Pure & Appl.Chem., Vol.54, No.12, pp.2409-2424, 1982.
Printed in Great Britain.

0033-4545/82/122409-16$03.00/0
Pergamon Press Ltd.

BIOLOGICALLY ACTIVE PEPTIDES AND THEIR MASS SPECTRA

Kenneth L. Rinehart, Jr., J. Carter Cook, Jr., Ramesh C. Pandey, Larry A. Gaudioso, Hsi Meng, Michael L. Moore, James B. Gloer, George R. Wilson, Robb E. Gutowsky, Paul D. Zierath, Lois S. Shield

School of Chemical Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

Li H. Li, Harold E. Renis, J. Patrick McGovren

The Upjohn Company, Kalamazoo, Michigan 49001, USA

Peter G. Canonico

Department of Antiviral Studies, US Army Medical Research Institute of Infectious Diseases, Fort Detrick, Frederick, Maryland 21701, USA

Abstract. A general procedure, relying heavily on mass spectrometry, has been developed for assigning structures to polypeptides. In the current version, amino acids are identified by HRFABMS, GC/MS, and GC on chiral columns, and quantitated by GC. FAB mass spectrometry assigns the peptide's molecular weight and amino acid sequence, supported sometimes by selective hydrolysis to oligopeptides. The procedure has been applied to several pore-forming peptaibophol antibiotics, as well as to the didemnins, cyclic depsipeptides isolated from a marine tunicate. The didemnins are active antitumor agents and show promise in the treatment of viral infections, including Rift Valley fever.

INTRODUCTION

Peptides certainly rank among the most biologically active natural products, providing man with his most potent toxins as well as with powerful hormones and useful antibiotics. Their chemistry and biology provide the subject matter of regular conferences in the United States (Ref. 1) and elsewhere (Ref. 2). Our own active interest in peptides commenced with an attempt to employ mass spectrometry to the fullest extent in assigning structures to peptide antibiotics, but more recently has turned to a study of some particularly potent peptides derived from marine sources. This presentation will describe both efforts, but first the mass spectrometric study.

STUDY OF PEPTAIBOPHOL ANTIBIOTICS BY MASS SPECTROMETRY

With their assemblages of small units which can be cleaved selectively by chemical means between the carbonyl and amino groups, peptides have long provided tempting targets for study by mass spectrometry, since one needs only to deduce from their mass spectra the order in which the amino acids are linked. A number of chemical procedures have been applied to the problem of inducing preferential cleavage of this type at or near the amide bonds, including lithium aluminum hydride reduction to a CH_2N grouping (Ref. 3) and N-methylation to convert secondary amides to tertiary amides (Ref. 4). In our own mass spectrometric studies of peptides we sought to minimize the number of chemical steps involved and required that the structures of the compounds should not be previously known. In addition, we preferred to work with peptides which for one reason or another might not be amenable to the elegant and

---

Abbreviations used: Mass spectrometry, MS; high resolution, HR; field desorption, FD; electron impact or electron ionization, EI; field ionization, FI; fast atom bombardment, FAB; chemical ionization, CI; gas chromatography, GC; high performance liquid chromatography, HPLC; and the residues for α-aminoisobutyric acid, Aib; 4-hydroxyproline, Hyp; isovaline, α-amino-α-methylbutyric acid, Iva; proline, Pro; phenylalaninol, Phol; phenylalanine, Phe; glycine, Gly; leucine, Leu; glutamic acid, Glu; glutamine, Gln; glutamine or glutamic acid, Glx; ornithine, Orn; threonine, Thr; N-methylleucine, MeLeu; statine, γ-amino-β-hydroxyisocaprylic acid, Sta; N,O-dimethyltyrosine, Me_2Tyr; α-(α-hydroxyisovaleryl)propionic acid, Hip; acetyl, Ac; alanine, Ala; valine, Val; tryptophan, Trp.

well developed chemical procedures, such as the Edman degradation, for dealing with small amounts of proteins.

Antiamoebins, emerimicins, alamethicins--first general procedure

The first peptide chosen for study was the antibiotic antiamoebin (Ref. 5) and the procedure ultimately developed in our laboratory for dealing with antiamoebin and other peptides is outlined in Table 1 (Ref. 6,7). By employing this procedure, structure 1 was reported for antiamoebin I in 1977 (Ref. 6), 2 for antiamoebin II in 1978 (Ref. 8), 3 and 4 for emerimicins III and IV (Ref. 9) and 5 and 6 for alamethicins I and II (Ref. 10) in 1977. A recent modification of the procedure involved compilation of a set of methylene unit values (Ref. 11) on two GC column substrates which can identify all of the standard and many unusual amino acids by GC alone (Ref. 12).

TABLE 1. General procedure I

1. Total hydrolysis of the antibiotic to amino acids, identified by a) HRFDMS; b) derivatization and GC/HREIMS; and c) derivatization and GC retention times on a chiral column
2. Quantitation of the amino acids by GC and amino acid analyzer
3. Assignment of a molecular formula to the antibiotic by molar ratios of amino acids and by HRFDMS on the molecular ion
4. Assignment of a partial sequence to the antibiotic from fragment peaks in its HREI mass spectrum
5. Partial hydrolysis of the antibiotic to oligopeptides, identified by a) derivatization and GC/HREIMS and GC/FIMS; and b) HRFDMS
6. Overlapping of the oligopeptides from step 5, combined with HREIMS on the intact antibiotic (step 4), to assign the structure

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19

1 Ac-Phe-Aib-Aib-Aib-Iva-Gly-Leu-Aib-Aib-Hyp-Gln-Iva-Hyp-Aib-Pro-Phol

2 Ac-Phe-Aib-Aib-Aib-Iva-Gly-Leu-Aib-Aib-Hyp-Gln-Iva-Pro-Aib-Pro-Phol

3 Ac-Phe-Aib-Aib-Aib-Val-Gly-Leu-Aib-Aib-Hyp-Gln-Iva-Hyp-Ala-Phol

4 Ac-Phe-Aib-Aib-Aib-Val-Gly-Leu-Aib-Aib-Hyp-Gln-Iva-Hyp-Aib-Phol

5 Ac-Aib-Pro-Aib-Ala-Aib-Ala-Gln-Aib-Val-Aib-Gly-Leu-Aib-Pro-Val-Aib-Aib-Glu-Gln-Phol

6 Ac-Aib-Pro-Aib-Ala-Aib-Aib-Gln-Aib-Val-Aib-Gly-Leu-Aib-Pro-Val-Aib-Aib-Glu-Gln-Phol

One of the more unusual techniques employed in General Procedure I is HRFDMS, used extensively throughout but most impressively in assigning the molecular formulas. Antiamoebin I contains the units Aib, Iva, L-Gly, L-Leu, L-Pro, L-Hyp, L-Glu, L-Phe and L-Phol in the approximate molar ratios 6:2:1:1:1:2:1:1:1, but Aib is difficult to quantitate. Adding the other amino acids and the C- and N-terminal groups and allowing for 5-7 Aib units predicted a molecular weight for antiamoebin I of either 1584, 1669 or 1754, but FDMS gave an ion in the molecular ion region at m/z 1692 and a similar ion in the triacetate at m/z 1818 (1692 + 3 x 42). The riddle of the anomalous molecular weight was solved by showing that both the 1692 and 1818 ions contained one atom of sodium. When other alkali metal salts were added, the "molecular" ion shifted down by 16 amu with lithium chloride and up by 16 amu with potassium chloride (Fig. 1) (Ref. 13). The molecular formula assigned the triacetate was confirmed by an HRFDMS measurement (1818.9596; Δ 6.2 mmu) carried out on the sodiated molecular ion.

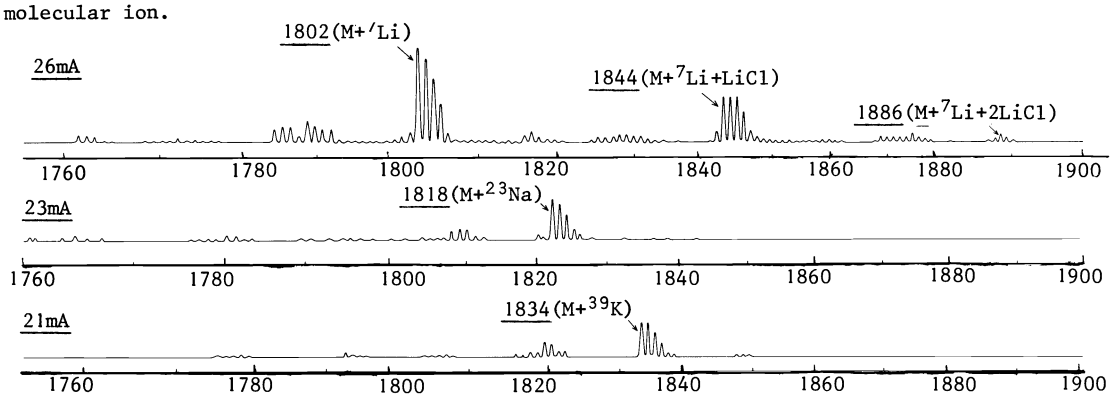
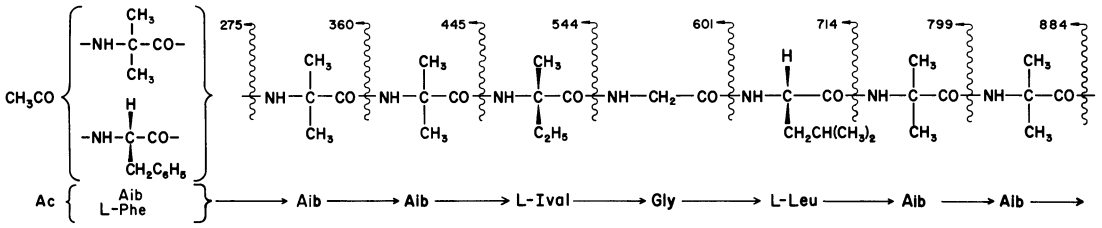


Fig. 1. Field desorption mass spectra of antiamoebin I triacetate in the molecular ion region: a, with added lithium chloride (26 mA); b, with added sodium chloride (23 mA); c, with added potassium chloride (21 mA). Peaks due to a small amount of a homologue, antiamoebin III, can be observed 14 amu below the major peaks.

In studying the sequence, the EI mass spectrum was investigated as far as ions could be observed. In contrast to many other peptides, antiamoebin showed rather clean CO-N fragmentation, allowing the partial sequence shown in Scheme 1 to be assigned, which was confirmed by HREIMS. Additional information was obtained from oligopeptides, with their spectra obtained by HRFDMS and their derivatives' by GC/HREIMS and GC/FIMS. In addition to derivatized oligopeptides obtained from the non-polar N-terminus (including Phe-Aib), oligopeptides identified from the polar C-terminus included Hyp → Gln → Iva, Glu $\xrightarrow{\alpha}$ Iva, Hyp → Aib → Pro and Pro → Phol. Overlapping these with the partial formula of Scheme 1 gave 1 as the structure of antiamoebin I.



Scheme 1.

The structure assignment rested nearly entirely on mass spectrometric data, with only two simple reactions being carried out on antiamoebin--total hydrolysis to the constituent amino acids and partial hydrolysis to a mixture of oligopeptides--plus derivatizations. It may be noted that to this point no products had been isolated.

In addition, a highly selective cleavage was effected by treatment of antiamoebin I with trifluoroacetic acid at room temperature (Scheme 2), which gave only three major products, each derived from selective cleavage between an Aib unit and the adjacent Hyp or Pro unit (Ref. 6). This appears to have been the first such selective use of trifluoroacetic acid. It is of some interest that the selectivity in the trifluoroacetolysis is mimicked by selectivity in the mass spectrometer, where preferential mass spectrometric cleavage between Aib and Hyp or Pro is also observed. Indeed, the last fragmentation seen in the EI mass spectrum of antiamoebin I (Scheme 1) is between Aib-9 and Pro-10 to give m/z 884.

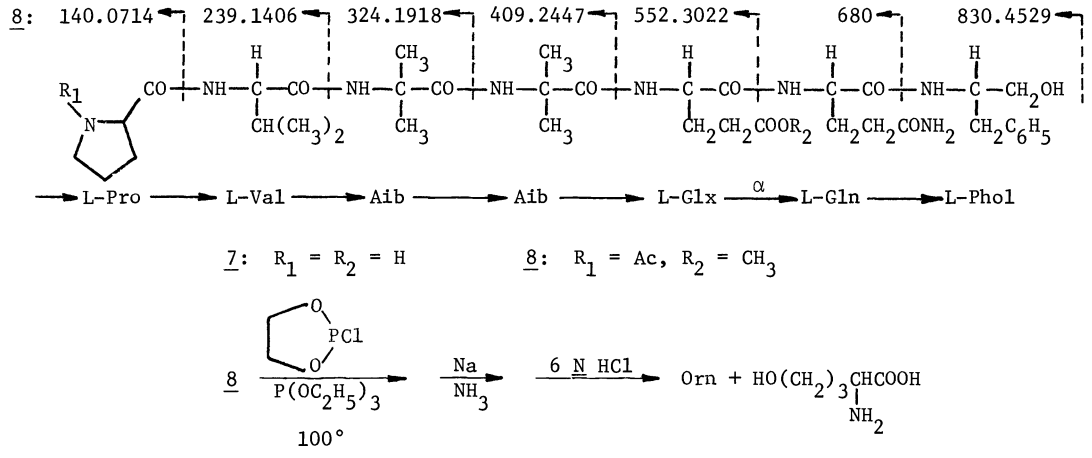
$$\text{Antiamoebin I} \xrightarrow[\text{r.t., 24 hr}]{CF_3COOH} \text{Hyp} \rightarrow \text{Gln} \rightarrow \text{Iva} + \text{Hyp} \rightarrow \text{Aib} + \text{Pro} \rightarrow \text{Phol}$$

$$+ \text{Ac} \rightarrow \text{Phe} \rightarrow \text{Aib} \rightarrow \text{Aib} \rightarrow \text{Aib} \rightarrow \text{Iva} \rightarrow \text{Gly} \rightarrow \text{Leu} \rightarrow \text{Aib} \rightarrow \text{Aib}$$

Scheme 2.

Precisely the same procedure outlined in Table 1 was employed in studying emerimicins III and IV (Ref. 9) and alamethicins I and II (Ref. 10). Alamethicin differs considerably from antiamoebin and emerimicin in its amino acid composition and its sequencing was somewhat more difficult than that of the antiamoebins or emerimicins. The sequence developed did not establish which of the Glx units were present as Gln and which as Glu, nor did it distinguish between α- and γ-linkages of the Glx units. To locate the single Glu residue, alamethicin was subjected to the trifluoroacetolysis technique developed previously for antiamoebin, giving as the major product the C-terminal "heptapeptide", 7. This was derivatized to 8 and subjected to HREIMS, which demonstrated that Gln was attached to Phol and that Glu was attached to Gln, as shown in Scheme 3. To distinguish between α- and γ-linkages, compound 7's methyl ester was dehydrated with diphenyl chlorophosphite and then reduced with sodium in liquid ammonia (Scheme 3) to convert carboxamide groups to primary amines and carbomethoxy groups to primary alcohols. Hydrolysis of 7 treated in this fashion identified Orn (derived from Gln) and 4-hydroxy-2-aminopentanoic acid (derived from Glu methyl ester). The structure of alamethicin I was assigned then as 5. The structure of alamethicin II (6) was deduced by noting the non-integral number of Ala residues and by observing homologous ions in the EI mass spectrum which pointed to substitution of Ala by Aib at amino acid 6. Syntheses of structure 5 have been carried out (Ref. 14,15) and biological properties of the natural product appear to match closely those of the synthetic material.

Since alamethicin had previously been recognized as having the property of forming pores in artificial membranes, facilitating transport of sodium and potassium ions, this pore-forming property was assayed for the antiamoebins and emerimicins and they were found to be active facilitators of ion transport (Ref. 16,17).



Scheme 3.

The first stage of our investigation of peptide structures by mass spectrometry was then completed with the assignment of structures to six peptaibophol antibiotics and the development of a methodology largely dependent on HRFDMS and GC/HREIMS.

Zervamicins--second general procedure

Investigations of the structures of zervamicins I and II (acidic and neutral, respectively) and emerimicin II (neutral) began almost as early as those of emerimicins III and IV, but these larger, more polar antibiotics proved considerably more difficult. Each contains one mole of tryptophan, a relatively fragile amino acid both hydrolytically and mass spectrometrically. Consequently, a number of the techniques developed for the other peptaibophols were unsuccessful with the zervamicins.

We have recently developed HPLC conditions for successfully separating complex mixtures of peptaibophol antibiotics, shown for zervamicin I and II in Fig. 2. The major component of zervamicin I is zervamicin IC, the major components of zervamicin II are zervamicins IIB and IIA, but there are numerous minor components, some of them separated only by recycling. In addition, emerimicin II was shown to consist of a mixture of components identical to those of zervamicin II, and the name emerimicin II has now been abandoned in favor of zervamicin II.

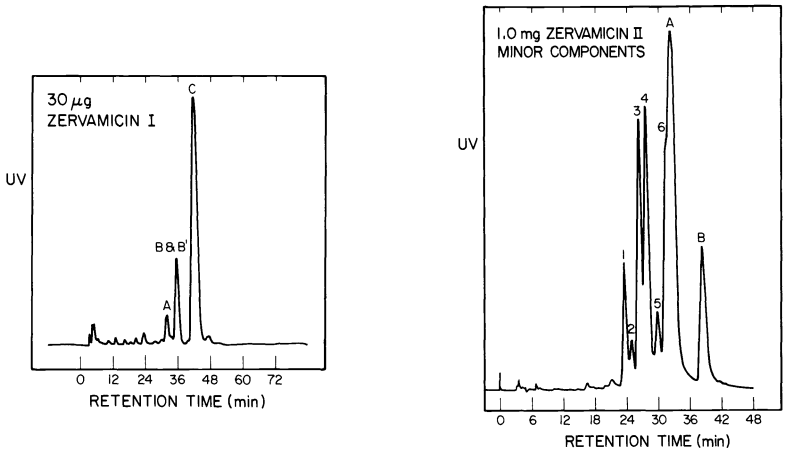


Fig. 2. HPLC separation of zervamicin components. Conditions: Waters Associates chromatograph, 215 nm UV detector, 25 cm x 10 mm Altex Ultrasphere-ODS column, methanol-water-2-propanol (50:34:16) mobile phase, 2.5 ml/min flow rate. For zervamicin I 0.1% of acetic acid was added to the solvent mixture.

The amino acid composition of zervamicin IC was successfully established by HRFDMS coupled with GC/MS, GC for quantitation, and GC on a chiral substrate for assignment of chirality, all of which argued the composition Aib:Iva:L-Ile:L-Leu:L-Pro:L-Thr:L-Hyp:L-Phol: L-Glx:L-Trp : 4:1:2:1:1:1:2:1:2:1. However, FDMS did not give reproducible molecular ions

on the antibiotics and EIMS gave only a few fragment ions, generally of insufficient intensity for high resolution measurements.

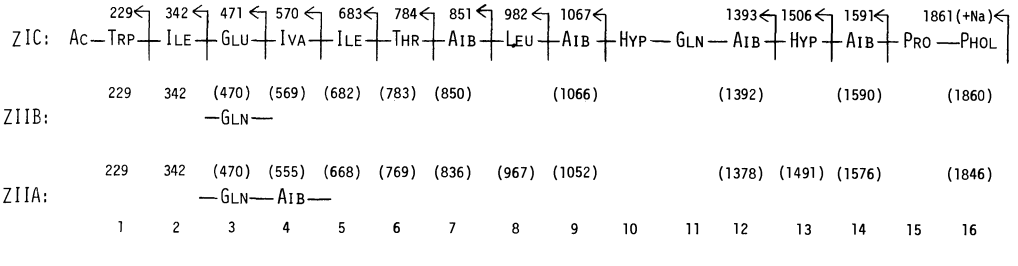
Introduction of fast atom bombardment mass spectrometry (Ref. 18,19) has altered our general procedure to that of Table 2, where it can be seen that one FAB measurement has essentially replaced two FD and one EI measurements.

TABLE 2. General procedure II

1. Total hydrolysis of the antibiotic to amino acids, identified by: a) fast atom bombardment mass spectrometry; b) derivatization and GC/HREIMS and GC/CIMS; c) derivatization and GC retention times on a chiral column
2. Quantitation of the amino acids by GC and amino acid analyzer
3. Assignment of a molecular formula to the antibiotic by molar ratios of amino acids and by HRFABMS on the molecular ion
4. Assignment of a partial sequence to the antibiotic from fragment peaks in its HRFAB mass spectrum
5. Partial hydrolysis of the antibiotic to oligopeptides, identified by: a) derivatization and GC/CIMS; and b) FABMS
6. Overlapping of the oligopeptides, combined with step 4, to assign the structure

Whereas FDMS had been only marginally successful in producing molecular ions from the zervamicins, requiring a special multichannel signal averager (Ref. 20), FABMS gave intense molecular ions, both in the positive ion and negative ion modes (Fig. 3). This was true not only of the major components, zervamicins IC, IIB, and IIA, but of all the minor components identifiable in Fig. 2. More importantly, FABMS has a major advantage in that it can give numerous interpretable fragment ion peaks as well as molecular ions. In the case of zervamicin IC, fragment ions for cleavage between most of the amino acids (CO-N cleavage) allowed a nearly complete assignment of its amino acid sequence, as shown in Scheme 4 (Ref. 21). Finally, in addition to molecular ion and fragment peaks, the FAB mass spectra of peptides contain a number of peaks at low masses characteristic of the individual amino acids present: for zervamicin IC at m/z 58 (Aib), 70 (Pro), 72 (Iva), 86 (Ile, Leu, Hyp), and 130 (Trp).

Remaining points of uncertainty involved the locations of Hyp, Leu and Ile. These were partially resolved by an HRFABMS measurement of the m/z 1067 peak (1067.6088), which demonstrated that Leu and Ile (but not Hyp) remain on the N-terminal side of this cleavage and both Hyp units on the C-terminal side. No fragment ions were detected resulting from cleavage between amino acids 10, 11 and 12, which must have been Hyp, Gln and Aib. In deciding the relative positions of these units (as well as placing Hyp and Pro at amino acids 13 and 15, respectively), a tentative conclusion was reached based on the selective cleavage between Aib and Hyp noted for other peptaibophols (Ref. 6). To confirm this tentative assignment, a careful study of the partial hydrolysis products was carried out on the closely related zervamicin IIB (see below). In this study the oligopeptides Hyp-Aib, Pro-Phol, and Hyp-Gln-Aib were identified by GC/MS, thus completing the assignment of the sequence of the amino acids in the C-terminal region and the overall assignment except for distinguishing between Leu and Ile at amino acids 2, 5 and 8. The major neutral components, zervamicins IIB and IIA, were sequenced to the same point by noting the replacement of Glu (in IC) by Gln (in IIB and IIA) and of an Iva (in IC) by Aib (in IIA), with appropriate mass shifts in their FAB mass spectra (Scheme 4).



Scheme 4.

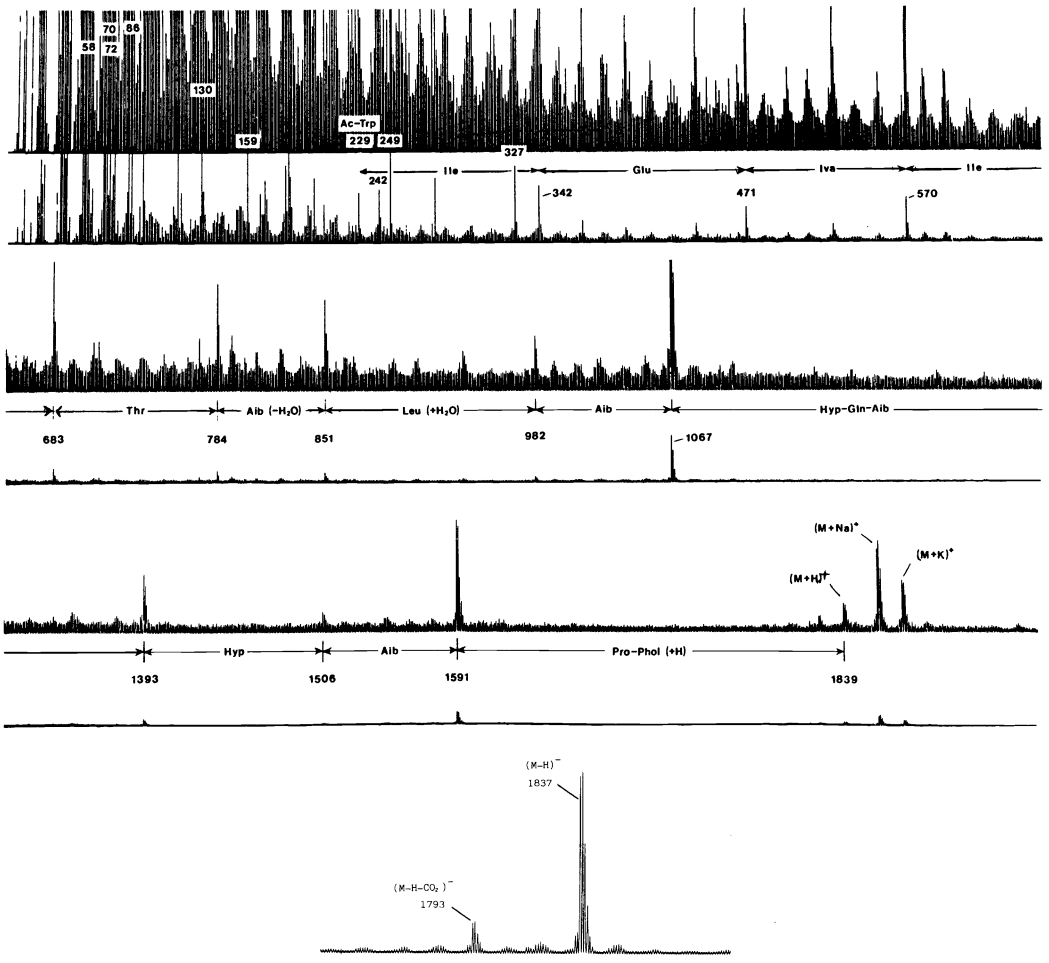


Fig. 3. FAB mass spectra of zervamicin IC: positive ion mode (upper) and negative ion mode (lower)

The locations of Leu and Ile were decided by drawing on the availability of the large number of components of the zervamicin complexes. Zervamicins IC, IIB and IIA, the three major components, all have two moles of Ile and one mole of Leu, whereas hydrolysis mixtures from a number of the minor components demonstrate that their amino acid compositions are identical to those of zervamicin IC or IIB or IIA, except in the replacement of one of the moles of Ile or Leu by a Val unit. FABMS fragmentation patterns of some of those minor components then serve to distinguish between Ile and Leu units, since wherever replacement of Ile (or Leu) by Val occurs, the appropriate mass shift is found, identifying the replaced unit as Ile or Leu. For example, zervamicins IA and IB contain one more Val and one fewer Ile than zervamicin IC. This replacement is clearly indicated by the fragmentation patterns of zervamicins IA and IB to be at amino acids 5 and 2, respectively. Therefore, the Ile's missing in zervamicin IC were at amino acids 5 and 2. Similar arguments can be made for the other replacements, e.g., of Leu by Val in zervamicins II-2 and II-5. By combining these replacement arguments, the locations of the Leu and Ile units in zervamicins IC, IIB and IIA were assigned, completing the structures of all the zervamicins as shown in Scheme 5 (Ref. 21).

The general procedure set forth in Table 2 is presently being investigated for peptides of other classes and should be of use in studying the structures of peptides with blocked N- and/or C-termini, as well as those difficult to cleave by enzymatic means due to the presence of unusual amino acids.

(ZIA): Ac-Trp-Ile-Glu-Iva-Val-Thr-Aib-Leu-Aib-Hyp-Gln-Aib-Hyp-Aib-Pro-Phol

(ZIB): Ac-Trp-Val-Glu-Iva-Ile-Thr-Aib-Leu-Aib-Hyp-Gln-Aib-Hyp-Aib-Pro-Phol

(ZIB'): Ac-Trp-Ile-Glu-Aib-Ile-Thr-Aib-Leu-Aib-Hyp-Gln-Aib-Hyp-Aib-Pro-Phol

(ZIC): Ac-Trp-Ile-Glu-Iva-Ile-Thr-Aib-Leu-Aib-Hyp-Gln-Aib-Hyp-Aib-Pro-Phol

(ZIIA): Ac-Trp-Ile-Gln-Aib-Ile-Thr-Aib-Leu-Aib-Hyp-Gln-Aib-Hyp-Aib-Pro-Phol

(ZIIB): Ac-Trp-Ile-Gln-Iva-Ile-Thr-Aib-Leu-Aib-Hyp-Gln-Aib-Hyp-Aib-Pro-Phol

(ZII-1): Ac-Trp-Ile-Gln-Aib-Val-Thr-Aib-Leu-Aib-Hyp-Gln-Aib-Hyp-Aib-Pro-Phol

(ZII-2): Ac-Trp-Ile-Gln-Aib-Ile-Thr-Aib-Val-Aib-Hyp-Gln-Aib-Hyp-Aib-Pro-Phol

(ZII-3): Ac-Trp-Val-Gln-Aib-Ile-Thr-Aib-Leu-Aib-Hyp-Gln-Aib-Hyp-Aib-Pro-Phol

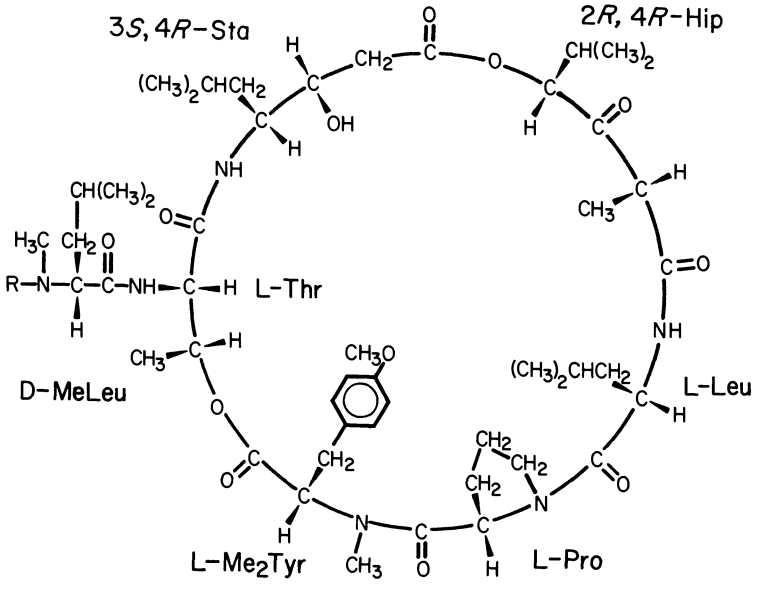
(ZII-4): Ac-Trp-Ile-Gln-Iva-Val-Thr-Aib-Leu-Aib-Hyp-Gln-Aib-Hyp-Aib-Pro-Phol

(ZII-5): Ac-Trp-Ile-Gln-Iva-Ile-Thr-Aib-Val-Aib-Hyp-Gln-Aib-Hyp-Aib-Pro-Phol

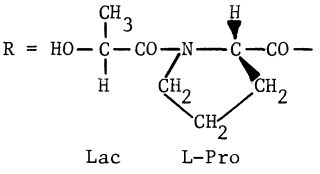
Scheme 5

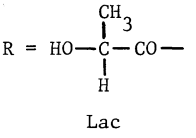
STUDY OF MARINE-DERIVED CYCLIC DEPSIPEPTIDES

We turn now to a rather different facet of our interest in peptides, to the isolation, structure proof and biological properties of the didemnins, a group of highly potent cyclic depsipeptides isolated from a Caribbean tunicate (also called an ascidian or sea squirt). The tunicate was originally demonstrated on shipboard to produce compounds having potent antiviral activity and cytotoxicity, as described elsewhere (Ref. 22). These compounds were subsequently isolated in the laboratory in Urbana and named didemnins after the tunicate, a Trididemnum species of the family Didemnidae. Preliminary reports of the structural assignments of didemnins A-C as 9-11 (here expanded to include known stereochemistry) (Ref. 23) and of their biological activities (Ref. 24) have already appeared. The structural assignments of the didemnins rest predominantly on mass spectrometric evidence and provide one of the best examples of the successful application of the methods developed for studying peptide structures by mass spectrometry.



(9) Didemnin A: R = H

(10) Didemnin B: R = 

(11) Didemnin C: R = 

Didemnin A

Although didemnin A is considerably less biologically active than didemnin B, it was chosen for the initial structural study because of its greater abundance and lower molecular weight. Didemnin A, suspected of being a peptide from its ^{1}H NMR spectrum, which contains NH protons near δ 8 ppm, was originally studied by the FD-based general procedure (Table 1). The mixture of amino acids obtained from total acid hydrolysis was converted to the amino acids' N-trifluoroacetyl n-butyl esters and these were analyzed by GC, GC/HREIMS and GC/CIMS and identified, in order of increasing retention time (Fig. 4), as derivatives of Thr, Leu, MeLeu, Sta, Pro, anhydrostatine (Sta-H_2O) and Me_2Tyr. The gas chromatogram indicated an approximately equimolar mixture of the six amino acids. Assignments of the structures of all of the compounds were confirmed by coinjection with authentic samples (Ref. 25) and statine was assigned by coinjection as the R,S (or S,R), erythro, isomer. The FD mass spectrum of the total acid hydrolyzate (Fig. 5) showed peaks at m/z 209, 175, 146, 132, 116 and 101 for the "molecular ions" (M, M + H or M - H_2O) of Me_2Tyr, Sta, MeLeu, Leu, Pro and Thr, plus a large peak at m/z 156.

The chirality of the amino acids was determined for Pro, Leu and Thr by resolution of their derivatives' racemic mixtures on a capillary column packed with a chiral substrate and comparison with the natural products' derivatives' retention times (Fig. 6), while the MeLeu and Me_2Tyr derivatives were not resolved, but were isolated by preparative GC and their rotations were compared with those of authentic samples. The stereochemistry at C-4 of statine was assigned as R by isolation of the anhydrostatine derivative (12) and comparison of its rotation with that obtained for the anhydrostatine derivative from 3S,4S-statine (Ref. 25).

Identification of the amino acids in didemnin A did not identify all the structural units. Addition of the unit weights for the six amino acids gave only 784 daltons, or the calculated precise mass 784.4891, while the molecular weight of didemnin A was determined to be 942.5678 by both HREIMS and HRFABMS. Thus, 156.0787 amu ($C_8H_{12}O_3$) remained. This piece of the molecule did not appear derivatized in the GC trace, but was observed in the FD mass spectrum of the mixture of amino acids as 156.0794 ($C_8H_{12}O_3$). Originally, the structure of the C_8 unit was assigned as -O-CH(iso-C_3H_7)COCH(CH_3)CO- by observation of its smaller sub-units in the ^{1}H and ^{13}C NMR spectra of didemnin A. Subsequently, the lactone of the hydroxy acid was isolated by extracting the aqueous hydrolyzate (amino acid mixture) with ether, and the ^{1}H NMR spectrum indicated cyclization to a lactone, 13.

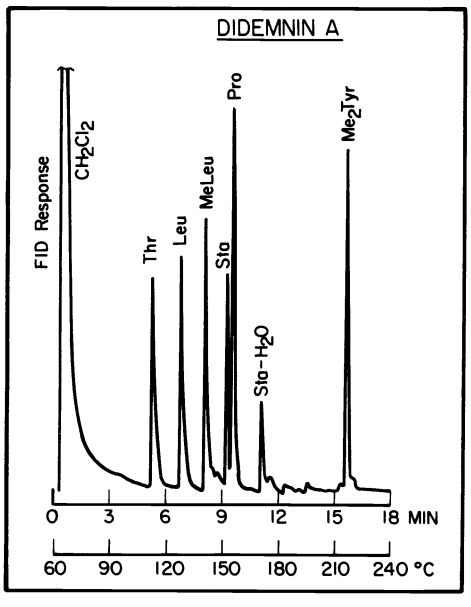


Fig. 4. GC trace of derivatives of amino acids from hydrolysis of didemnin A.

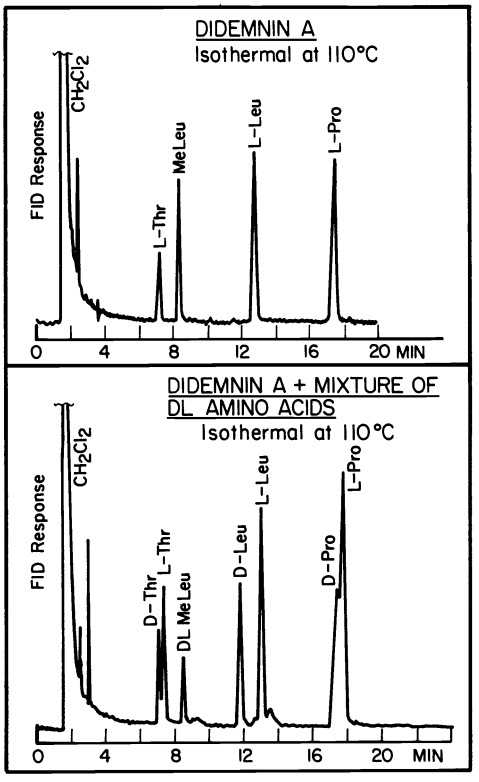


Fig. 6. GC traces of derivatives of amino acids on a chiral column: above, mixture of amino acids from didemnin A; below, same plus mixture of D,L-amino acids.

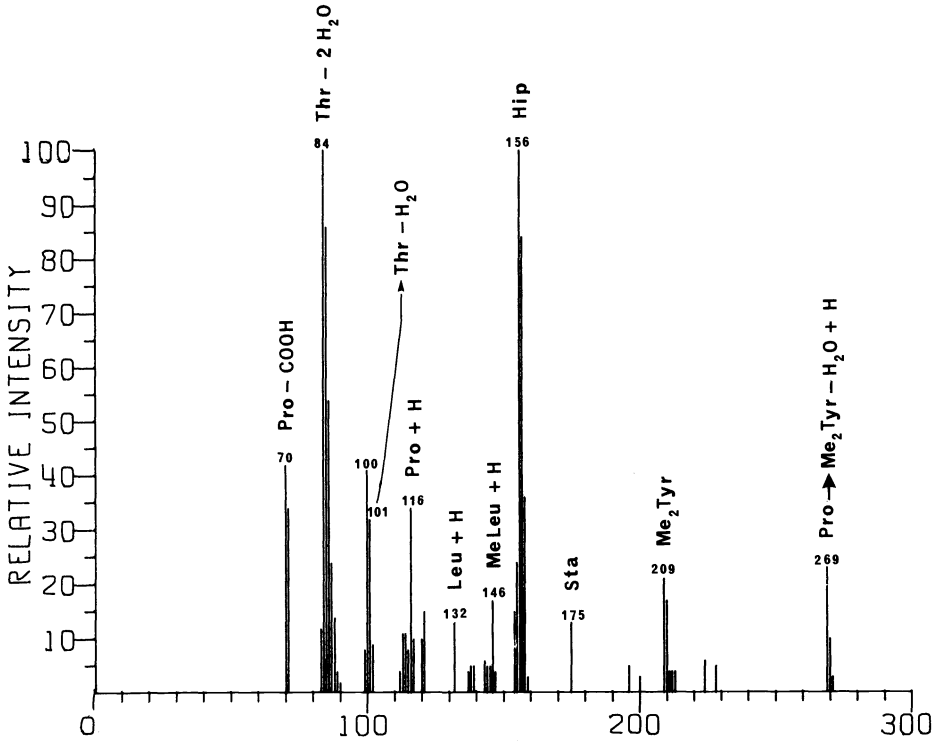
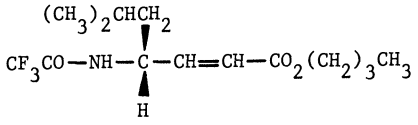
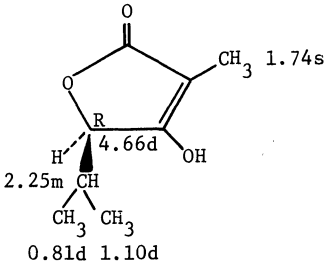


Fig. 5. FD mass spectrum of mixture of amino acids resulting from hydrolysis of didemnin A



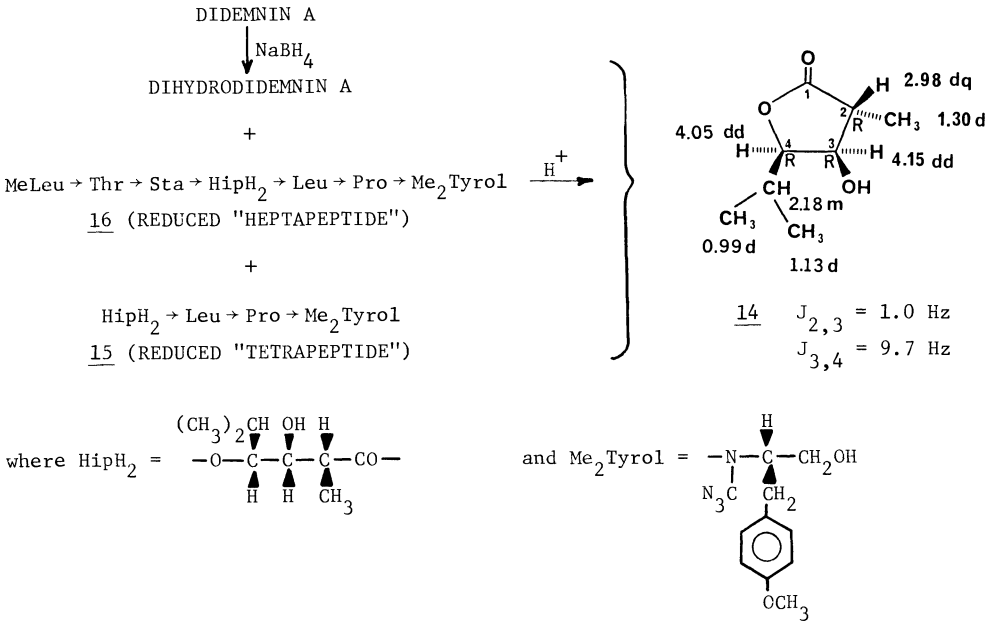
12



13

Absolute stereochemistry of the lactone was assigned as R at C-4 of 13 by comparison with known values for a similar tetronic acid with a 4-methyl instead of a 4-isopropyl group (Ref. 26). Relative (hence, absolute) stereochemistry at C-2 was derived from a reduction product obtained by reducing didemnin A with sodium borohydride (Scheme 6) to give a mixture of products in each of which the β-keto grouping of Hip had been reduced to a secondary alcohol. Hydrolysis of the mixture of borohydride-reduced products gave, among other compounds, the dihydrolactone 14, whose relative stereochemistry was assigned on the basis of ^1H-^1H coupling constants. While five-membered rings are flexible and the coupling constants for cis-trans pairs, in general, variable, an extremely small coupling constant ($J \leq 1$ Hz) can only be observed when hydrogens are trans on the five-membered ring (Ref. 27). The H-2, H-3 coupling constant observed for 14 argues for trans stereochemistry at those positions. On the other hand, $J_{3,4}$ was 9.7 Hz. Molecular models show that this coupling constant only agrees with H-3 and H-4 being cis. Thus, the complete stereochemistry of the dihydrolactone must be as shown for 14, which indicates that of Hip must be as shown in 9, completing the stereochemical assignment of the structural units of didemnin A.

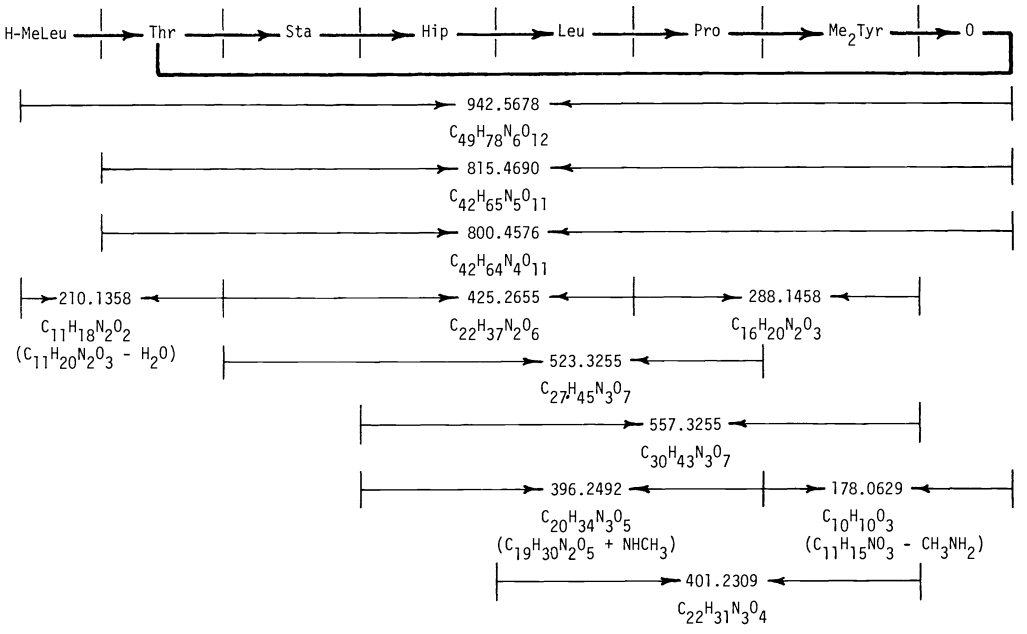
Completing the structural assignment of didemnin A required assigning the sequence of the six amino acids and the single keto acid present. The molecular formula of didemnin A is completed by the seven known structural units, allowing for no terminal hydroxyl group as in



Scheme 6.

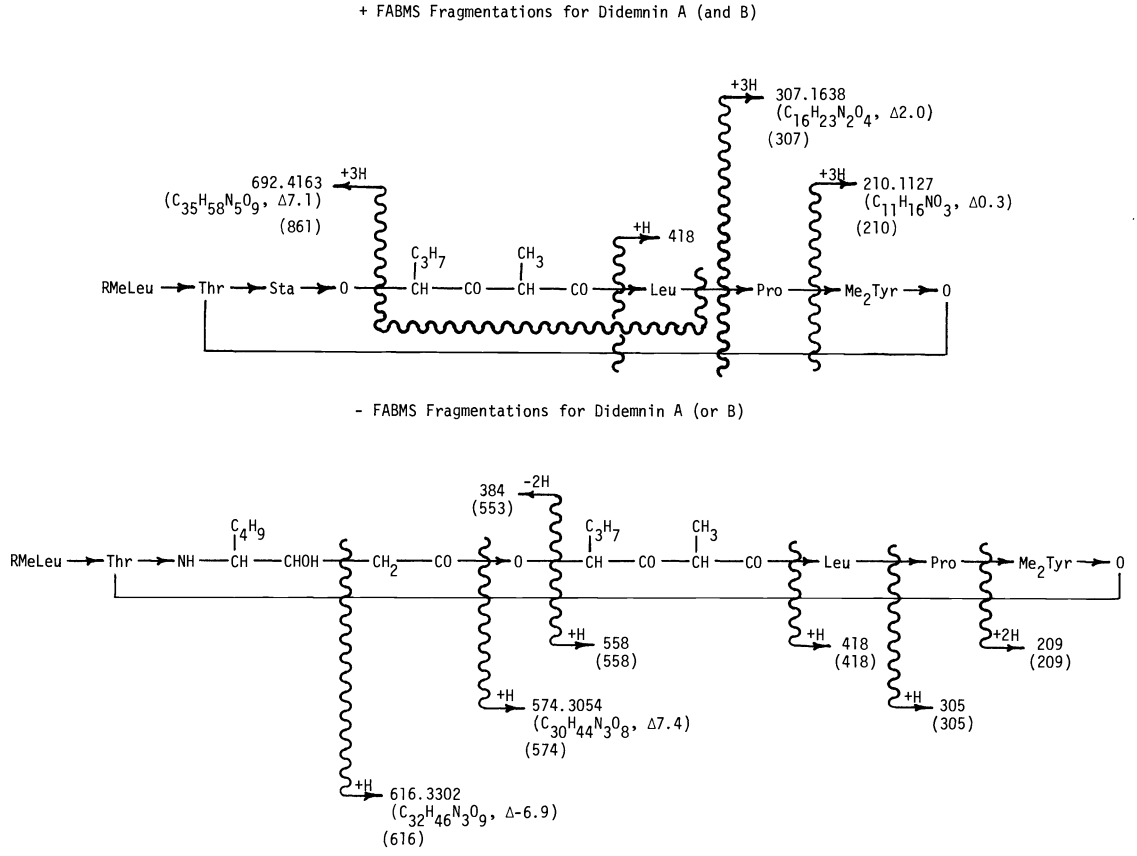
a linear peptide; thus, the peptide was deemed to be cyclic. From the ^1H NMR spectrum MeLeu was assigned as having a free N-CH_3 group (N-terminal amino acid) and statine as having a free hydroxyl group. Following the normal procedure for peptide structure assignment (Table 1), we turned to the HREI mass spectrum of didemnin A. Here, a key peak was at m/z 178 ($C_{10}H_{10}O_3$), which can only be due to a methoxycinnamic acid ion, assuring that Me_2Tyr is bound in an ester linkage to the hydroxyl group of Thr (since the hydroxyl group of Sta is free).

Beginning with this observation, the sequences Pro-Me_2Tyr, Leu-Pro-Me_2Tyr, Hip-Leu-Pro-Me_2Tyr, Sta-Hip-Leu-Pro and Thr-Sta-Hip-Leu-Pro-Me_2Tyr were assigned from the HREIMS peaks at m/z 288, 401, 557, 523, and 815, respectively (Scheme 7), thus assigning the complete structure of didemnin A.



Scheme 7.

The structure of didemnin A was assigned from these observations before the advent of FABMS (Ref. 22,23), but the FAB mass spectra have recently been studied and they confirm the sequence. The positive and negative ion FAB mass spectral fragmentations of didemnin A (Scheme 8) are extremely revealing, both in that they could de novo have assigned the sequence of didemnin A, and in the differences between the FAB fragmentations and the EI fragmentations in Scheme 7. It was noted above that positive and negative ion FAB spectra are often complementary and this is illustrated particularly well for didemnin A. Other major points of interest are cleavage at the nitrogen of Pro, a cleavage noted above to be significant in the assignment of structures to the peptaibophols, and an especially prominent elimination of the acyloxy group of Thr as Me_2TyrOH and related fragments.

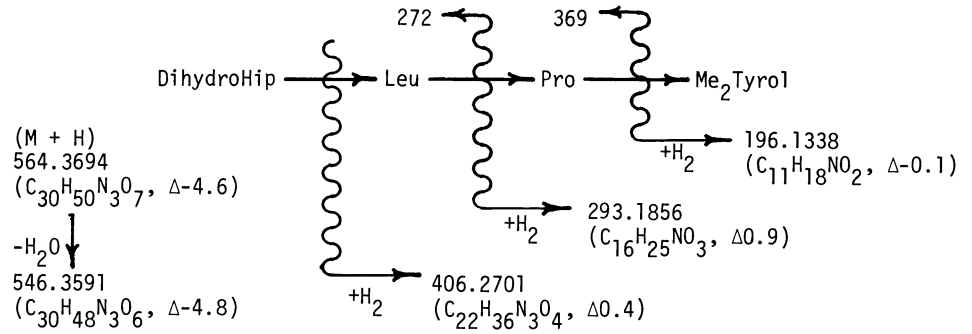


Scheme 8.

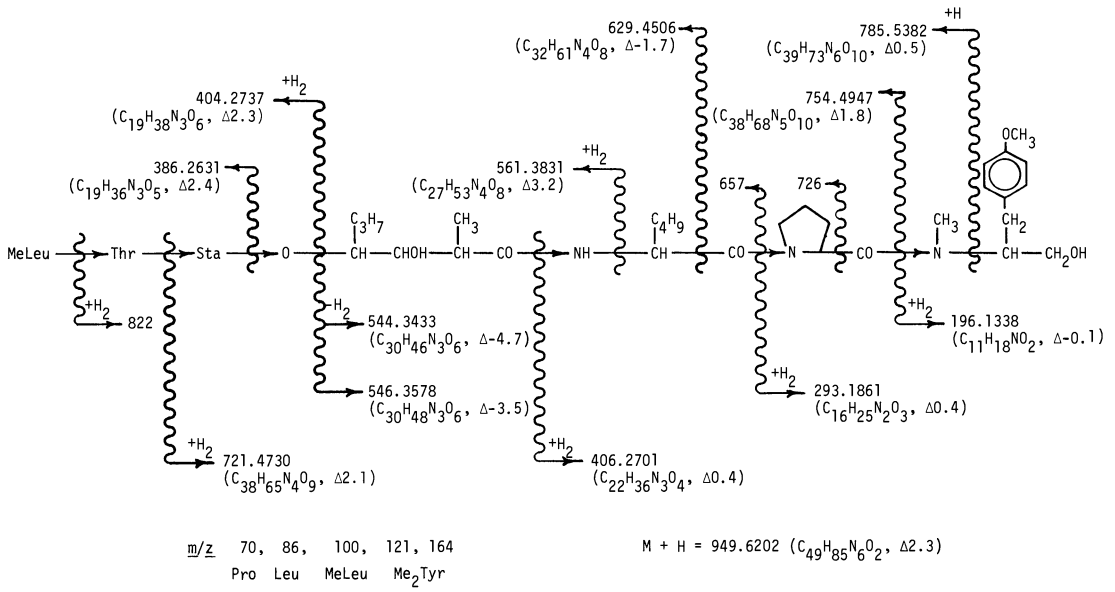
The sequences derived from EI or FAB mass spectra of didemnin A are based on spectra of cyclic depsipeptides, which have previously given some difficulty due to rearrangements. To confirm these assignments, didemnin A was converted to two linear peptides. The sodium borohydride reduction, which gave the dihydrolactone 14 after hydrolysis, yielded a number of primary products, as shown in Scheme 6. Among these, the "tetrapeptide" 15 was isolated in which Hip had been reduced and both of the lactone linkages had been cleaved by reduction. Thus, hydrolysis of 15 gave dihydrolactone 14 (cycloHipH_2), Leu, Pro and Me_2Tyrol. Since HipH_2 must have been N-terminal and Me_2Tyrol must have been C-terminal, only two structures are allowed. These are readily distinguished by the FAB fragmentation pattern of 15 (Scheme 9), which could, for that matter, have assigned the structure of the reduced tetrapeptide de novo even in the absence of knowledge of the N-terminal and C-terminal units. Significant low-mass peaks were also identified at m/z 70, 86, 121 and 164, characteristic of Pro, Leu and Me_2Tyrol (or Me_2Tyr).

In addition to 15, the exceedingly significant reduced "heptapeptide" 16 was isolated, in which, as usual, Hip had been reduced, but only a single lactone cleaved, giving an N-terminal MeLeu and a C-terminal Me_2Tyrol unit, as judged by hydrolysis of 16 and identification of the components by GC/MS. The amino acid sequence of 16, and thus of didemnin A, would have been totally derivable by interpretation of its FAB mass spectrum. The fragment peaks shown in Scheme 9 indicate ions are derived from both the C-terminal and N-terminal regions of the peptide. Even without the knowledge that MeLeu is N-terminal, since it is N-terminal in didemnin A, and Me_2Tyrol is C-terminal, the high resolution data completely define the structure of 16.

+ FABMS Fragmentations of Reduced "Tetrapeptide"



+ FABMS Fragmentations of Reduced "Heptapeptide"



Scheme 9.

Didemnin B

With the structure of didemnin A in hand, attention shifted to the structure of didemnin B, the more active component. Hydrolysis of didemnin B gave the same amino acids as didemnin A but with an extra mole of Pro, which was shown to be the L-isomer by chiral GC. The molecular formula of didemnin B ($C_{57}H_{89}N_7O_{15}$) was in accord with the presence of a $C_3H_4O_2$ unit in addition to the extra Pro unit. This was assigned as a lactic acid unit, Lac, which was not isolated but was identified in the ^1H NMR spectrum of didemnin B by its CH_3CHOH-group and in its ^{13}C NMR spectrum by an additional carboxyl carbon. The mass spectra of didemnin B (Scheme 8) are sufficiently similar to those of didemnin A to argue persuasively for a nearly identical structure. The methylamino group of MeLeu is acylated in didemnin B, as judged by the ^1H NMR spectrum, so the branching point must be MeLeu and, since both Pro units are acylated, structure 10 was assigned. In accord with the assignment, acetylation of didemnin B effects changes in the chemical shifts of the Lac unit but not the Pro unit.

Didemnin C

The third component of the didemnins, didemnin C, was shown to contain the same amino acids as didemnin A and in the same ratios, but didemnin C has an additional $C_3H_4O_2$ unit, which from the ^1H NMR spectrum must be attached to MeLeu. Again, as with didemnin B, this was assigned as a Lac unit although it was not isolated. Experiments to determine the stereochemistry of the Lac unit have not been conclusive.

Biological activities of the didemnins

As noted at the outset of this discussion of the didemnins, the extracts of the tunicate were initially noted on shipboard during the Alpha Helix Caribbean Expedition 1978 for their antiviral activity against Herpes simplex virus, type 1, as well as for their cytotoxicity. These activities were subsequently confirmed at The Upjohn Company, where the antiviral activity was expanded to include a number of DNA and RNA viruses and cytotoxicity was demonstrated for L1210 leukemia cells at a very low concentration (Ref. 22).

The L1210 cell studies revealed that didemnin B was the most active of the three (ID_{50} 0.0011 μg/ml), didemnin A the least active, and didemnin C of intermediate activity. In studies in mice, didemnin B was shown to extend the lifetime of mice infected with P388 leukemia by up to 100% and of mice infected with a melanoma by up to 60% (Table 3) (Ref. 28,29). These activities were confirmed by the National Cancer Institute and have resulted in a decision to test didemnin B in clinical trials. An adequate sample of didemnin B (15 grams) has now been isolated, and preliminary results should be available within the next year.

TABLE 3. Antitumor activities of didemnin B in vivo

Schedule of administration	Dose (mg/kg/ injection)	P388 leukemia T/C[a] (%)	P388 leukemia Body weight change (g)	B16 melanoma T/C[b] (%)	B16 melanoma Body weight change (g)
Days 1-9, i.p.	0.03			135	+0.8
	0.06	143	+0.1	144	-0.4
	0.25	138	-2.2	157	-3.3
	0.5	T[c]	-3.4	T	-4.5
	1	T	T		
Days 1,5,9, i.p.	0.03	125	+1.6		
	0.06	140	+1.2	139	+1.2
	0.25	157	-0.5	155	+0.3
	0.5	176	-1.7	146	-1.3
	1	199	-3.8	160	-3.2

[a]Median death of untreated animals inoculated (i.p.) with P388 leukemia (control) = 10.2 days. [b]Median death of untreated animals inoculated (i.p.) with B16 melanoma (control) = 19.4 days. [c]Toxic.

As an antiviral drug, didemnin B is exceedingly potent against both DNA and RNA viruses (Fig. 8). Here again, didemnin B is the most active; for example, among DNA viruses it reduces virus titers by about 2 logs for Herpes simplex virus, types 1 and 2, at 0.5 μM (Ref. 28,30). In vivo, didemnin B has been shown to protect mice with vaginal infections of Herpes simplex virus, type 2, by topical application (Table 4).

TABLE 4. Protection of female mice from genital HSV-2 infection by didemnins A and B[a]

Treatment	Drug concentration (mg/ml)	Death/ total	Survival (%)	Mean death (day)
Saline	0	13/14	7.1	6.5
DMSO (10%)	0	14/14	0	6.7
Didemnin A	1.0	6/14	57.1[b]	9.3[b]
	0.1	14/14	0	6.9
Didemnin B	0.23	4/14	71.4[b]	8.0[b]

[a]Mice infected with intravaginal inoculation of HSV-2 were treated intravaginally 3x/day x 3 days beginning 1 hr after inoculation with 0.1 ml of drug. Illnesses and deaths were recorded daily for 21 days. HSV-2 (strain 35D, 9.0 x 10^4 PFU/ 0.1 ml) was inoculated at T_0. [b]$p < 0.01$.

Studies at the University of Puerto Rico have recently shown didemnin A to reduce the titers of dengue virus (breakbone fever virus), a debilitating tropical DNA virus, at concentrations of 0.2-0.6 μg/ml (Ref. 31).

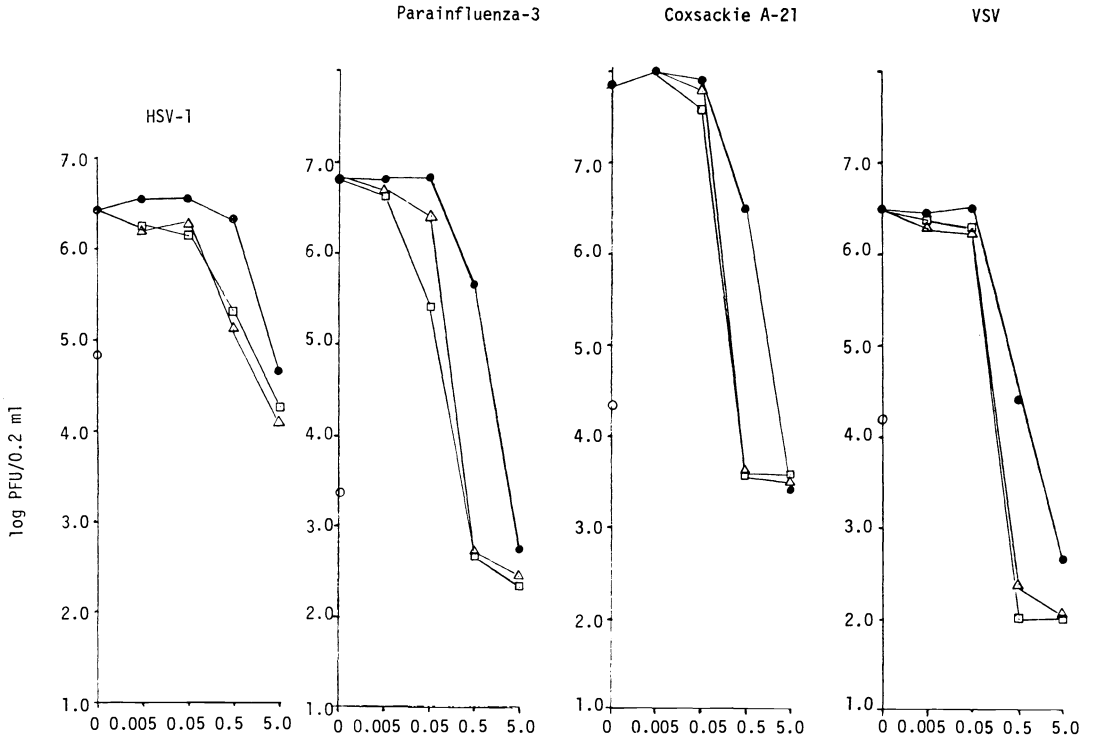


Fig. 8. Virus yields from infected cultures treated with different concentrations of didemnins A (●), B (□) and C (Δ). VSV = vesicular stomatitis virus.

Among the RNA viruses, the didemnins are active against influenza virus, para-influenza virus, Coxsackie virus and an equine rhinovirus, again at very low concentrations (Ref. 28, 30). Most recently, the didemnins have been demonstrated to have considerable activity against a number of quite potent RNA viruses including Venezuelan equine encephalomyelitis, yellow fever, sand fly fever, and, most impressively, Rift Valley fever virus and a Pichinde virus related to Lassa fever virus (Fig. 9) (Ref. 32). It is especially satisfying to report the last two results in a lecture in Africa, since Rift Valley fever and Lassa fever are East African and West African diseases, commonly fatal in man, and there are no effective treatments at the present time. With Rift Valley fever virus, mice were protected by didemnin B at relatively low concentrations (Table 5) and preparations are underway for carrying out tests of these drugs against primates.

Although DNA and RNA syntheses are also affected, the principal mode of action of the didemnins appears to be inhibition of protein synthesis, which is affected to a degree paralleling the *in vitro* activity with didemnin B being the most active of the group (Ref. 28-30).

Although the clinical utility of the didemnins is at present unclear, preliminary results from antitumor testing should be available within the next year. As with most potent compounds, extensive chemical manipulation will probably be required to optimize the therapeutic index. Regardless of their ultimate utility, however, there is no doubt that the didemnins provide yet another example of highly potent biologically active peptides and of the utility of mass spectrometry in elucidating their structures.

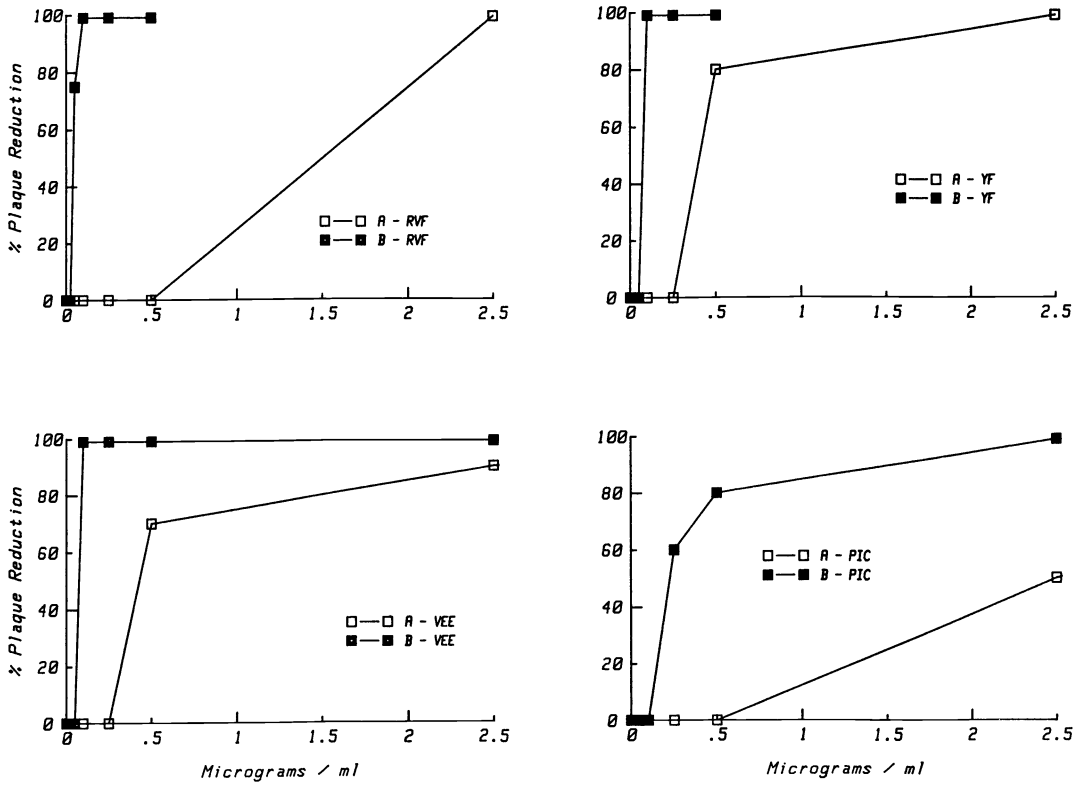


Fig. 9. Viral plaque reduction by didemnins A and B. RVF = Rift Valley fever virus; YF = yellow fever virus; VEE = Venezuelan equine encephalomyelitis; PIC = a Pichinde virus

TABLE 5. Survival of Rift Valley fever virus-infected mice treated with didemnins

	% Survival by day postinfection						
Drug dosage (mg/kg/day)	1	3	5	7	10	14	21
Didemnin A							
Virus inoculated							
0 (control)	100	90	10	0	0	0	0
1.25	100	90	70	50	50	50	50
5.0	100	100	90	60	50	50	50
10.0	100	100	60	10	0	0	0
Uninoculated (drug toxicity controls)							
1.25	100	100	100	100	100	100	100
10.0	100	100	100	100	100	100	100
Didemnin B							
Virus inoculated							
0 (control)	100	90	10	0	0	0	0
0.1	100	100	30	10	10	10	10
0.2	100	100	100	50	50	40	40
0.25	100	100	100	90	90	90	90
1.0	100	10	0	0	0	0	0
Uninoculated (drug toxicity controls)							
0.2	100	100	100	100	100	100	100
0.25	100	90	90	90	90	90	90
1.0	100	10	0	0	0	0	0

Acknowledgment -- Our efforts toward mass spectrometric solutions of peptide structure and identification of biologically active marine natural products have been supported by grants from the National Institute of Allergy and Infectious Diseases (AI 04769, AI 01278), the National Institute of General Medical Sciences (GM 27029, GM 16864), the National Science Foundation (PCM 77-12584), and by a contract with the National Cancer Institute (263-81-C-0449).

REFERENCES

1. K.L. Rinehart, Jr., M.L. Moore, L.A. Gaudioso, M. Barber, R.S. Bordoli, R.D. Sedgwick, A.N. Tyler and B.N. Green, in Peptides: Synthesis-Structure-Function, D.H. Rich and E. Gross, Eds., pp 757-760, Pierce Chemical Co., Rockford, IL (1981).
2. Peptides 1980, K. Brunfeldt, Ed., Scriptor, Copenhagen (1981).
3. K. Biemann, in Biochemical Applications of Mass Spectrometry, First Supplementary Volume, G.R. Waller and O.C. Dermer, Eds., pp 469-525, Wiley, New York (1980).
4. B.C. Das, S.D. Gero and E. Lederer, Biochem. Biophys. Res. Commun. 29, 211-215 (1967).
5. M.J. Thirumalachar, Hindustan Antibiot. Bull. 10, 287-289 (1968).
6. R.C. Pandey, H. Meng, J.C. Cook, Jr., and K.L. Rinehart, Jr., J. Am. Chem. Soc. 99, 5203-5205 (1977).
7. K.L. Rinehart, Jr., R.C. Pandey, M.L. Moore, S.R. Tarbox, C.R. Snelling, J.C. Cook, Jr., and R.H. Milberg, in Peptides: Structure and Biological Function, E. Gross and J. Meienhofer, Eds., pp 59-71, Pierce Chemical Co., Rockford, IL (1979).
8. R.C. Pandey, J.C. Cook, Jr., and K.L. Rinehart, Jr., J. Antibiot. 31, 241-243 (1978).
9. R.C. Pandey, J.C. Cook, Jr., and K.L. Rinehart, Jr., J. Am. Chem. Soc. 99, 5205-5206 (1977).
10. R.C. Pandey, J.C. Cook, Jr., and K.L. Rinehart, Jr., J. Am. Chem. Soc. 99, 8469-8483 (1977).
11. E. Kováts, Helv. Chim. Acta 41, 1915-1932 (1958).
12. S.J. Thomas, B.Sc. Thesis, University of Illinois, Urbana (1982).
13. K.L. Rinehart, Jr., J.C. Cook, Jr., H. Meng, K.L. Olson and R.C. Pandey, Nature 269, 832-833 (1977).
14. B.F. Gisin, D.G. Davis, A.K. Borowska, J.E. Hall and S. Kobayashi, J. Am. Chem. Soc. 103, 6373-6377 (1981).
15. G.R. Marshall and T.M. Balasubramanian, in Ref. 7, pp 639-646.
16. P. Mueller, Eastern Pennsylvania Psychiatric Institute, Philadelphia, PA, personal communication to K.L. Rinehart, Jr.
17. M. Eisenberg, State University of NY at Stony Brook, Long Island, NY 11794, personal communication to K.L. Rinehart, Jr.
18. M. Barber, R.S. Bordoli, R.D. Sedgwick and A.N. Tyler, J. Chem. Soc., Chem. Commun., 325-327 (1981).
19. D.J. Surman and J.C. Vickerman, J. Chem. Soc., Chem. Commun., 324-325 (1981).
20. C.R. Snelling, Jr., J.C. Cook, Jr., R.M. Milberg and K.L. Rinehart, Jr., Abstracts, 29th Ann. Conf. Mass Spectrom. Allied Topics, Minneapolis, MN, May 24-29, 1981, p. 602.
21. K.L. Rinehart, Jr., L.A. Gaudioso, M.L. Moore, R.C. Pandey, J.C. Cook, Jr., M. Barber, R.D. Sedgwick, R.S. Bordoli, A.N. Tyler and B.N. Green, J. Am. Chem. Soc. 103, 6517-6520 (1981).
22. K.L. Rinehart, Jr., P.D. Shaw, L.S. Shield, J.B. Gloer, G.C. Harbour, M.E.S. Koker, D. Samain, R.E. Schwartz, A.A. Tymiak, D.L. Weller, G.T. Carter, M.H.G. Munro, R.G. Hughes, Jr., H.E. Renis, E.B. Swynenberg, D.A. Stringfellow, J.J. Vavra, J.H. Coats, G.E. Zurenko, S.L. Kuentzel, L.H. Li, G.J. Bakus, R.C. Brusca, L.L. Craft, D.N. Young and J.L. Connor, Pure Appl. Chem. 53, 795-817 (1981).
23. K.L. Rinehart, Jr., J.B. Gloer, J.C. Cook, Jr., S.A. Mizsak and T.A. Scahill, J. Am. Chem. Soc. 103, 1857-1859 (1981).
24. K.L. Rinehart, Jr., J.B. Gloer, R.G. Hughes, Jr., H.E. Renis, J.P. McGovren, E.B. Swynenberg, D.A. Stringfellow, S.L. Kuentzel and L.H. Li, Science 212, 933-935 (1981).
25. Samples of Me_2Tyr and statine--erythro (3R,4S) and threo (3S,4S) isomers--were obtained from Drs. R.E. Moore, University of Hawaii and D.H. Rich, University of Wisconsin, respectively.
26. P.W. Clutterbuck, H. Raistrick and F. Reuter, Biochem. J. 29, 1300-1309 (1935).
27. K.L. Rinehart, Jr., W.S. Chilton, M. Hichens and W. von Phillipsborn, J. Am. Chem. Soc. 84, 3216-3217 (1962).
28. K.L. Rinehart, Jr., J.B. Gloer, G.R. Wilson, R.G. Hughes, Jr., L.H. Li, H.E. Renis and J.P. McGovren, Fed. Proc., in press.
29. L.H. Li, H.E. Renis, J.P. McGovren and K.L. Rinehart, Jr., Proc. Am. Assoc. Cancer Res., 22, 255 (1981).
30. H.E. Renis, B.A. Court, E.E. Eidson, E.B. Swynenberg, J.B. Gloer and K.L. Rinehart, Jr., Abstracts, 21st Intersci. Conf. Antimicrob. Agents Chemother., Chicago, IL, November 4-6, 1981, No. 189.
31. E. Maldonado, J.A. Lavergne and E. Kraiselburd, Puerto Rico Health Sci. J., in press.
32. P.G. Canonico, W.L. Pannier, J.W. Huggins and K.L. Rinehart, Jr., Antimicrob. Agents Chemother., submitted.