6.30) |
| | 4 | 26.36, CH$_2$ | 1.0001 ddq (2.21, −14.81, 7.10) |
| | | | 1.2670, ddq (7.00, −14.81, 7.74) |
| | 5 | 19.36, CH$_3$ | 0.7522, d (6.30) |
| | 6 | 15.19, CH$_3$ | 0.7235, dd (7.10, 7.74) |
| | 7 | 31.52, CH$_3$ | 3.2279, s |
| L-T | 1 | 171.92, C | |
| | 2 | 53.37, CH | 5.1338, d (2.12) |
| | 3 | 70.80, CH | 5.78360, dq (2.12, 6.54) |
| | 4 | 17.80, CH$_3$ | 1.3146, d (6.54) |
| *N*-Me-L-T | 1 | 171.42, C | |
| | 2 | 63.92, CH | 5.0278, d (3.63) |
| | 3 | 67.83, CH | 4.4355, dq (3.63, 7.13) |
| | 4 | 19.89, CH$_3$ | 0.8860, d (7.13) |
| | 5 | 35.28, CH$_3$ | 3.3201, s |
| L-V[2] | 1 | 174.93, C | |
| | 2 | 56.87, CH | 4.82, d (9.22) |
| | 3 | 31.84, CH | 2.3565, dqq (9.22, 6.60, 6.50) |
| | 4 | 19.64, CH$_3$ | 1.0956, d (6.60) |
| | 5 | 19.70, CH$_3$ | 0.9807, d (6.50) |
| *N*-Me-L-L | 1 | 173.54, C | |
| | 2 | 56.80, CH | 5.1683, dd (7.43, 7.42) |
| | 3 | 40.50, CH$_2$ | 1.2917, ddd (7.43, −14.05, 1.87) |
| | | | 1.4257, ddd (7.42, −14.05, 5.78) |
| | 4 | 26.52, CH | 0.9942, ddqq (1.87, 5.78, 6.64, 6.64) |
| | 5 | 21.96, CH$_3$ | 0.2241, d (6.64) |
| | 6 | 23.54, CH$_3$ | 0.3218, d (6.64) |
| | 7 | 31.63, CH$_3$ | 3.2632, s |
| L-V[3] | 1 | 173.63, C | |
| | 2 | 55.59, CH | 4.6016, d (9.22) |
| | 3 | 32.74, CH | 2.0383, dqq (9.22, 6.41, 6.20) |
| | 4 | 19.3, CH$_3$ | 0.9147, d (6.41) |
| | 5 | 19.48, CH$_3$ | 0.8677, d (6.20) |
| *N*-Me-L-V | 1 | 171.09, C | |
| | 2 | 71.60, CH | 3.0543, d (7.60) |
| | 3 | 31.57, CH | 2.5878, dqq (7.60, 6.97, 6.51) |
| | 4 | 19.75, CH$_3$ | 0.9807, d (6.97) |
| | 5 | 22.04, CH$_3$ | 1.0956, d (6.51) |
| | 6 | 40.51, CH$_3$ | 3.1464, s |
| *N*-Me-4-OMe-L-Y | 1 | 171.57, C | |
| | 2 | 71.09, CH | 4.0950, dd (11.17, 4.49) |
| | 3 | 22.87, CH$_2$ | 3.5471, dd (11.17, −13.59) |
| | | | 3.6911, dd (4.49, −13.59) |
| | 4 | 113.30, C | |
| | 5 | 124.9, CH | 6.6976, d (0.87) |
| | 6 | 156.00, C | |
| | 7 | 107.00, CH | 6.9207, ddd (0.87, 8.23, 0.90) |
| | 8 | 124.30, CH | 6.9815, dd (8.23, 7.80) |
| | 9 | 100.80, CH | 6.4408, dd (0.90, 7.80) |
| | 10 | 140.90, C | |
| | 11 | 118.90, C | |
| | 12 | 55.74, CH$_3$ | 3.8183, s |
| | 13 | 41.92, CH$_3$ | 2.1653, s |
| L-V[4] | 1 | 174.11, C | |
| | 2 | 59.17, CH | 4.4070, d (8.33) |
| | 3 | 33.62, CH | 2.1896, dqq (8.33, 6.03, 7.96) |
| | 4 | 20.16, CH$_3$ | 0.9806, d (6.03) |
| | 5 | 19.79, CH$_3$ | 0.9806, d (7.96) |
| L-F | 1 | 174.29, C | |
| | 2 | 58.25, CH | 4.7800, dd (12.15, 10.60) |
| | 3 | 38.20, CH$_2$ | 3.3500, dd (12.15, −14.13) |
| | | | 2.8412, dd (10.60, −14.13) |
| | 4 | 140.90, C | |
| | 5, 9 | 131.22, CH | 7.0927, dddd (7.89, 1.29, 0.38, 1.86) |
| | 6, 8 | 130.47, CH | 7.2024, dddd (7.89, 7.38, 1.98, 0.38) |
| | 7 | 128.30, CH | 7.1495, dd (1.29, 7.38) |
| L-V[5] | 1 | 175.45, C | |
| | 2 | 59.36, CH | 4.3690, d (8.95) |
| | 3 | 33.62, CH | 1.9660, dqq (8.95, 7.27, 6.83) |
| | 4 | 19.54, CH$_3$ | 0.9367, d (7.27) |
| | 5 | 19.78, CH$_3$ | 0.9199, d (6.83) |

[a]Interpretation of $^{1}H$ NMR data was done by means of $^{1}H$ iterative full-spin analysis.

sample amounts, an ultrahigh-field NMR instrument (700 MHz) equipped with a 1.7 mm cryoprobe designed for small samples was used in an attempt to acquire high-quality 2D NMR data. Unfortunately, still not all 2D NMR experiments (see Supporting Information) could yield the information to provide unequivocal NMR evidence for the structures. In addition, when analyzed at 900 MHz $^{1}H$ NMR using instrumentation equipped with a 5 mm cryoprobe, only $^{1}H$ and COSY experiments yielded useful information (data not shown). As a result, the ability of the HiFSA structural sequencing method to extract key NMR parameters from the heavily overlapped $^{1}H$ NMR spectra and support the structure elucidation process is invaluable. Available information from the COSY, HSQC, and HMBC spectra of **5**−**7** was consistent with the proposed structures. The low signal-to-noise ratio in the carbonyl region of the HMBC spectra and limited signal intensity of the semiselective HMBC spectra (data not shown) precluded assignment of all carbon signals in these molecules.

**Stereochemistry of the Ecumicin Analogues.** The isolated peptides are close analogues to the well-studied nonribosomally encoded peptide ecumicin (**1**), which suggests that all are synthesized by the same NRPS. The lack of an epimerization domain[15] in the NRPS responsible for **1** suggests that all amino acids in the cyclic peptide analogues also have the L-configuration, leading to tentative assignments of stereochemistry for **2**−**7** as shown in Figure 1.

**HiFSA Fingerprints of 2−6.** The HiFSA profiles of **2**−**6**, which were generated as a byproduct of the HiFSA structural



DOI: 10.1021/acs.jnatprod.7b00207
J. Nat. Prod. 2017, 80, 2630−2643

Table 2. $^{1}$H and $^{13}$C NMR Data of Norecumicin (3; $CD_3OD$, 900 MHz for $^{1}$H, 225 MHz for $^{13}$C)

| Moiety and position | | $\delta_C$, type | $\delta_H$, $J$ in Hz[a] |
|---|---|---|---|
| *N,N*-Me$_2$-V | 1 | 173.4, C | |
| | 2 | 75.75, CH | 2.6754, d (9.19) |
| | 3 | 28.81, CH | 2.0426, dqq (9.19, 6.75, 6.46) |
| | 4,5 | 42.38, CH$_3$ | 2.3153, s |
| | 6 | 19.83, CH$_3$ | 1.0925, d (6.75) |
| | 7 | 19.89, CH$_3$ | 0.8459, d (6.46) |
| L-V[1] | 1 | 175.11, C | |
| | 2 | 55.94, CH | 4.6720, d (8.76) |
| | 3 | 31.78, CH | 2.1028, dqq (8.76, 6.78, 6.69) |
| | 4 | 20.11, CH$_3$ | 0.9772, d (6.78) |
| | 5 | 19.34, CH$_3$ | 1.0558, d (6.69) |
| *N*-Me-L-V | 1 | 172.66, C | |
| | 2 | 63.10, CH | 4.8340, d (11.07) |
| | 3 | 28.04, CH | 2.1756, dqq (11.07, 6.42, 6.61) |
| | 4 | 22.15, CH | 0.8099, d (6.42) |
| | 5 | 19.48, CH$_3$ | 0.7938, d (6.61) |
| | 6 | 31.56, CH$_3$ | 3.2486, s |
| **L-T** | 1 | 172.01, C | |
| | 2 | 53.37, CH | 5.1740, d (2.29) |
| | 3 | 70.06, CH | 5.7859, dq (2.29, 6.51) |
| | 4 | 16.89, CH$_3$ | 1.3130, d (6.51) |
| *N*-Me-L-T | 1 | 171.20, C | |
| | 2 | 63.15, CH | 5.0085, d (4.01) |
| | 3 | 66.84, CH | 4.4263, dq (4.01, 6.61) |
| | 4 | 20.17, CH$_3$ | 0.9197, d (6.61) |
| | 5 | 34.34, CH$_3$ | 3.3232, s |
| L-V[2] | 1 | 174.83, C | |
| | 2 | 56.78, CH | 4.8533, d (8.92) |
| | 3 | 31.92, CH | 2.3514, dqq (8.92, 7.23, 6.84) |
| | 4 | 19.50, CH$_3$ | 1.0938, d (7.23) |
| | 5 | 19.36, CH$_3$ | 0.9780, d (6.84) |
| *N*-Me-L-L | 1 | 173.32, C | |
| | 2 | 55.88, CH | 5.0844, dd (6.77, 8.59) |
| | 3 | 39.50, CH$_2$ | 1.2189, ddd (6.77, −13.57, 6.30) |
| | | | 1.4660, ddd (8.59, −13.57, 6.73) |
| | 4 | 25.76, CH | 0.9918, ddqq (6.30, 6.73, 6.62, 6.57) |
| | 5 | 21.73, CH$_3$ | 0.1480, d (6.62) |
| | 6 | 23.62, CH$_3$ | 0.3357, d (6.57) |
| | 7 | 31.83, CH$_3$ | 3.2603, s |
| L-V[3] | 1 | 173.46, C | |
| | 2 | 55.38, CH | 4.5957, d (8.91) |
| | 3 | 33.04, CH | 2.0272, dqq (8.91, 6.83, 6.82) |
| | 4 | 19.51, CH$_3$ | 0.9124, d (6.83) |
| | 5 | 19.58, CH$_3$ | 0.85921, d (6.82) |
| *N*-Me-L-V | 1 | 171.06, C | |
| | 2 | 71.49, CH | 3.0639, d (7.58) |
| | 3 | 30.08, CH | 2.5819, dqq (7.58, 7.16, 6.12) |
| | 4 | 19.79, CH$_3$ | 0.9801, d (7.16) |
| | 5 | 19.76, CH$_3$ | 1.0939, d (6.12) |
| | 6 | 40.50, CH$_3$ | 3.1378, s |
| *N*-Me-4-OMe-L-Y | 1 | 171.65, C | |
| | 2 | 71.23, CH | 4.1043, dd (11.17, 4.67) |
| | 3 | 27.10, CH$_2$ | 3.5450, dd (11,17, −13.72) |
| | | | 3.6934, dd (4.67, −13.72) |
| | 4 | 112.53, C | |
| | 5 | 123.95, CH | 6.6982, d (0.81) |
| | 6 | 155.35, C | |
| | 7 | 106.08, CH | 6.9175, ddd (0.81, 8.17, 0.57) |
| | 8 | 123.44, CH | 6.9816, dd (8.17, 7.75) |
| | 9 | 99.87, CH | 6.4422, dd (0.57, 7.75) |
| | 10 | 140.01, C | |
| | 11 | 118.45, C | |
| | 12 | 55.71, CH$_3$ | 3.8269, s |
| | 13 | 41.05, CH$_3$ | 2.1564, s |
| L-V[4] | 1 | 174.16, C | |
| | 2 | 59.20, CH | 4.5363, d (7.86) |
| | 3 | 33.78, CH | 2.1906, dqq (7.86, 6.81, 6.66) |
| | 4 | 19.53, CH$_3$ | 1.0268, d (6.81) |
| | 5 | 20.12, CH$_3$ | 0.9933, d (6.66) |
| Ph-L-*threo*-S | 1 | 173.80, C | |
| | 2 | 59.74, CH | 4.8839, d (1.59) |
| | 3 | 73.04, CH$_2$ | 5.3260, d (1.59) |
| | 4 | 142.80, C | |
| | 5,9 | 127.19, CH | 7.2292, dddd (7.72, 1.00, 0.78, 0.70) |
| | 6,8 | 129.43, CH | 7.2522, dddd (7.72, 7.50, 0.20, 0.78) |
| | 7 | 128.41, CH | 7.1984, dd (1.00, 7.50) |
| L-V[5] | 1 | 175.27, C | |
| | 2 | 59.24, CH | 4.4022, d (8.83) |
| | 3 | 33.02, CH | 1.9633, dqq (8.83, 6.73, 6.79) |
| | 4 | 19.58, CH$_3$ | 0.9370, d (6.73) |
| | 5 | 19.25, CH$_3$ | 0.9224, d (6.79) |

[a]Interpretation of $^{1}$H NMR data was done by means of $^{1}$H iterative full-spin analysis.

sequencing procedure, explain the vast majority of intensities of the experimental $^{1}$H NMR spectra, with root-mean-square (RMS) difference values of 0.089%, 0.074%, 0.070%, 0.269%, and 0.287%, respectively. The $^{1}$H and $^{13}$C assignments for **2**−**6** are given in Tables 1−5, respectively. Highly comprehensive $^{1}$H assignments based on HiFSA interpretations could only be achieved for **2**−**4**. Although a good agreement was achieved between the experimental and calculated spectra of **5** and **6**, the existence of a large number of yet uncharacterized variables for signals associated with impurities, especially in the heavily overlapped methyl region (0.80−1.10 ppm), reduced the reliability of iterated chemical shifts of the methyl groups. Because of this, the chemical shifts of these methyl proton signals are only reported to two decimal places.

In general, the HiFSA-based sequencing methodology is very sensitive for the detection of subtle structure differences. As shown by the study of the silybin and isosilybin isomers,[20] HiFSA is able to differentiate near identical molecules such as regio- and stereoisomers. The present study of ecumicin analogues again demonstrates the ability of HiFSA to distinguish closely related molecules and to perform such analyses for several congeners in the 1600 amu range. The HiFSA structural sequencing method solved the structures of complex peptides by recognizing the essential $^{1}$H NMR parameters of each constituent amino acid present in their crowded $^{1}$H NMR spectra.

The major limitation of this method is the lack of *direct* evidence for the amino acid sequence. However, due to available evidence for the reference molecules, the connectivity of the amino acids could be inferred from indirect information, i.e., that the ecumicin family of peptides are biosynthesized by the same NRPS. This deductive reasoning could be confirmed


DOI: 10.1021/acs.jnatprod.7b00207
J. Nat. Prod. 2017, 80, 2630−2643

Table 3. $^{1}H$ and $^{13}C$ NMR Data of Nordeoxyecumicin (4; $CD_3OD$, 900 MHz for $^{1}H$, 225 MHz for $^{13}C$)

| moiety and position | | $\delta_C$, type | $\delta_H$, $J$ in Hz[a] |
|---|---|---|---|
| *N*,*N*-$Me_2$-V | 1 | 173.41, C | |
| | 2 | 75.50, CH | 2.7900, d (8.77) |
| | 3 | 28.80, CH | 2.0856, dqq (8.77, 6.72, 7.02) |
| | 4,5 | 42.38, $CH_3$ | 2.3820, s |
| | 6 | 19.45, $CH_3$ | 1.0485, d (6.72) |
| | 7 | 20.13, $CH_3$ | 0.9837, d (7.02) |
| L-V[1] | 1 | 174.95, C | |
| | 2 | 56.02, CH | 4.6704, d (8.73) |
| | 3 | 31.60, CH | 2.0912, dqq (8.73, 6.57, 7.14) |
| | 4 | 19.18, $CH_3$ | 0.8581, d (6.57) |
| | 5 | 20.04, $CH_3$ | 0.9892, d (7.14) |
| *N*-Me-L-V | 1 | 172.83, C | |
| | 2 | 63.04, CH | 4.8569, d (11.03) |
| | 3 | 28.16, CH | 2.1606, dqq (11.03, 6.60, 6.58) |
| | 4 | 19.45, $CH_3$ | 0.7997, d (6.60) |
| | 5 | 19.43, $CH_3$ | 0.7924, d (6.58) |
| | 6 | 31.71, $CH_3$ | 3.2486, s |
| L-T | 1 | 171.86, C | |
| | 2 | 53.46, CH | 5.1524, d (2.33) |
| | 3 | 69.82, CH | 5.7910, dq (2.33, 6.35) |
| | 4 | 16.86, $CH_3$ | 1.3100, d (6.35) |
| *N*-Me-L-T | 1 | 171.31, C | |
| | 2 | 63.03, CH | 5.0160, d (3.85) |
| | 3 | 66.91, CH | 4.4076, dq (3.85, 6.46) |
| | 4 | 19.79, $CH_3$ | 0.8855, d (6.46) |
| | 5 | 34.35, $CH_3$ | 3.3160, s |
| L-V[2] | 1 | 174.90, C | |
| | 2 | 56.84, CH | 4.8360, d (9.05) |
| | 3 | 31.87, CH | 2.3632, dqq (9.05, 6.67, 6.97) |
| | 4 | 19.83, $CH_3$ | 1.0983, d (6.67) |
| | 5 | 19.90, $CH_3$ | 0.9748, d (6.97) |
| *N*-Me-L-L | 1 | 173.57, C | |
| | 2 | 55.90, CH | 5.1321, dd (7.92, 7.14) |
| | 3 | 39.55 $CH_2$ | 1.4392, ddd (7.92, −12.12, 4.16) |
| | | | 1.2809, ddd (7.14, −12.12, 5.74) |
| | 4 | 25.74, CH | 0.9573, ddqq (4.16, 5.74, 6.54, 6.59) |
| | 5 | 23.51, $CH_3$ | 0.3178, d (6.54) |
| | 6 | 21.84, $CH_3$ | 0.1932, d (6.59) |
| | 7 | 31.89, $CH_3$ | 3.2645, s |
| L-V[3] | 1 | 173.61, C | |
| | 2 | 55.50, CH | 4.5811, d (9.17) |
| | 3 | 32.92, CH | 2.0448, dqq (9.17, 6.77, 6.74) |
| | 4 | 19.67, $CH_3$ | 0.9145, d (6.77) |
| | 5 | 19.40, $CH_3$ | 0.8671, d (6.74) |
| *N*-Me-L-V | 1 | 171.14, C | |
| | 2 | 71.57, CH | 3.0619, d (7.74) |
| | 3 | 29.99, CH | 2.5911, dqq (7.74, 6.61, 6.84) |
| | 4 | 22.18, $CH_3$ | 1.0938, d (6.61) |
| | 5 | 19.77, $CH_3$ | 0.9788, d (6.84) |
| | 6 | 40.56, $CH_3$ | 3.1458, s |
| *N*-Me-4-OMe-L-Y | 1 | 171.58, C | |
| | 2 | 71.24, CH | 4.0958, dd (4.73, 11.18) |
| | 3 | 27.06 $CH_2$ | 3.6876, dd (4.73, −13.72) |
| | | | 3.5437, dd (11.18, −13.72) |
| | 4 | 112.52, C | |
| | 5 | 123.96, CH | 6.6951, d (0.81) |
| | 6 | 155.35, C | |
| | 7 | 106.09, CH | 6.9177, ddd (0.81, 8.17, 0.33) |
| | 8 | 123.44, CH | 6.9801, dd (8.17, 7.80) |
| | 9 | 99.87, CH | 6.4381, dd (0.33, 7.80) |
| | 10 | 140.01, C | |
| | 11 | 118.44, C | |
| | 12 | 55.71, $CH_3$ | 3.8158, s |
| | 13 | 41.10, $CH_3$ | 2.1622, s |
| L-V[4] | 1 | 174.05, C | |
| | 2 | 59.42, CH | 4.4060, d (8.47) |
| | 3 | 33.80, CH | 2.1757, dqq (8.47, 6.72, 7.18) |
| | 4 | 19.72, $CH_3$ | 1.0255, d (6.72) |
| | 5 | 20.21, $CH_3$ | 0.9819, d (7.18) |
| L-F | 1 | 174.30, C | |
| | 2 | 57.40, CH | 4.8021, dd (10.84, 2.65) |
| | 3 | 37.58 $CH_2$ | 2.8333, dd (10.84, −14.34) |
| | 4 | 140.01, C | |
| | 5,9 | 130.38, CH | 7.0772, dddd (6.77, 1.31, 0.24, 1.13) |
| | 6,8 | 129.64, CH | 7.2010, dddd (6.77, 7.55, 0.80, 0.24) |
| | 7 | 127.47, CH | 7.1497, dd (1.31, 7.55) |
| L-V[5] | 1 | 175.47, C | |
| | 2 | 59.24, CH | 4.3644, d (9.07) |
| | 3 | 33.29, CH | 1.9599, ddd (9.07, 6.85, 6.68) |
| | 4 | 19.31, $CH_3$ | 0.9224, d (6.85) |
| | 5 | 19.54, $CH_3$ | 0.9383, d (6.68) |

[a]Interpretation of $^{1}H$ NMR data was done by means of $^{1}H$ iterative full-spin analysis.

by the 2D semiselective HMBC experiments for the more abundant peptides (**2**−**4**) and provides reasonable albeit not unequivocal evidence for structure elucidation of mass limited samples (**5**−**7**) where supportive 2D NMR data are unavailable due to NMR sensitivity considerations.

## ■ CONCLUSION

The present work is an initial attempt to rely almost exclusively on 1D $^{1}H$ NMR information to solve the structures of the large complex 13-amino acid cyclic peptides that only differ in the presence of a hydroxy group or a methylene group or even differing in the location of a methylene group. The three homoecumicins (**5**−**7**) were indistinguishable even by high-resolution mass spectrometer, and fully inclusive 2D NMR data sets were not achievable despite the advanced instrumentation due to mass-limited samples. The ability of the HiFSA sequencing approach to extract key NMR parameters from even the heavily overlapped $^{1}H$ NMR spectra and support the structure elucidation process is invaluable. In five out of the six cases presented (**2**−**6**), a study and interpretation of the $^{1}H$ NMR spectrum led to successful structure elucidations of several unknown cyclic peptides. All structures proposed from this process were subsequently confirmed by available 1D $^{13}C$ and 2D NMR experiments, which also validated the HiFSA structural sequencing method.

The HiFSA structural sequencing method simultaneously fulfills two main goals of $^{1}H$ NMR interpretation: (1) to facilitate structural elucidation of small sample quantities of relatively large molecules and (2) to enable the reproducibility of $^{1}H$ NMR data for dereplication.[11] Naturally, a wealth of information in the $^{1}H$ spectrum is encoded in the rather complex coupling patterns ("multiplets"). The information is


DOI: 10.1021/acs.jnatprod.7b00207
J. Nat. Prod. 2017, 80, 2630−2643

Table 4. $^{1}$H and $^{13}$C NMR Data of Homoecumicin-ile11 (5; $CD_3OD$, 700 MHz for $^{1}$H, 175 MHz for $^{13}$C)

| moiety and position | | $\delta_C$[a] | $\delta_H$, $J$ in Hz[b] |
|---|---|---|---|
| *N*,*N*-Me$_2$-V | 1 | | |
| | 2 | 75.61 | 2.7053, d (9.03) |
| | 3 | 28.68 | 2.0536, dqq (9.03, 7.33, 6.66) |
| | 4,5 | 42.26 | 2.3315, s |
| | 6 | 19.4 | 0.9800, d (7.33) |
| | 7 | 20.0 | 0.8500, d (6.66) |
| L-V$^{1}$ | 1 | | |
| | 2 | 55.91 | 4.6702, d (8.37) |
| | 3 | 31.62 | 2.1085, dqq (8.37, 6.49, 6.58) |
| | 4 | 19.6 | 1.0600, d (6.49) |
| | 5 | 19.5 | 1.0000, d (6.58) |
| *N*-Me-L-*allo*-L | 1 | | |
| | 2 | 61.76 | 4.9142, d (11.32) |
| | 3 | 34.33 | 1.9542, dddq (11.32, 2.21, 7.00, 6.30) |
| | 4 | 26.61 | 0.9972, ddq (2.21, −14.81, 7.10) |
| | | | 1.2908, ddq (7.00, −14.81, 7.74) |
| | 5 | 11.5 | 0.7600, d (6.30) |
| | 6 | 15.1 | 0.7600, dd (7.10, 7.74) |
| | 7 | 31.4 | 3.2355, s |
| L-T | 1 | | |
| | 2 | 53.39 | 5.1688, d (2.12) |
| | 3 | 69.82 | 5.7758, dq (2.12, 6.54) |
| | 4 | 16.76 | 1.3100, d (6.54) |
| *N*-Me-L-T | 1 | | |
| | 2 | 63.00 | 4.9994, d (3.63) |
| | 3 | 66.93 | 4.4466, dq (3.63, 7.13) |
| | 4 | 19.9 | 0.9100, d (7.13) |
| | 5 | 34.30 | 3.3255, s |
| L-V$^{2}$ | 1 | | |
| | 2 | 56.77 | 4.8639, d (9.22) |
| | 3 | 31.76 | 2.3657, dqq (9.22, 6.60, 6.50) |
| | 4 | 19.6 | 1.09, d (6.60) |
| | 5 | 19.7 | 0.98, d (6.50) |
| *N*-Me-L-L | 1 | | |
| | 2 | 55.59 | 5.1318, dd (7.43, 7.42) |
| | 3 | 39.37 | 1.2328, ddd (7.43, −14.05, 1.87) |
| | | | 1.4775, ddd (7.42, −14.05, 5.78) |
| | 4 | 25.64 | 0.9488, ddqq (1.87, 5.78, 6.64, 6.64) |
| | 5 | 21.67 | 0.1806, d (6.64) |
| | 6 | 23.47 | 0.3640, d (6.64) |
| | 7 | 31.63 | 3.2644, s |
| L-V$^{3}$ | 1 | | |
| | 2 | 55.34 | 4.6096, d (9.22) |
| | 3 | 32.88 | 2.0082, dqq (9.22, 6.41, 6.20) |
| | 4 | 19.7 | 0.92, d (6.41) |
| | 5 | 19.7 | 0.86, d (6.20) |
| *N*-Me-L-V | 1 | | |
| | 2 | 71.37 | 3.0759, d (7.60) |
| | 3 | 29.94 | 2.5760, dqq (7.60, 6.97, 6.51) |
| | 4 | 19.8 | 0.96, d (6.97) |
| | 5 | 22.0 | 1.10, d (6.51) |
| | 6 | 40.37 | 3.1360, s |
| *N*-Me-4-OMe-L-Y | 1 | | |
| | 2 | 71.11 | 4.1183, dd (11.17, 4.49) |
| | 3 | 26.97 | 3.5645, dd (11.17, −13.59) |
| | | | 3.6913, dd (4.49, −13.59) |
| | 4 | 112.47 | |
| | 5 | 124 | 6.7018, d (0.87) |
| | 6 | 140.04 | |
| | 7 | 106.01 | 6.9195, ddd (0.87, 8.23, 0.90) |
| | 8 | 123.3 | 6.9827, dd (8.23, 7.80) |
| | 9 | 99.77 | 6.4434, dd (0.90, 7.80) |
| | 10 | 155.23 | |
| | 11 | 118.36 | |
| | 12 | 55.59 | 3.8302, s |
| | 13 | 40.90 | 2.1577, s |
| L-I | 1 | | |
| | 2 | 58.28 | 4.6459, d (7.18) |
| | 3 | 40.40 | 1.9631, dddq (7.18, 11.65, 8.65, 6.80) |
| | 4 | 20.07 | 1.5771, ddq (11.64, −15.20, 6.51) |
| | | | 1.2221, ddq (8.65, −15.20, 6.81) |
| | 5 | 19.70 | 1.0300, d (6.80) |
| | 6 | 19.70 | 0.9500, dd (6.51, 6.81) |
| Ph-L-*threo*-S | 1 | | |
| | 2 | 59.78 | 4.9196, d (1.96) |
| | 3 | 72.97 | 5.3191, d (1.96) |
| | 4 | 142.68 | |
| | 5,9 | 127.06 | 7.2289, dddd (7.89, 1.29, 0.38, 1.86) |
| | 6,8 | 129.28 | 7.2492, dddd (7.89, 7.38, 1.98, 0.38) |
| | 7 | 128.25 | 7.1962, dd (1.29, 7.38) |
| L-V$^{5}$ | 1 | | |
| | 2 | 59.26 | 4.4086, d (8.95) |
| | 3 | 33.19 | 1.9674, dqq (8.95, 7.27, 6.83) |
| | 4 | 19.10 | 0.9300, d (7.27) |
| | 5 | 19.40 | 0.9200, d (6.83) |

[a] $^{13}$C NMR parameters were interpreted from HSQC and HMBC spectra. [b] Interpretation of $^{1}$H NMR data was done by means of $^{1}$H iterative full-spin analysis. However, due to high level of impurities in the sample, chemical shifts for methyl signals (0.84−1.10 ppm) were interpreted visually.

rarely used at present, as various 2D NMR experiments (COSY, TCOSY, HSQC, HMBC, etc.) provide the necessary information that could be derived more readily by other means. However, while often neglected, such information normally embedded in $^{1}$H spectra holds major value to support the molecular structure, as seen in the cases presented in this and earlier works.[21−23] For a structure elucidation process that depends on indirect evidence and deductive reasoning methods, such as those involving NMR experiments, it is critically important to interpret *all* of the available data exhaustively. Admittedly, visual observation is not sufficient to extract all information packed into a complex $^{1}$H NMR spectrum, and this is likely the reason that this has not been exploited more extensively. With the capability of interpreting $^{1}$H NMR spectra of peptides at an early stage by means of the HiFSA sequencing method, this information can provide both early clues and additional proof for a structure.

The 1D $^{1}$H NMR experiment is the most sensitive NMR experiment and allows the collection of spectra with excellent signal-to-noise ratio (SNR) within relatively short total acquisition times. As cryomagnetic NMR instrumentation has evolved, contemporary entry to moderately advanced level (400−600 MHz $^{1}$H) instruments can be considered "basic" NMR spectrometers. Such equipment is normally sufficient to acquire excellent-quality $^{1}$H NMR spectra of even sub-milligram amounts of samples. The large molecular weight (∼1600 amu) and numbers of nonexchangeable protons (∼126) of the investigated tridecapeptides required the initial


DOI: 10.1021/acs.jnatprod.7b00207
J. Nat. Prod. 2017, 80, 2630−2643

Table 5. $^{1}$H and $^{13}$C NMR Data of Homoecumicin-ile13 (6; $CD_3OD$, 700 MHz for $^{1}$H, 175 MHz for $^{13}$C)

| moiety and position | | $\delta_C$[a] | $\delta_H$, $J$ in Hz[b] |
|---|---|---|---|
| *N,N*-Me$_2$-V | 1 | | |
| | 2 | 75.61 | 2.6902, d (9.17) |
| | 3 | 28.68 | 2.0536, dqq (9.17, 6.59, 6.62) |
| | 4,5 | 42.26 | 2.3255, s |
| | 6 | 19.4 | 0.9800, d (6.59) |
| | 7 | 20.0 | 0.8500, d (6.62) |
| L-V$^{1}$ | 1 | | |
| | 2 | 55.91 | 4.6651, d (8.75) |
| | 3 | 31.62 | 2.0998, dqq (8.75, 6.78, 6.72) |
| | 4 | 19.6 | 1.0600, d (6.78) |
| | 5 | 19.5 | 1.0000, d (6.72) |
| *N*-Me-L-*allo*-L | 1 | | |
| | 2 | 61.76 | 4.9111, d (11.23) |
| | 3 | 34.33 | 1.9497, dddq (11.23, 0.46, 2.93, 6.63) |
| | 4 | 26.61 | 0.9550, ddq (0.46, −12.22, 7.29) |
| | | | 1.2968, ddq (2.93, −12.22, 7.63) |
| | 5 | 11.5 | 0.7600, d (6.63) |
| | 6 | 15.1 | 0.7500, dd (7.29, 7.63) |
| | 7 | 31.4 | 3.2340, s |
| L-T | 1 | | |
| | 2 | 53.39 | 5.1610, d (2.34) |
| | 3 | 69.82 | 5.7726, dq (2.34, 6.51) |
| | 4 | 16.76 | 1.3105, d (6.51) |
| *N*-Me-L-T | 1 | | |
| | 2 | 63.00 | 5.0081, d (3.73) |
| | 3 | 66.93 | 4.4466, dq (3.73, 6.47) |
| | 4 | 19.9 | 0.9100, d (6.47) |
| | 5 | 34.30 | 3.3291, s |
| L-V$^{2}$ | 1 | | |
| | 2 | 59.26 | 4.8399, d (8.91) |
| | 3 | 33.19 | 1.9647, dqq (8.91, 6.84, 6.41) |
| | 4 | 19.1 | 0.93, d (6.84) |
| | 5 | 19.4 | 0.92, d (6.41) |
| *N*-Me-L-L | 1 | | |
| | 2 | 55.59 | 5.1161, dd (6.24, 8.38) |
| | 3 | 39.37 | 1.2376, ddd (6.24, −13.30, 8.21) |
| | | | 1.4582, ddd (8.38, −13.30, 5.79) |
| | 4 | 25.64 | 0.9488, ddqq (8.21, 5.79, 6.55, 6.59) |
| | 5 | 21.67 | 0.1690, d (6.55) |
| | 6 | 23.47 | 0.3366, d (6.59) |
| | 7 | 31.63 | 3.2606, s |
| L-V$^{3}$ | 1 | | |
| | 2 | 55.34 | 4.6086, d (8.87) |
| | 3 | 32.88 | 2.0262, dqq (8.87, 6.55, 6.78) |
| | 4 | 19.7 | 0.92, d (6.55) |
| | 5 | 19.7 | 0.86, d (6.78) |
| *N*-Me-L-V | 1 | | |
| | 2 | 71.37 | 3.0749, d (7.62) |
| | 3 | 29.94 | 2.5843, dqq (7.62, 6.50, 6.78) |
| | 4 | 19.8 | 0.98, d (6.50) |
| | 5 | 22.0 | 1.10, d (6.78) |
| | 6 | 40.37 | 3.1455, s |
| *N*-Me-4-OMe-L-Y | 1 | | |
| | 2 | 71.11 | 4.1016, dd (11.16, 4.71) |
| | 3 | 26.97 | 3.5604, dd (11.16, −13.73) |
| | | | 3.6909, dd (4.71, −13.73) |
| | 4 | | |
| | 5 | 124 | 6.6984, d (0.48) |
| | 6 | | |
| | 7 | 106.01 | 6.9188, ddd (0.67, 8.18, 0.48) |
| | 8 | 123.3 | 6.9823, dd (8.18, 7.79) |
| | 9 | 99.77 | 6.4424, dd (0.67, 7.79) |
| | 10 | | |
| | 11 | | |
| | 12 | 55.59 | 3.8302, s |
| | 13 | 40.9 | 2.1577, s |
| L-V$^{4}$ | 1 | | |
| | 2 | 59.21 | 4.5249, d (7.71) |
| | 3 | 33.7 | 2.1996, dqq (7.71, 6.75, 6.81) |
| | 4 | 19.7 | 1.03, d (6.75) |
| | 5 | 19.7 | 0.99, d (6.81) |
| Ph-L-threo-S | 1 | | |
| | 2 | 59.78 | 4.8391, d (1.90) |
| | 3 | 72.97 | 5.3191, d (1.90) |
| | 4 | | |
| | 5,9 | 127.06 | 7.2383, dddd (7.65, 1.33, 0.78, 1.28) |
| | 6,8 | 129.28 | 7.2551, dddd (7.65, 0.78, 7.47, 1.70,) |
| | 7 | 128.25 | 7.2022, dd (1.28, 7.47) |
| L-I | 1 | | |
| | 2 | 59.65 | 4.5147, d (8.89) |
| | 3 | 39.33 | 1.7852, dddq (8.89, 6.69, 7.03, 7.03) |
| | 4 | 26.97 | 1.4004, ddq (6.69, −14.00, 7.03) |
| | | | 1.1204, ddq (7.03, −14.00, 7.03) |
| | 5 | 12.50 | 0.91, d (7.03) |
| | 6 | 15.69 | 0.91, dd (7.03, 7.03) |

[a] $^{13}$C NMR parameters were interpreted from HSQC and HMBC spectra. [b] Interpretation of $^{1}$H NMR data was done by $^{1}$H iterative full-spin analysis. However, due to high level of impurities in the sample, chemical shifts for methyl signals (0.84−1.10 ppm) were interpreted visually.

use of higher magnetic field strength (800−900 MHz) to resolve the resonances in the small 0−8 ppm window. However, once the HiFSA profile of a reference peptide has been established, HiFSA sequencing can be performed at lower magnetic field strengths, as shown here with 600 and 700 MHz data for **2**−**7**. Therefore, when 2D NMR spectra are not available, e.g., due to limited sample availability, a high-resolution and good SNR 1D $^{1}$H NMR spectrum can provide critical structure information and be generated even with a basic NMR spectrometer. Considering that the investigated molecules are located at the higher end of "small-molecule" peptides, molecules with <1000 amu are fully amenable to HiFSA sequencing when using entry-level magnetic field strengths (400−500 MHz).

Whereas 2D NMR spectra can be developed into powerful dereplication tools, such as using HSQC and HMBC in 2D barcoding for chemical identification of triterpenes,[24] the HiFSA fingerprint of a new peptide, which was generated as a byproduct of this HiFSA-based structure elucidation process, also bears the potential for the quantitation and dereplication of isolated products.[11,14,18] Importantly, when supported by independent evidence for amino acid sequence (e.g., HMBC, MS-MS, or X-ray crystallography), such HiFSA profiles also become reference points for the elucidation of analogues. The subtlety of the intramolecular substituent chemical shift effects in peptides, especially those containing aromatic residues, makes the s.c.s. topology of related molecules highly sensitive indicators of structural similarity, including sequence.



DOI: 10.1021/acs.jnatprod.7b00207
J. Nat. Prod. 2017, 80, 2630−2643

The present study of ecumicin (**1**) analogues demonstrates that HiFSA structural sequencing is suitable for cyclic peptides as large as 13 amino acids, representing the largest molecules for which full-spin analysis has been performed to date. In fact, HiFSA sequencing might be suitable for the study of larger peptides. Once the HiFSA fingerprints are fully understood, it may be possible to study the effect of interactions between small molecule(s) and large(r) peptides by observing the small molecule as a separate spin system attached to a larger peptide. Moreover, as the HiFSA structural sequencing method is established based on the fact that each amino acid contains an individual $^{1}$H spin system isolated by amide groups, its application could be extended to all oligomeric compounds that consist of chains of uncoupled monomers, such as oligosaccharides[14] and proanthocyanidins.[25]

## ■ EXPERIMENTAL SECTION

**General Experimental Procedures.** The HiFSA structural sequencing of the ecumicin analogues was carried out with the PERCH NMR software tools (PERCH Solutions Inc., Kuopio, Finland). The $^{1}$H, COSY, HSQC, and HMBC NMR spectra of **2** were measured with a Bruker AVANCE-600 NMR, and the spectra of **3** and **4** on a Bruker AVANCE-II-900 NMR spectrometer. Both spectrometers were equipped with 5 mm TCI triple resonance inverse detection cryoprobes and *z*-axis pulse field gradients. The $^{13}$C spectrum of **2** was acquired on a Bruker DPX-400 NMR spectrometer equipped with a 5 mm QNP (4-nucleus) probe. The $^{1}$H COSY, HSQC, and HMBC NMR spectra of **5**−**7** were acquired a Bruker AV-III 700 NMR spectrometer, equipped with a 1.7 mm TCI triple resonance inverse detection cryoprobe and a *z*-axis pulse field gradient. The IR spectra were recorded on a Nicolet 6700 FT-IR instrument. Exact mass-to charge ratios were determined on a Shimadzu IT-TOF instrument. The semipreparative HPLC was carried out on a Waters Deakta 600 system with a Waters 996 photodiode array detector, using a YMC ODS-AQ 2.5 $\mu$m, 10 × 250 mm RP column. UV spectra were extracted from respective HPLC data. The samples were weighed with a Mettler Toledo XS105 Dual Range analytical balance. A Pressure-Lok gas syringe (VICI Precision Sampling Inc., Baton Rouge, LA, USA) was used for NMR sample preparation.

**Isolation.** Ecumicin (**1**) and its analogues (**2**−**7**) were isolated from the methanolic mycelial extract of *Nonomuraea* sp. MJM5123 by a four-step, bioassay-guided fractionation scheme: vacuum liquid chromatography using RP-18 silica gel, Sephadex LH-20 size exclusion chromatography, high-speed countercurrent chromatography using hexanes, ethyl acetate, methanol, and water (3:7:5:5 v/v/v/v) as the solvent system, and semipreparative HPLC. The first three steps were carried out as previously reported.[15] A two-solvent linear or isocratic gradient was used. Solvent A was 1% formic acid in water, and solvent B was 1% formic acid in acetonitrile. The flow rate was kept at 3 mL/min. For the purification of **2**−**4**, the mobile phase was kept at 49% B for 2 min, then in 15 min, ramped up to 67% B, and washed with 95% B for 5 min, and finally the column was re-equilibrated with 49% B for 5 min. The retention times were 14.8, 15.2, and 15.9 min for **3**, **4**, and **2**, respectively. For the isolation and purification of **5**−**7**, the mobile phase was kept at 51% B for 30 min. The retention times were 20.3, 21.1, and 22.5 min for **5**, **6**, and **7**, respectively.

**NMR Experiments.** Samples of deoxyecumicin (**2**), norecumicin (**3**), and nordeoxyecumicin (**4**) were prepared by weighing 1.15 mg (±0.01 mg), 2.00 mg (±0.01 mg), and 1.10 mg (±0.01 mg) into microcentrifuge tubes, followed by addition of 200, 150, and 130 $\mu$L of methanol-$d_4$ (99.8%), respectively. The solutions were transferred to 3 mm NMR tubes. The final concentrations of the samples were 5.75, 13.3, and 8.46 mg/mL, respectively. The $^{1}$H NMR experiments were acquired at 298 K (25 °C) using the standard Bruker pulse sequence ("zg"), with qHNMR parameter optimization. The probes were frequency tuned and impedance matched prior to data collection. Chemical shifts ($\delta$) are expressed in ppm relative to the residual protonated solvent signal ($\delta$ 3.310 ppm), and coupling constants ($J$) are given in hertz (Hz).

Samples of homoecumicin-ile11 (**5**), homoecumicin-ile13 (**6**), and homoecumicin-ile6 (**7**) were prepared by weighing 0.41 mg (±0.01 mg), 0.21 mg (±0.01 mg), and 0.37 mg (±0.01 mg) into microcentrifuge tubes, respectively, followed by addition of 40 $\mu$L of methanol-$d_4$ (99.8%). The solutions were transferred to a 1.7 mm NMR tube. The final concentration of the samples was 10.25, 5.25, and 9.25 mg/mL, respectively. The $^{1}$H NMR experiments were acquired at 298 K (25 °C) using the standard Bruker pulse sequence ("zg"). The probes were frequency tuned and impedance matched prior to data collection. Chemical shifts ($\delta$) are expressed in ppm relative to the residual protonated solvent signal ($\delta$ 3.310 ppm), and coupling constants ($J$) are given in hertz (Hz).

**HiFSA Structural Sequencing.** The HiFSA structural sequencing of the ecumicin analogues involved four major steps:

Step 1: the HiFSA profile of ecumicin was used as a reference and fitted to the $^{1}$H NMR spectrum of an unknown analogue, generating the unknown's pseudo HiFSA profile. The resolution-enhanced $^{1}$H NMR spectrum of the unknown was imported into the PERCH shell as a JCAMP-DX file and subjected to baseline correction, peak picking, and integration. The HiFSA profile of **1** was loaded to the PERCHit as the starting parameters for the unknown. Using both the integral-transform (D) mode and the total-line-fitting (T) mode, PERCHit refined these parameters exhaustively to the lowest possible RMS value. The resulting HiFSA profile was used as the unknown's pseudo HiFSA profile. Major differences between this pseudo HiFSA profile and the experimental spectrum were identified by visual comparison, which would presumably represent the structural difference(s) between the reference molecule and the unknown.

Step 2: the pseudo HiFSA profile was opened as a text file, and all the lines describing one different amino acid were deleted, leaving a modified HiFSA profile that represented the common amino acids in the reference molecule and the unknown.

Step 3: comparison of the common amino acids' HiFSA profile and the experimental spectrum of the unknown compound led to the isolation of the subspectrum of the different amino acid that is only present in the analogue.

Step 4: the structure of this amino acid was solved based on its unique $^{1}$H NMR profile, and the presumed structure of the analogue was arrived at by locating this amino acid in place of the subtracted amino acid in **1**. The HiFSA profile, information on this amino acid, was then manually added to the HiFSA profile of the common amino acids. This modified HiFSA profile served as the starting parameters for PERCH, which were refined by PERCHit using both the integral-transform (D) and the total-line-fitting (T) modes, until excellent agreement between the observed and simulated spectra was achieved.

If necessary, steps 2 to 4 were repeated until the structures of all the different amino acids were solved.

**MICs versus *M. tb*.** All MICs were determined using the microplate Alamar Blue assay as described previously.[16]

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jnatprod.7b00207.

UV, COSY, HSQC, and HMBC spectra of **2**−**7**; IR spectra for **5**−**7**; as well as $^{13}$C and semiselective HMBC NMR spectra of **2**−**4**; for compound **7**, the $^{1}$H NMR spectrum, detailed structure elucidation procedure, and the NMR table; HiFSA profiles mentioned in the text (PDF)


DOI: 10.1021/acs.jnatprod.7b00207
J. Nat. Prod. 2017, 80, 2630−2643

## AUTHOR INFORMATION


### Corresponding Author

*Tel: +1 (312) 355-1949. Fax: +1 (312) 355-2693. E-mail: gfp@uic.edu.

### ORCID

James B. McAlpine: 0000-0003-1696-8828
José G. Napolitano: 0000-0003-1250-2262
David C. Lankin: 0000-0003-3166-5980
Charlotte Simmler: 0000-0002-6923-2630
Shao-Nong Chen: 0000-0003-0748-0863
Guido F. Pauli: 0000-0003-1022-4326


### Notes

The authors declare no competing financial interest.
The original $^{1}$NMR data (FIDs) are made available at http://dx.doi.org/10.7910/DVN/92D3C9.

## ACKNOWLEDGMENTS

The authors are grateful to Mr. M. Niemitz, Perch Solutions Limited, Kuopio, Finland, for helpful discussions regarding iterative spin analysis, and Dr. R. Phansalkar, Institute for Tuberculosis Research, UIC, for acquisition of the IR data. The construction of the UIC−CSB NMR facility and the purchase of the 600 and 900 MHz NMR spectrometers were funded by NIHGNS grant P41 GM068944 awarded to Dr. P. G. W. Gettins. We would also like to acknowledge Dr. B. Ramirez for his assistance and support of the CSB instrumentation. This work was supported in part by "Cooperative Research Program for Agriculture Science & Technology Development (PJ01133003)" Rural Development Administration, Republic of Korea.

## REFERENCES


(1) Fosgerau, K.; Hoffmann, T. *Drug Discovery Today* **2015**, *20* (1), 122−8.
(2) Kaspar, A. A.; Reichert, J. M. *Drug Discovery Today* **2013**, *18* (17−18), 807−17.
(3) Wuthrich, K. *Biopolymers* **1983**, *22* (1), 131−8.
(4) Norwood, T. J.; Crawford, D. A.; Steventon, M. E.; Driscoll, P. C.; Campbell, I. D. *Biochemistry* **1992**, *31* (27), 6285−90.
(5) Edwards, D. J.; Marquez, B. L.; Nogle, L. M.; McPhail, K.; Goeger, D. E.; Roberts, M. A.; Gerwick, W. H. *Chem. Biol.* **2004**, *11* (6), 817−33.
(6) Nolis, P.; Parella, T. *J. Magn. Reson.* **2005**, *176* (1), 15−26.
(7) Medina, R. A.; Goeger, D. E.; Hills, P.; Mooberry, S. L.; Huang, N.; Romero, L. I.; Ortega-Barria, E.; Gerwick, W. H.; McPhail, K. L. *J. Am. Chem. Soc.* **2008**, *130* (20), 6324−5.
(8) Leet, J. E.; Li, W.; Ax, H. A.; Matson, J. A.; Huang, S.; Huang, R.; Cantone, J. L.; Drexler, D.; Dalterio, R. A.; Lam, K. S. *J. Antibiot.* **2003**, *56* (3), 232−42.
(9) *Modern NMR Approaches to the Structure Elucidation of Natural Products: Volume 1: Instrumentation and Software*; 2016; Vol. *1*, pp P001−329.
(10) An, T.; Kumar, T. K.; Wang, M.; Liu, L.; Lay, J. O., Jr.; Liyanage, R.; Berry, J.; Gantar, M.; Marks, V.; Gawley, R. E.; Rein, K. S. *J. Nat. Prod.* **2007**, *70* (5), 730−5.
(11) Pauli, G. F.; Chen, S. N.; Lankin, D. C.; Bisson, J.; Case, R. J.; Chadwick, L. R.; Godecke, T.; Inui, T.; Krunic, A.; Jaki, B. U.; McAlpine, J. B.; Mo, S.; Napolitano, J. G.; Orjala, J.; Lehtivarjo, J.; Korhonen, S. P.; Niemitz, M. *J. Nat. Prod.* **2014**, *77* (6), 1473−87.
(12) Napolitano, J. G.; Godecke, T.; Rodriguez-Brasco, M. F.; Jaki, B. U.; Chen, S. N.; Lankin, D. C.; Pauli, G. F. *J. Nat. Prod.* **2012**, *75* (2), 238−48.
(13) Napolitano, J. G.; Lankin, D. C.; McAlpine, J. B.; Niemitz, M.; Korhonen, S. P.; Chen, S. N.; Pauli, G. F. *J. Org. Chem.* **2013**, *78* (19), 9963−8.
(14) Napolitano, J. G.; Simmler, C.; McAlpine, J. B.; Lankin, D. C.; Chen, S. N.; Pauli, G. F. *J. Nat. Prod.* **2015**, *78* (4), 658−65.
(15) Gao, W.; Kim, J. Y.; Chen, S. N.; Cho, S. H.; Choi, J.; Jaki, B. U.; Jin, Y. Y.; Lankin, D. C.; Lee, J. E.; Lee, S. Y.; McAlpine, J. B.; Napolitano, J. G.; Franzblau, S. G.; Suh, J. W.; Pauli, G. F. *Org. Lett.* **2014**, *16* (23), 6044−7.
(16) Gao, W.; Kim, J. Y.; Anderson, J. R.; Akopian, T.; Hong, S.; Jin, Y. Y.; Kandror, O.; Kim, J. W.; Lee, I. A.; Lee, S. Y.; McAlpine, J. B.; Mulugeta, S.; Sunoqrot, S.; Wang, Y.; Yang, S. H.; Yoon, T. M.; Goldberg, A. L.; Pauli, G. F.; Suh, J. W.; Franzblau, S. G.; Cho, S. *Antimicrob. Agents Chemother.* **2015**, *59* (2), 880−9.
(17) Gao, W.; Napolitano, J. G.; Lankin, D. C.; Kim, J. Y.; Jin, Y. Y.; Lee, H.; Suh, J. W.; Chen, S. N.; Pauli, G. F. *Magnetic Resonance in Chemistry*; MRC, 2016.
(18) Pauli, G. F.; Chen, S. N.; Simmler, C.; Lankin, D. C.; Godecke, T.; Jaki, B. U.; Friesen, J. B.; McAlpine, J. B.; Napolitano, J. G. *J. Med. Chem.* **2014**, *57* (22), 9220−31.
(19) Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. *J. Org. Chem.* **1997**, *62* (21), 7512−7515.
(20) Napolitano, J. G.; Lankin, D. C.; Graf, T. N.; Friesen, J. B.; Chen, S. N.; McAlpine, J. B.; Oberlies, N. H.; Pauli, G. F. *J. Org. Chem.* **2013**, *78* (7), 2827−39.
(21) Molina-Salinas, G. M.; Rivas-Galindo, V. M.; Said-Fernandez, S.; Lankin, D. C.; Munoz, M. A.; Joseph-Nathan, P.; Pauli, G. F.; Waksman, N. *J. Nat. Prod.* **2011**, *74* (9), 1842−50.
(22) Niemitz, M.; Laatikainen, R.; Chen, S. N.; Kleps, R.; Kozikowski, A. P.; Pauli, G. F. *Magn. Reson. Chem.* **2007**, *45* (10), 878−82.
(23) Pauli, G. F.; Niemitz, M.; Bisson, J.; Lodewyk, M. W.; Soldi, C.; Shaw, J. T.; Tantillo, D. J.; Saya, J. M.; Vos, K.; Kleinnijenhuis, R. A.; Hiemstra, H.; Chen, S. N.; McAlpine, J. B.; Lankin, D. C.; Friesen, J. B. *J. Org. Chem.* **2016**, *81* (3), 878−89.
(24) Qiu, F.; McAlpine, J. B.; Lankin, D. C.; Burton, I.; Karakach, T.; Chen, S. N.; Pauli, G. F. *Anal. Chem.* **2014**, *86* (8), 3964−72.
(25) Nam, J. W.; Phansalkar, R. S.; Lankin, D. C.; McAlpine, J. B.; Leme-Kraus, A. A.; Vidal, C. M.; Gan, L. S.; Bedran-Russo, A.; Chen, S. N.; Pauli, G. F. *J. Org. Chem.* **2017**, *82* (3), 1316−1329.




DOI: 10.1021/acs.jnatprod.7b00207
*J. Nat. Prod.* 2017, 80, 2630−2643