TILDEN LECTURE

Structural Studies on Bio-active Molecules

By Dudley H. Williams
UNIVERSITY CHEMICAL LABORATORY, LENSFIELD ROAD,
CAMBRIDGE, CB2 1EW

1 Introduction

The structure elucidation of organic molecules can be regarded as falling into three areas. One area is that of *X*-ray crystallography; this approach requires single crystals, and therefore quantities of the substance of the order of mg or greater. It is also required that the crystals diffract *X*-rays satisfactorily, that they do not rapidly deteriorate in the *X*-ray beam, and further that the structure can be solved from the acquired data. In the absence of a heavy atom derivative, this last problem can be formidable if the asymmetric unit contains more than around 100 atoms.

A second area is that of solution methods to determine the structures of the standard biopolymers, particularly DNA, RNA, and proteins (or peptides). Astonishingly sensitive and rapid methods have been developed to determine the sequence of bases (in DNA or RNA) or of amino acids (in peptides and proteins). These methods determine only the primary structure (sequence), although it is true that in a limited number of cases, secondary and tertiary structure (*i.e.* the whole three-dimensional picture) has been determined by the powerful method of *X*-ray crystallography.

The third area has to attempt to encompass all the remaining problems. It includes all compounds that are not standard biopolymers and, additionally, suitable crystals cannot be obtained. Thus, we must include not only those compounds which, no matter how much is available, refuse to crystallize, but also those cases where crystallization is effectively impossible because of the small amount of substance available. This area will include an enormous number of physiologically important compounds such as hormones, vitamins, neuropeptides, toxins, and antibiotics (and their metabolites). Whereas, two decades ago, such problems had to be solved by tedious and demanding chemical degradations, recently a number of relatively rapid and powerful methods for their solution have evolved. Most notably, these methods include mass spectrometry and nuclear magnetic resonance. It is with this area of structure elucidation that the present article is concerned.

The techniques will be illustrated by the solution of four quite different kinds of structures, each typifying a characteristic problem.

2 The Metabolite Problem: Structure Elucidation of 1α,25-dihydroxyvitamin D_3

The problem of structure determination of metabolites is often characterized by two features: the amount of metabolite available is small ($< 100\,\mu g$), and the

metabolite is related to the known structure of the starting material by a relatively simple chemical change *e.g.* hydroxylation, demethylation, sulphation. Such problems are ideally suited to study by mass spectrometry since mass spectrometry is a sensitive technique (molecular weight determination in the μg to ng range). In principle, ^{1}H n.m.r. can be usefully applied to samples as small as 10 μg, but in practice such samples are frequently contaminated with other materials and satisfactory spectra may not be obtained. In cases such as the latter, mass spectrometry has the further advantage that a temperature gradient can be applied to the sample when it has been loaded onto the mass spectrometer probe. In this way, impurities can often be fractionated from the substance of interest, and a good spectrum obtained of a single component.

By 1970, it had been established that vitamin D_3 (cholecalciferol) undergoes several metabolic transformations before it can affect calcium metabolism in its target tissues such as intestine and bone.[1-3] In particular, a metabolite is produced by the kidney which seemed to be obligatory for vitamin D activity. Work by the group of Dr. E. Kodicek,[4] at the Dunn Nutritional Laboratory, Cambridge, had shown that following intravascular injection, [4-^{14}C,1-^{3}H] cholecalciferol was converted, in the kidney, into a more polar metabolite which had lost ^{3}H. Furthermore, [4-^{14}C,1-^{3}H]-25-hydroxycholecalciferol (which is produced by 25-hydroxylation of cholecalciferol in liver) is converted by kidney into the same polar metabolite (again with concomitant loss of ^{3}H).

The identification of the polar metabolite from the small amount found in animal tissues (< 3 ng/g) would have required extraction of several kg of tissue. However, the synthesis of the metabolite by homogenates of kidney provided a method which allowed isolation of 63 μg of the metabolite.[5] The molecular ions of the metabolite and its TMS ether at *m/z* 416 and 632, respectively, established that two additional oxygen atoms had been inserted in the cholecalciferol molecule in the form of two hydroxy groups. Cleavage of the C(24)—C(25) bond gave intense peaks at *m/z* 59 and 131 in the spectra of the free steroid and the TMS ether respectively (Figure 1), confirming the presence of the hydroxy group at C(25). The spectrum of the free steroid has peaks at 287 and 269 corresponding to loss of the side chain, followed by the elimination of water. This indicates that the third hydroxy group [like that at C(3), originally present in cholecalciferol itself] is in the steroid nucleus. The mass spectrum of cholecalciferol, 25-hydroxycholecalciferol, and indeed all molecules so far examined with the vitamin D conjugated triene system, have intense peaks at *m/z* 136 and 118 (*m/z* 136 − H_2O), and in the case of their TMS ethers at *m/z* 208 and 118 (*m/z* 208 − trimethylsilanol). The peaks at *m/z* 136 and 208 have been attributed to the fragment formed by the A ring plus

[1] J. W. Blunt, H. F. DeLuca, and H. K. Schnoes, *Biochemistry*, 1968, **7**, 3317.
[2] M. H. Haussler, J. F. Myrtle, and A. W. Norman, *J. Biol. Chem.*, 1968, **243**, 4055.
[3] D. E. M. Lawson, P. W. Wilson, and E. Kodicek, *Nature*, 1969, **222**, 171.
[4] D. R. Fraser and E. Kodicek, *Nature*, 1970, **228**, 764.
[5] D. E. M. Lawson, D. R. Fraser, E. Kodicek, H. R. Morris, and D. H. Williams, *Nature*, 1971, **230**, 228.

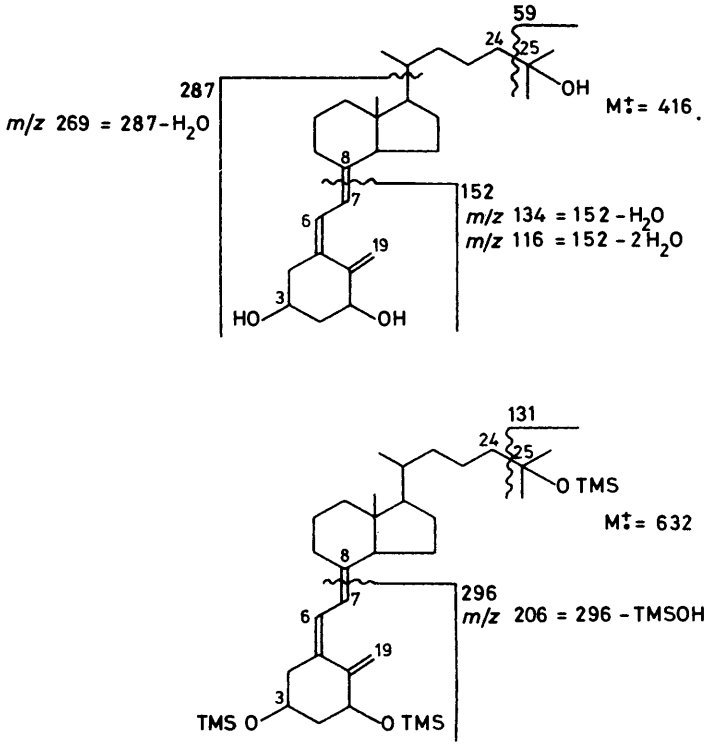


Figure 1 *Fragmentation of* 1,25-*dihydroxycholecalciferol and its tri-TMS derivative*

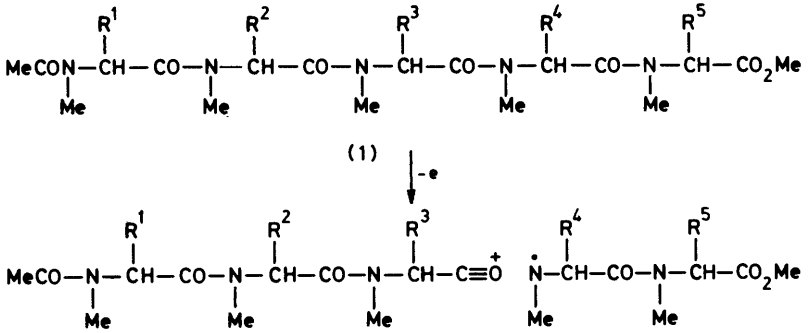
C(6), C(7), and C(19) [cleavage of the C(7)—(8) bond].[6] The mass spectrum of the metabolite had intense peaks at *m/z* 134 and 116, and that of the TMS ether derivative at *m/z* 206 (Figure 1). The mass spectrum of a mixture of the TMS ether and the deuteriated TMS ether of the metabolite had intense peaks at *m/z* 206 and 215 (^{2}H(6)-206), and supported the interpretation that the peak at *m/z* 206 was due to the A-ring plus C(6), C(7), and C(19). We interpreted the difference for the metabolite and cholecalciferol (or their TMS derivatives) of two mass units between *m/z* 134 and 136, or *m/z* 206 and 208, as being caused by the elimination of water or trimethylsilanol, respectively. This interpretation demands that the third hydroxy group is in the A-ring plus C(6), C(7), and C(19). Since the u.v. absorption of the metabolite and cholecalciferol are not only similar in shape and intensity but also in λ_{max} (269 nm *vs.* 266 nm), it is concluded that the vitamin D triene chromophore is present in both compounds. Therefore, hydroxylation in the kidney must have occurred at C(1), C(2), or C(4). Since the hydroxylation in kidney occurs with loss of tritium from C(1) (and, in particular, from C(1α) since the

[6] T. Suda, H. F. DeLuca, H. K. Schnoes, G. Ponchon, Y. Tanaka, and M. F. Holick, *Biochemistry*, 1970, **9**, 2917.

tritium had originally been introduced with 85% stereoselectivity into the 1α-position[7]), it was concluded that the metabolite was 1α,25-dihydroxycholecalciferol. Note that this conclusion calls on the fact that hydroxylation of hydrocarbon chains normally occurs *via* oxygen insertion into the appropriate C—H bond, *i.e.* it occurs with retention of configuration. The structure of the metabolite has subsequently been confirmed by synthesis. The metabolite is a hormone controlling calcium metabolism. It is of clinical importance, for example when renal insufficiency causes problems in calcium metabolism.

3 The Structure Elucidation of Unusual Linear Peptides

In the case of a structure which is not only unknown but also novel, it is rare that mass spectrometry alone will provide extensive structural information. This is because the fragmentation patterns of complex molecules are enormously varied, and are more often rationalized with the value of hindsight rather than being used to determine structure. A notable exception is the sequence determination of linear peptides.[8] It is established that such molecules, when converted into their *N*-acetylpermethyl derivatives (1), fragment at each amide bond in a relatively reproducible manner. This cleavage, occurring with charge retention by the *N*-terminal portion is shown below; it has arbitrarily been selected between residues 3 and 4, but occurs also at the other peptide bonds.



This fragmentation, shown above for the case of electron impact (EI) mass spectrometry, also occurs under chemical ionization (CI) conditions. Given the masses of the twenty common amino acids, the ions corresponding to fragments containing the amino acids associated with R^1, R^1R^2, $R^1R^2R^3$, *etc.* are sufficient to sequence the peptide.[8] The method is limited (in EIMS) to peptides containing up to *ca.* 10 amino acids, and requires $\geqslant$25 nmol of peptide. In our own

[7] P. A. Bell and E. Kodicek, *Biochem. J.*, 1970, **116**, 755.
[8] H. R. Morris, D. H. Williams, G. Midwinter, and B. S. Hartley, *Biochem. J.*, 1974, **141**, 701.

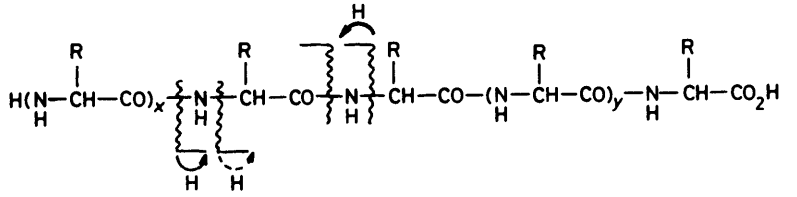
laboratory, and others, it has been used to sequence, or partially sequence, proteins. The proteins are broken into suitable fragments by enzymic digestion, and the fragments sequenced as above. Long sequences are then generated by repeating the process with enzymes of differing specificity, and overlapping (by means of portions of common sequence) all the fragments so obtained.[9] However, the same results can be achieved with comparable or greater efficiency by solution methods. The unique advantage of mass spectrometry in this area lies in studying unusual linear peptides. The unusual features include the identification of naturally occurring *N*-terminal blocking groups (which preclude the use of the Edman degradation). For example, an *N*-terminally blocked peptide was isolated *via* enzymic digestion of an alcohol dehydrogenase from *Drosophila melanogaster* N-11. It was acetylated with $(CD_3CO)_2O$ prior to permethylation; the mass spectrum of the product established the structural unit CH_3CO-Ser-Phe-.[10] Thus, the natural blocking group in the enzyme is determined to be acetyl; being already present in the enzyme, it precludes further reaction at the *N*-terminus with the deuterated analogue of acetic anhydride.

A more complex example is found in the recent determination of the structure of an unusual peptide isolated from larvae of the sawfly.[11] The poisoning of cattle by ingestion of larvae of an Australian species of sawfly (*Lophyrotoma interrupta*) is a serious problem in several of the grazing areas of Queensland.[12,13] Generally, the disorder occurs in late winter or spring when there is a heavy infestation of the silver-leaf ironbark tree (*Eucalyptus melanophloia*), the principal host of the sawfly larvae. Large accumulations of dead larvae may collect on the ground. Animals eat such material, and develop a considerable liking for it. It has been suggested that it may act as a protein source. However, the end result is poisoning, and critically affected animals usually die within two days.

The LD_{100} of the essentially pure toxin is 2 mg/kg. Amino acid analysis shows the presence of the following amino acids (molar ratios given in parenthesis): Asp (1.89), Glu (1.93), Ala (1.00), Val (1.16), Ile (0.82), Phe (0.92) (taking Ala = 1.00). Initial work, relying on the principle of electron-impact sequencing outlined above established the presence of benzoyl as an *N*-terminal blocking group, and led to a partial sequence PhCO-Ala-Phe-Val-Ile... However, a typical problem of EI sequence determination, that of the higher mass ions being of low or negligible abundance (a particularly severe problem in this case), precluded extension of the sequence. Fortunately, at this period of the study, the technique of fast atom bombardment mass spectrometry (FAB MS) became available.[14] In this technique, a few μg (say 1—100 μg, depending on the amount of sample available, and the suitability of the compound for the technique) of sample are dissolved in

[9] A. Auffret, T. J. Blake, and D. H. Williams, *Eur. J. Biochem.*, 1981, **113**, 333.
[10] A. D. Auffret, D. H. Williams, and D. R. Thatcher, *FEBS Lett.*, 1978, **90**, 324.
[11] D. H. Williams, S. Santikarn, F. DeAngelis, R. J. Smith, D. R. Reid, P. B. Oelrichs, and J. K. MacLeod, *J. Chem. Soc., Perkin Trans. 1*, 1983, 1869.
[12] F. H. S. Roberts, *Queensl. Agric. J.*, 1932, **37**, 41.
[13] L. L. Callow, *Queensl. Agric. J.*, 1955, **81**, 155.
[14] M. Barber, R. S. Bordoli, R. D. Sedgwick, and A. N. Taylor, *J. Chem. Soc., Chem. Commun.*, 1981, 325.

a viscous, low-volatility, polar solvent such as glycerol or thioglycerol. The solution, on a mass spectrometer probe tip, is bombarded with xenon atoms of 4—10 keV translation energy. Although the technique is not a panacea for non-polar molecules, polar molecules in the molecular weight range (say) 300—5000 daltons are projected (as ions) into the gas phase with astonishing efficiency. These ions, normally MH^+ in positive ion spectra, or $[M-H]^-$ in negative ion spectra (M being the molecule of interest), are then detected in the mass spectrometer in the usual way. In FAB spectra, MH^+ {or $[M-H]^-$} are normally of high abundance, with fragment ions being of lower abundance. The observed fragmentation processes correspond to low energy cleavages and, as for CI, give important sequence information in the case of linear peptides. For example, underivatized peptides give molecular weight information that is readily obtained up to 3000 daltons; and, with the appropriate apparatus and technique, to 5000 daltons. In favourable cases, underivatized peptides of up to *ca.* 15 residues have been shown to give sequence information. Some of the most useful fragmentations are given in (3).



(3)

Where the horizontal line leaving the 'wavy' line does so to the left, these fragmentations occur with charge retention by the *N*-terminal fragment; where this horizontal line is to the right of the 'wavy' line, charge retention is by the *C*-terminal fragment. The symbol $\overset{H}{\frown}$ (or $\underset{H}{\smile}$) indicates that the cleavage is associated with hydrogen migration to the charged fragment. The symbol $\overset{H}{\cdots}$ indicates that the cleavage is sometimes observed with associated hydrogen migration, and sometimes without. The mass differences associated with these cleavages have been tabulated.[15]

These points are illustrated by the positive and negative ion FAB spectra of the toxin produced in Figures 2 and 3, respectively. Cleavage (ii) [see (3)] gives ions corresponding to the acylium ions (4), occurring at *m/z* 105, 176, 323, 422, 535, 650, 765, and 894 (and MH^+ at 1040, Figure 2) and indicate the sequence PhCO-Ala-Phe-Val-Ile-Asp-Asp-Glu-Gln; or the *C*-terminal Gln could equally be replaced by the isomeric amino acid carrying the amide on the backbone carbonyl group and the carboxy function in the side chain [iso-Gln; see (5) and (6)].

[15] D. H. Williams, C. V. Bradley, S. Santikarn, and G. Bojesen, *Biochem. J.*, 1982, **201**, 105.

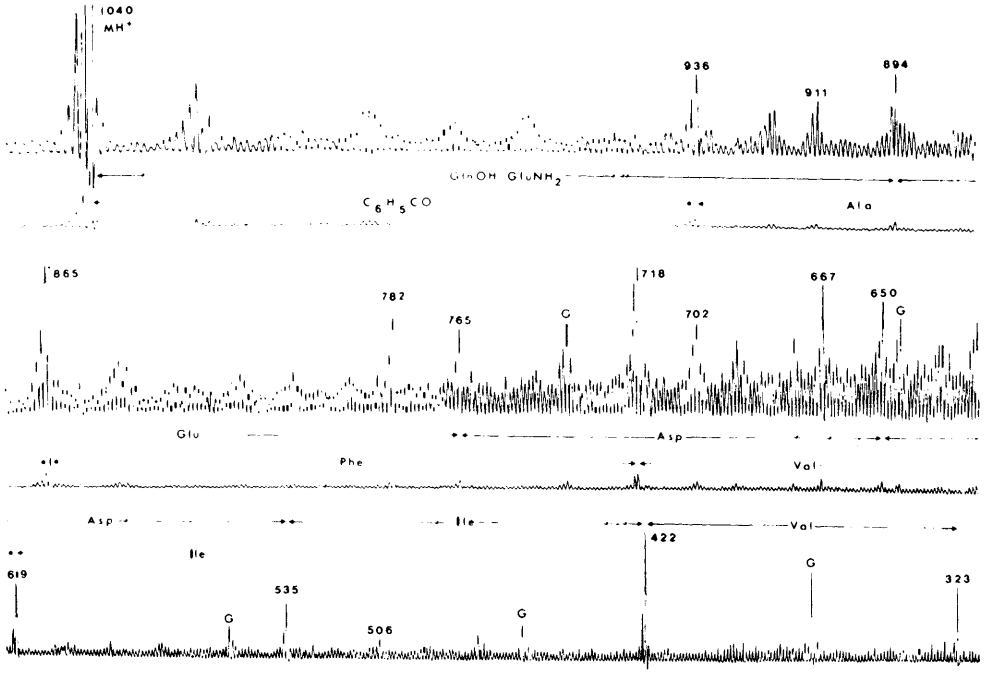


Figure 2 *The positive ion FAB mass spectrum (m/z 320—1050) of the toxic peptide from sawfly larvae. The peaks marked G represent protonated oligomers of glycerol (the solvent used in the determination)*

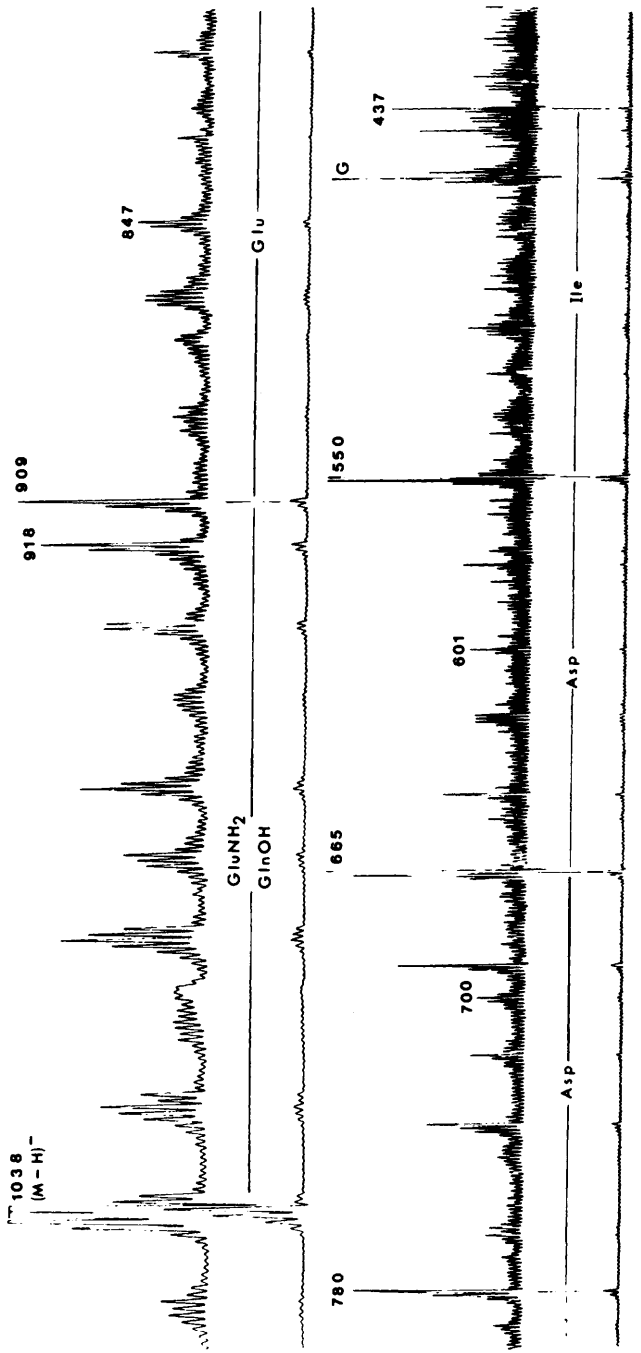
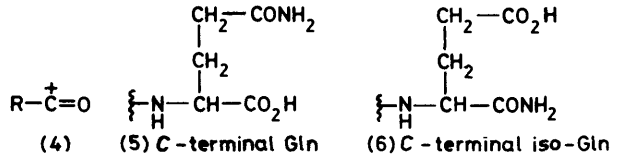
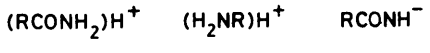


Figure 3 *The negative ion FAB mass spectrum (m/z 400—1060) of the toxic peptide from sawfly larvae*



(4) (5) *C*-terminal Gln (6) *C*-terminal iso-Gln

Fragments formed *via* cleavage (i) [see (3)], and corresponding to fragment ions of the general formula (7), at *m/z* 667, 782, and 911 (in conjunction with MH^+, Figure 2) confirm the nature of the three amino acids constituting the *C*-terminal portion.



(7) (8) (9)

Those sequence ions which contain the *C*-terminal portion of the peptide are formed by cleavage (iv) [see (3)]. Cleavage (iv) gives rise to protonated amine fragments (8), which occur at *m/z* 865, 718, 619, and 506 in Figure 2; they indicate the sequence X-Phe-Val-Ile-Y, where the masses of X and Y are in accord with the sequence proposed above. The negative ion FAB spectrum (Figure 3) shows an extremely abundant $(M-H)^-$ ion at *m/z* 1038. A series of sequence ions corresponding to cleavage (i) [see (3)], and most generally expressed in terms of the anion (9), occur at *m/z* 338, 437, 550, 665, 780, and 909. These, together with the molecular weight information, indicate the sequence X-Val-Ile-Asp-Asp-Glu-Gln (or, as before, the *C*-terminal residue may be iso-Glu). These data attest to the power of FAB MS not only in determining the molecular weights of relatively large polar molecules, but also in sequence determination of peptides (which is especially valuable when they are *N*-terminally blocked).

The above established sequence might be regarded as giving a 'structure' if the peptide were a product of ribosomal synthesis of a peptide. In such cases, it can be reasonably assumed that Asp, Glu, and Gln are incorporated as such, and not as their iso-structures; and further, that all the amino acids have L-configuration. However, in our early experiments on the toxin (prior to the availability of FAB MS), an enzymic digestion with chymotrypsin had been attempted, in the expectation that cleavage at the *C*-terminal side of the Phe residue would occur, thus giving a *C*-terminal fragment which might conceivably have been sequenced by EI MS. The attempted enzymic digestion failed. One possible explanation for this failure is that one or more of the amino acids in the toxin might have D-stereochemistry.

Mass spectrometry cannot throw light on this question, but it can be probed using gas chromatography on a chiral column. The toxin was hydrolysed (6M-HCl at 110 °C for 72 h), and the resulting amino acids converted into their *N*-trifluoroacetyl isopropyl ester derivatives; these volatile derivatives were chromatographed on a chiral GC column ('Chirasil-Val'). Comparison of the retention times of the components with those of authentic samples established that Val, Ile,

one Asp, and a Glu (or Gln) have the L-configuration; and that Ala, Phe, one Asp, and a Glu (or Gln) have the D-configuration. The ambiguity between Glu and Gln arises because Gln is converted into Glu by the total hydrolysis procedure.

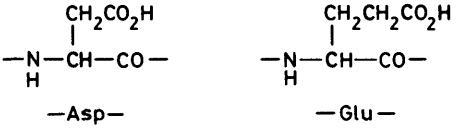
The above results increased the fascination and importance of the problem. Prior to this study, no peptide isolated from an animal had contained more than a single D-amino acid. In 'higher' organisms (the eucaryotes, including the animals), the synthesis of peptides and proteins occurs on ribosomes; in this process, only the 20 standard amino acids, in their L-configurations, are known to participate. However, in lower organisms (the procaryotes, including bacteria), peptides synthesis can, on occasions take place without ribosomes; this type of synthesis is effected on multi-functional enzymes, and can include the utilization of unusual amino acids (*e.g.* phenylglycine), and amino acids in their D-configuration. Subsequently, we will return to this problem of the procaryotic or eucaryotic origin of the peptide toxin although, unfortunately, we will not answer it!

The above data give a further twist to the problem. The occurrence, in the hydrolytic products, of one L-Asp, one D-Asp, one L-Glu, and one D-Glu gives rise to no less than four combinations of stereochemistry at the *C*-terminus; *viz.*,

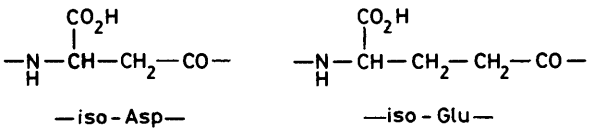
. . D-Asp-L-Asp-D-Glu-L-Gln, . . . L-Asp-D-Asp-D-Glu-L-Gln,

. . . D-Asp-L-Asp-L-Glu-D-Gln, . . . L-Asp-D-Asp-L-Glu-D-Gln.

Additionally, since the peptide is 'non-standard', we have to consider that Asp and Glu may be incorporated not only as they normally are in proteins [(10); see also (5)], but as their iso-structures, in which the usual side chain becomes part of the peptide backbone [(11); see also (6)]. Thus, determination of the amino acid sequence still leaves a total of 64 possible structures. The determination of a unique structure will now be considered.



(10)



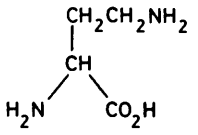
(11)

The stereochemistries of the Asp residues were determined first. A possible strategy is to isolate a peptide fragment containing only one of the Asp residues, and then to determine the stereochemistry of that residue. Fortunately, it is known that acid hydrolysis of peptides occurs selectively as Asp residues, and therefore the toxin was subject to relatively mild acid hydrolysis (10% aqueous HCl, 100 °C.

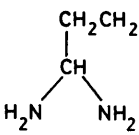
1.5 h), and the resulting peptides fragments were isolated *via* preparative HPLC. The molecular weights of the isolated fragments were then determined by FAB MS.[11] One fraction contained abundant ions only at *m/z* 493 and 247 in the positive ion FAB spectrum. These ions correspond to MH^+ signals from the peptides Phe-Val-Ile-Asp and Ile-Asp. Since it can be argued reasonably that one number does not guarantee securely these sequences, a few μg of the fraction was subjected to *N*-acetylation (Ac_2O–H_2O, 1:4, room temperature 4 h), and to esterification (1% HCl in MeOH, room temperature, 10 h). The resulting derivatives, upon FAB MS analysis showed MH^+ peaks at *m/z* 535 and 289; and *m/z* 521 and 275, respectively. These peaks correspond to the anticipated mono-*N*-acetyl derivatives (+42 daltons) and dimethyl ester (+28 daltons) derivatives, and therefore support the structure assignments. Note that these reactions occur in good yield, and the excess of reagents can be removed under vacuum, thus avoiding any tube-to-tube transfers or other manipulations in the derivatizations. Therefore, functional-group determination can be carried out on a μg scale by FAB MS; the importance and power of this technique is worthy of emphasis.

The above peptide fraction, containing only the aspartic acid residue at position 5 (Asp^5, numbering amino acids of the toxin from the *N*-terminus) was then hydrolysed. When volatile derivatives (prepared as given earlier) of the resulting amino acids were analysed on the chiral GC column, Asp^5 was shown to have the D-configuration. Therefore, Asp^6 has the L-configuration.

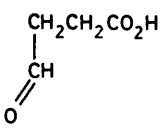
The stereochemical assignments of Glu^6 and Gln^7 can be determined if, prior to hydrolysis, Gln^7 is modified such that the product of this modification does not produce a Glu residue upon hydrolysis. A suitable modification was achieved *via* a Hofmann degradation,[11] which brings about the conversion $RCONH_2 \rightarrow RNH_2$. Thus, Gln will afford 2,4-diaminobutyric acid (12) upon hydrolysis, whereas iso-Gln would transiently produce 4,4-diaminobutyric acid (13), and hence the ninhydrin-negative 4-oxobutyric acid (14).



(12)



(13)



(14)

Following Hofmann degradation of the toxin, and acid hydrolysis of the product, amino acid analysis showed the production of (12). Moreover, analysis of a volatile derivative of (12) on a chiral GC column established it as possessing the L-configuration. Therefore, the *C*-terminal residue is Gln (and not iso-Gln) and has the L absolute configuration. It follows that the penultimate residue, producing Glu upon hydrolysis, has the D-configuration.

The ambiguity between normal and iso-residues at positions 5, 6, and 7 in the toxin was resolved by successively exposing the toxin to (i) Hofmann degradation and (ii) Ac_2O–pyridine–D_2O. The latter treatment causes racemization of, and

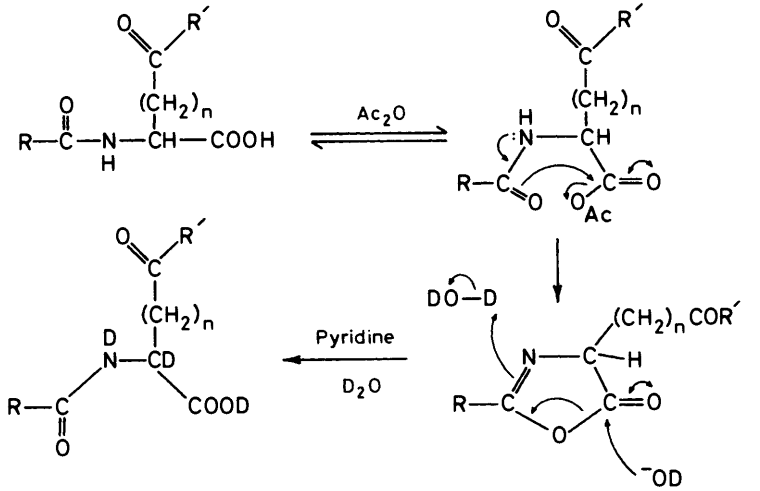


Figure 4 *The chemical basis for racemization of, and deuterium incorporation into, amino acids possessing a free carboxy group attached to the* α-CH

deuterium incorporation into, the α-position of those amino acids which have free carboxy groups attached to the α-carbon,[16] *i.e.* those in-chain amino acids which are incorporated as iso-structures [see (11)], and normally constituted *C*-terminal amino acids [see, for example, (5), which has already been shown to be present in the toxin].

The chemical basis for the racemization and deuterium incorporation is given in Figure 4. When the free carboxy group is attached to the α-CH, the mixed anhydride formed upon reaction with Ac_2O is subjected to intramolecular nucleophilic attack by the amide function incorporating the carbonyl group of the adjacent amino acid. As a result, the acetate anion is displaced, and a 5-membered ring oxazolone is formed (Figure 4). In this ring system, the hydrogen of the α-CH group is relatively acidic, being not only adjacent to a carbonyl group, but also to an olefinic double bond. In the presence of a base (pyridine) and D_2O, it will therefore be replaced by deuterium. Additionally, the intermediate enolate anion is sp^2 hydridized at the original α-CH group, and a further consequence of deuterium incorporation is racemization at the α-carbon of the amino acid. Thus, base-catalysed hydrolysis of the oxazolone (Figure 4) produces an amino acid which is not only deuteriated at the α-position, but also racemic.

Toxin treated in the above manner yielded 2,4-diaminobutyric acid which incorporated deuterium and was racemized. Such deuterium incorporation and racemization was not observed for the Asp residues, nor for the Glu residue. Therefore, these residues are not incorporated as iso-structures. The complete structure of the toxin is therefore that reproduced in Figure 5.

In summary, the slowly uncovered structural features of the toxin, which dic-

[16] G. N. Holcomb, S. A. Jones, and D. N. Ward, *Biochemistry*, 1968, 7, 1291.

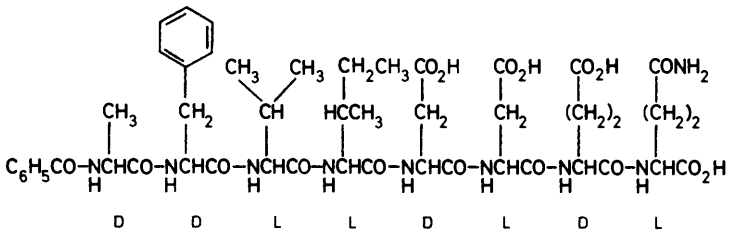


Figure 5 *Structure of the peptide toxin from the sawfly* Lophyrotoma interrupta

tated those methods which could successfully tackle the problem are (i) the presence of the *N*-terminal blocking group, precluding the use of the Edman degradation, (ii) the presence of D-amino acids, precluding the use of enzymic digestion, (ii) the presence of a linear peptide, allowing the application of EI and, particularly in this case, FAB MS sequencing. Note that stereochemical problems are rarely amenable to analysis by mass spectrometry, and the successful use of a chiral column required standards of known absolute configurations. Nuclear magnetic resonance (n.m.r.) was not called upon significantly in the structure elucidation because the structural sub-units were of established types. Additionally, 'communication' between these sub-units, although in principle possible by means of the nuclear Overhauser effect (NOE, see later), could not be successfully exploited in the above example. In the linear molecule, the conformation was too floppy, and the molecular weight in 'no-man's land', between lower and higher values which could have been exploited.

The biosynthetic origins (procaryotic or eucaryotic) of the toxin remain an enigma. It has recently been established that a fungus lives symbiotically with the sawfly larvae, in the gut.[17] It is possible that the fungus is responsible for the production of the D-amino acids, or even later stages of the biosynthetic pathway; but this is speculation, and a definitive conclusion will have to await further experimentation.

4 The Structure Elucidation of Unusual Monocyclic Peptides

The previously discussed sequencing of (unusual) linear peptides by mass spectrometry is a useful general method. The strategy is not directly applicable to monocyclic peptides because fragmentation of an amide bond produces a linear peptide and no fragments. It might be thought that the internal energy of the linear product would be sufficient to produce fragments *via* further amide bond cleavages; but such is not generally the case in FAB spectra (low internal energy for ions). Further fragmentation often occurs in EI spectra, but the spectra are more difficult to interpret, and high resolution data are usually needed (see later).

An alternative general strategy has therefore to be sought. One approach is to cleave chemically an amide bond of the monocyclic peptide, and then use EI and/or

[17] Professor H. Kleinkauf, Technical University of Berlin, personal communication.

FAB MS sequencing of the linear product. Such an approach was applied successfully to determine the structure of a peptide toxin from the blue-green alga, *Microcystis aeruginosa*.[18,19]

Since the end of the last century, heavy blooms of the blue-green alga *Microcystis aeruginosa* have been known to cause deaths among livestock consuming heavily contaminated water. The lethal doses (LD_{50}) of the purified toxins are *ca.* 50 μg/kg for mice on intraperitoneal injection.[20] The toxic peptides of *M. aeruginosa* are normally contained within the algal cells, and are released only when the cell is damaged. It has recently been concluded[20] that an increase in liver damage was indicated in a human population during the period of a bloom of *M. aeruginosa* in the water-supply reservoir.

Although enzymatic experiments and amino acid analysis (equimolar ratios of L-leucine, D-glutamic acid, *erythro*-β-methyl-D-aspartic acid, L-alanine, and D-alanine)[21] established the peptidic nature of the toxin, no sequence ions were seen in the FAB mass spectrum (which established a molecular weight of 909 daltons). Additionally, it was shown that the peptide lacked a free basic *N*-terminus because the molecular weight was unchanged following treatment with Ac_2O–H_2O. The lack of fragmentation in the FAB spectrum, and the absence of a free *N*-terminus, together suggest the possibility of a cyclic peptide. However, it is clear that the identified amino acids fall far short of accounting for the observed molecular weight. Since the toxin was available in a quantity of *ca.* 2 mg, it was possible to throw light on the non-standard amino acid components of the toxin by high-field ^1H n.m.r. at 400 MHz.

Before discussing the results of the n.m.r. study, it is necessary to outline the principle of an n.m.r. phenomenon which has assumed great importance recently. It is the nuclear Overhauser effect (NOE). Although this effect has been known for many years, and was first exploited in organic chemistry in the 1960s, it is the advent of difference spectroscopy which has resulted in its widespread use during the last few years.[22]

When nuclei which behave like bar magnets are placed in a magnetic field, they can occupy a high- and a low-energy state. A nuclear magnetic resonance absorption signal is obtained from such nuclei when, upon supplying electromagnetic radiation of a suitable frequency, ν, there is a net passage of nuclei from the low to the high energy state. Nuclei may pass back from the high to the low energy state by a process known as relaxation. Proton nuclei are normally relaxed by a mechanism which involves neighbouring protons. The effectiveness of such neighbouring protons in bringing about relaxation depends upon r^{-6}, where r is the internuclear distance between the proton being relaxed and the proton effecting the relaxation.

[18] S. Santikarn, D. H. Williams, R. J. Smith, S. J. Hammond, D. P. Botes, A. Tuinman, P. L. Wessels, C. C. Viljoen, and H. Kruger, *J. Chem. Soc., Chem. Commun.*, 1983, 652.

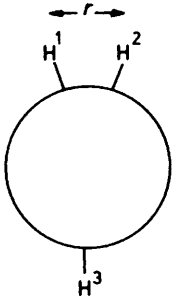
[19] D. P. Botes, A. Tuinman, P. L. Wessels, C. C. Viljoen, H. Kruger, S. Santikarn, D. H. Williams, R. J. Smith, and S. J. Hammond, *J. Chem. Soc., Perkin Trans. 1*, in press.

[20] I. R. Falconer, A. M. Beresford, and M. T. C. Runnegar, *Med. J. Aust.*, 1983, 511.

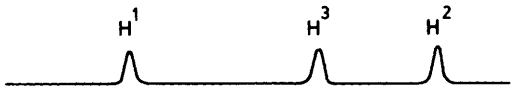
[21] D. P. Botes, C. C. Viljoen, H. Kruger, P. L. Wessels, and D. H. Williams, *Toxicon*, 1982, **20**, 1037.

[22] See, for example, J. K. M. Sanders and J. D. Mersh, in 'Progress in Nuclear Magnetic Resonance Spectroscopy', Vol. 15, pp. 353—400, Pergamon Press, Oxford, 1983.

If the intensity of the resonance of one proton (H^1) is normally I, then if $r_{1,2}$ is relatively small (as a useful guide for our present purposes say, < 0.3 nm), the effect of irradiating a second proton (H^2) before recording the intensity of the resonance of H^1 is to change its intensity to I', *i.e.* $I \neq I'$. This change in intensity is called the NOE. The effect is indicated schematically in (15)–(18). Three hydrogen nuclei in a molecule are indicated in (15). The signals which arise in the ^1H n.m.r. spectrum of (15) in the absence of pre-irradiating any of the protons is given in (16). Upon pre-irradiation of H^1 at its resonance frequency ν_1, the ^1H n.m.r. spectrum is modified to (17). Note that in (17), the resonance of H^1 has disappeared, because its resonance is 'saturated', *i.e.* the populations of upper and lower energy levels are equalized, by the pre-irradiation. The H^2 resonance intensity is reduced due to the NOE. The NOE difference spectrum [NOEDS, (18)] is obtained by on-line computer subtraction of (17) from (16).



(15)



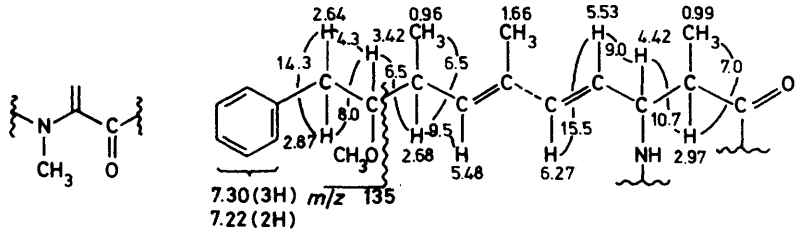
(16)

(17)

(18)

Note that since H^3 is distant from H^1, no observable NOE occurs for H^3. The value of the difference spectrum is evident for molecules containing many protons; the NOEDS then contains only signals due to the irradiated proton H^1, and less intense signals due to those protons which are proximate to H^1. Such experiments are extremely valuable in determining structures of molecules in solution.

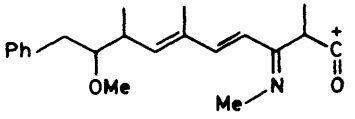
Analysis of the ^1H 400 MHz spectrum of the toxin in D_2O established from chemical shift values and spin decoupling experiments the presence of the *N*-methyldehydroalanine (19) and of units (20) and (21). The connectivity between (20) and (21) was established to occur *via* the C—C bond indicated by a broken line (using NOEs, determined in [$^2H_6$]DMSO solution). Thus, pre-irradiation of the $\delta 5.53$ resonance led to a 12% reduction in the intensity of the $\delta 1.66$ resonance {normally expressed in the concise form [5.53] → 1.66 (−12%)}. Similarly, [6.27] → 5.48 (−15%) and [5.48] → 6.27 (−15%).



(19) (20) (21)

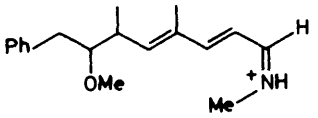
The above combination of the two units provides an excellent example of how connectivity can be established from NOEs when evidence is not available from coupling constants (because four or more σ-bonds occur between adjacent protons).

Note that in the above case, the n.m.r. experiment does not determine the nature of the electronegative substituents (OMe, NH, C=O) on the chain; but only that some electronegative substituent must be attached to the carbons bearing the protons at δ3.42, 4.42, and 2.97. The nature of these substituents was established from fragment ions occurring in the EI mass spectrum of the toxin and/or permethylated toxin. These spectra afforded ions at *m/z* 135 [$PhCH_2CH{=}\dot{O}Me$, with or without permethylation], *m/z* 326 [(22), following permethylation] and *m/z* 272 [(23), following permethylation]. The atomic compositions of the fragments (22) and (23) are supported by high resolution measurements. When the permethylation procedure utilises CD_3I instead of CH_3I, the *m/z* values of (22) and (23) shift to 329 and 275, respectively. Thus, the nitrogen atom is methylated in the derivatization.



m/z 326.2112
(calc. 326.2120)

(22)



m/z 272.2011
(calc. 272.2014)

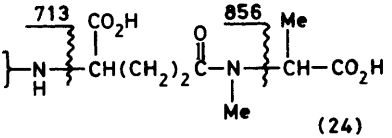
(23)

In view of the presumed cyclic nature of the toxin, the hydrolysis of a single amide bond in the peptide might be attempted by limited acid hydrolysis.

Since the *N*-methyldehydroalanine unit (19) might be modified by even quite mild acid treatment (Michael addition or enamine hydrolysis), it was deemed wise to reduce this residue prior to the pulse acid hydrolysis treatment. The reduction of (19) was effected by conjugate addition of hydride ion, from $NaBH_4$, to give *N*-methylalanine. This procedure has the convenience that $NaBD_4$ can alternatively be used, and thus provide an isotopic label for the reduced residue. The

reduced toxin was pulse hydrolysed (6M-HCl, 5 min, 100 °C) and a product of molecular weight 812 daltons (by FAB MS) isolated. Clearly, if the toxin has been ring opened, a portion of it has also been lost. Indeed, when the two-step procedure was repeated using $NaBD_4$, a product of molecular weight 812 daltons was again obtained. The product therefore corresponds to reduction (+2), ring opening (+18), loss of *N*-methylalanine (−85) and methanol [−32 daltons by elimination from (20)]. Micro-derivatization experiments followed by FAB MS (see earlier) show the product to contain one amino and three carboxy groups. Acetylation and permethylation of the product, followed by EI MS sequencing established the sequence Ala-Leu-'Glu'-Ala-*β*aa'-'Glu', where *β*aa' is the novel *β*-amino acid (20)/(21) minus methanol; and 'Glu' represents Glu or *β*-Me-Asp (or their iso-structures), since these units are all isomeric. When the linear peptide was subjected to one cycle of the Edman degradation, *N*-terminal alanine was cleanly removed (as shown by FAB MS);[23] and chiral column GC–MS experiments (see earlier) on the amino acids derived from the product of Edman degradation established the absence of D-Ala and the presence of L-Ala. As expected, a second Edman degradation removed L-Leu, but the third Edman cycle failed. Evidently, the third amino acid, in the linear peptide, is not incorporated as a standard amino acid. However, in the product of the second Edman cycle, the amino group of the third amino acid is present as a free, basic group. This group can therefore be converted into the sulphonamide by reaction with dansyl chloride. Acid hydrolysis of the product produced the dansyl derivative of *β*-methylaspartic acid.[18] Since this amino acid is not incorporated as a standard amino acid in the linear peptide, it must be incorporated as *β*-methyl-iso-aspartic acid. Note that this does not necessarily imply that this acid is present as the iso-structure in the original toxin; since the Asp → iso-Asp conversion can be brought about by either mild acid or mild base, the isomerization could have occurred in the course of the conversion of the toxin into the linear peptide. Since residue -3 on the linear peptide is *β*-Me-iso-Asp, it follows that the *C*-terminal residue is Glu (or iso-Glu).

It remains to establish the points of attachment of the *N*-methyldehydroalanine residue which, following $NaBH_4$ reduction of the toxin, was lost during pulse hydrolysis. This was achieved[19] by exposing the reduced toxin to a milder pulse hydrolysis (moist trifluoroacetic acid at room termperature for 24 h). The negative ion FAB mass spectrum of this hydrolysis product showed $[M-H]^-$ at *m/z* 928, indicating that the dihydro-toxin (molecular weight 911 daltons) had undergone simple hydrolytic addition of water, without loss of an amino acid residue. Moreover, *N*-terminal amide sequence ions [see (3), cleavage (i)] in this spectrum occurred at *m/z* 856 and 713, establishing that *N*-methyl-Ala is attached to a Glu (or iso-Glu) residue (24).[19]



(24)

[23] C. V. Bradley, D. H. Williams, and M. R. Hanley, *Biochem. Biophys. Res. Commun.*, 1982, **104**, 1223.

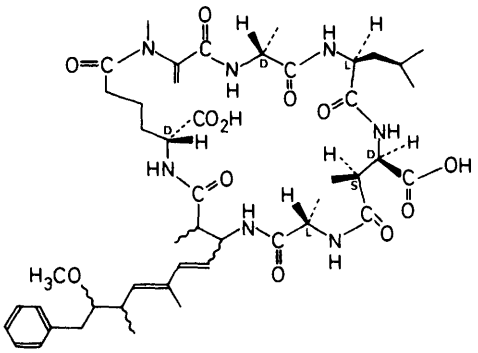


Figure 6 *Structure of the peptide toxin BE-4 from* Microcystis aeruginosa

The presence of iso-Glu in the toxin was indicated not only by the change in chemical shift of the α-CH proton upon titration of the carboxy group of the glutamic acid residue, but also by the chemical method outlined in Figure 4.[19] Similar chemical experiments on the unreduced toxin demonstrated that β-methyl-Asp is incorporated as the iso-structure in the toxin.[19]

Thus, a monocyclic structure for the toxin can be completed by attachment of the carbonyl of *N*-Me-dehydro-Ala to the amino group of D-Ala (Figure 6). Note that in the published work,[18] only a partial structure for the toxin variant BE-4 had been assigned, and that the tentative attachment[18] of *N*-Me-dehydro-Ala as an exocyclic residue is now shown[19] to be incorrect. Experiments to determine the stereochemistry of the asymmetric centres in the β-amino acid have yet to be carried out.

As in the case of the sawfly liver toxin, note how complex details of stereochemistry and connectivity can be proved by use of GC, enzymic methods, and MS, so long as the sub-units have standard structures. Where a sub-unit of novel structure was involved, it was necessary to turn to high field ^{1}H n.m.r.; and in particular, to spin decoupling and NOEs.

5 The Structure Elucidation of Polycyclic Peptides; The Vancomycin Group of Antibiotics and their Mode of Action

It will be clear from the foregoing two sections that attempted structure elucidation by mass spectrometry of polycyclic peptides is likely to fail. Fragmentation would require multiple bond rupture, which may not be feasible energetically. Such was our experience with the vancomycin group of antibiotics, which are a series of tri- and tetra-cyclic peptides. Thus, the rôle of mass spectrometry was largely in the determination of molecular weights; both of the antibiotics, and of fragments obtained by chemical degradation. By far the most important rôles in the structure

elucidation of vancomycin[24–28] and ristocetin[29–32] were played by chemical degradation, *X*-ray crystallography, and ^{1}H n.m.r.

Most of these experiments have been reviewed previously,[33] and therefore it will suffice here to state the general principles involved. The structure of vancomycin was partially determined by the characterization of fragments produced by degradation,[24] and some of the bonds broken in hydrolytic experiments then 'reconstituted' by use of spin decoupling experiments and NOEs (and other techniques).[25] However, the numbers of asymmetric centres of undetermined stereochemistry were sufficiently large to preclude a total solution of the problem by these methods. The structure of a degradation production of vancomycin, CDP-I, was determined by *X*-ray crystallography.[26] It was initially assumed that in the mild conditions for conversion of vancomycin into CDP-I (80 °C, 3 days, pH 4.2), the only chemical change was $RCONH_2 \rightarrow RCO_2H$. However, it has been shown subsequently that a chlorine-containing aromatic ring undergoes a rotation of *ca.* 180 °;[27] and that an Asn residue of vancomycin isomerizes to an iso-Asp residue in CDP-I.[28] The currently accepted structure of vancomycin is reproduced in Figure 7.

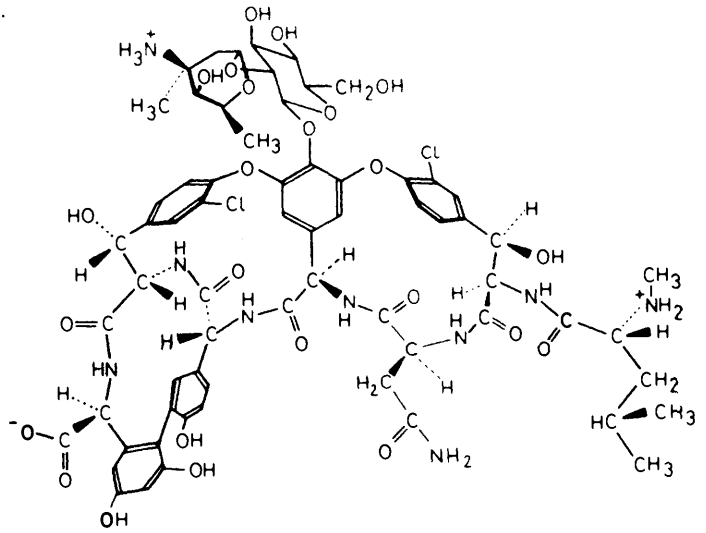


Figure 7 *Proposed structure of vancomycin*

[24] G. A. Smith, K. A. Smith, and D. H. Williams, *J. Chem. Soc., Perkin Trans. 1*, 1975, 2108.
[25] D. H. Williams and J. R. Kalman, *J. Am. Chem. Soc.*, 1977, **99**, 2768.
[26] G. M. Sheldrick, P. G. Jones, O. Kennard, D. H. Williams, and G. A. Smith, *Nature*, 1978, **271**, 223.
[27] M. P. Williamson and D. H. Williams, *J. Am. Chem. Soc.*, 1981, **103**, 6580.
[28] C. M. Harris and T. M. Harris, *J. Am. Chem. Soc.*, 1982, **104**, 4293.
[29] D. H. Williams, V. Rajananda, and J. R. Kalman, *J. Chem. Soc., Perkin Trans. 1*, 1979, 787.
[30] J. R. Kalman and D. H. Williams, *J. Am. Chem. Soc.*, 1980, **102**, 897.
[31] D. H. Williams, V. Rajananda, G. Bojesen, and M. P. Williamson, *J. Chem. Soc., Chem. Commun.*, 1976, 906.
[32] C. M. Harris and T. M. Harris, *J. Am. Chem. Soc.*, 1982, **104**, 363.
[33] D. H. Williams, V. Rajananda, M. P. Williamson, G. Bojesen, in 'Topics in Antibiotic Chemistry', ed. P. Sammes, Horwood, Chichester, UK, 1980, Vol. 5, Part B.

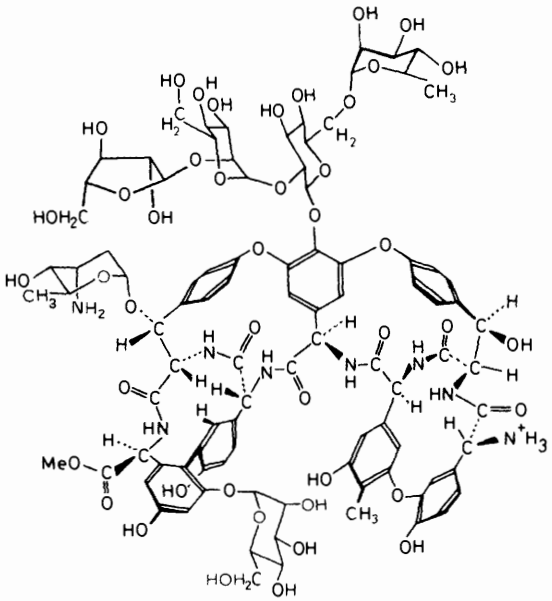


Figure 8 *Proposed structure of ristocetin A*

The structure elucidation of a second member of the vancomycin group, ristocetin A, has not been aided directly by *X*-ray crystallography because no crystals suitable for structure determination have been obtained. However, there are remarkable similarities between many of the proton chemical shifts, coupling constants, and NOEs observed in the ^{1}H n.m.r. spectra of vancomycin and ristocetin. Hence, using both chemical[29] and n.m.r. techniques,[29,30] it was possible to propose a structure for ristocetin A.[31] The first structure proposal[31] has been revised by inversion of the stereochemistry at the *N*-terminus (as a result of chemical degradative experiments[32]), and the currently accepted structure is reproduced in Figure 8. Much definitive work on the structure of the saccharide portion of ristocetin A was carried out by Sztaricskai *et al.*[34] Some general features of the structures reproduced in Figures 7 and 8 are (i) both are heptapeptides in which three rings are made by phenol oxidative coupling (two by C—O bond coupling, and one by C—C bond coupling to give a biphenyl unit); a fourth ring in ristocetin A is formed by a third C—O bond coupling, (ii) the stereochemical arrangements in the 'left-hand portion' (as presented) of both structures, in the aglycone unit, are the same, although ristocetin A has a methylated *C*-terminus which vancomycin lacks, (iii) the stereochemistries at the α-CH centres of the seven amino acids are the same, and, from the *N*-terminus are *R,R,S,R,R,S,S*. Note,

[34] See, for example, F. Sztaricskai, A. Neszmelyi, and R. Bognar, *Tetrahedron Lett.*, 1980, 2983, and references cited therein.

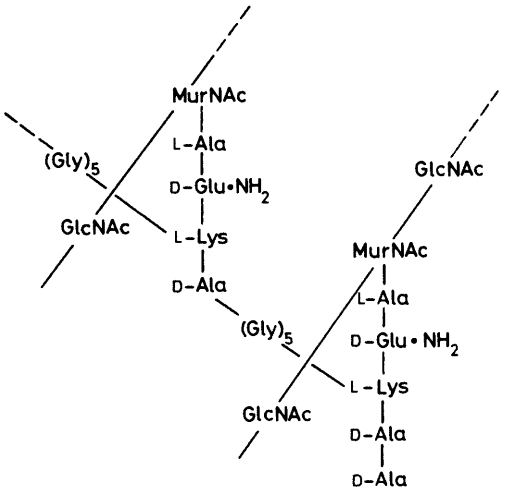
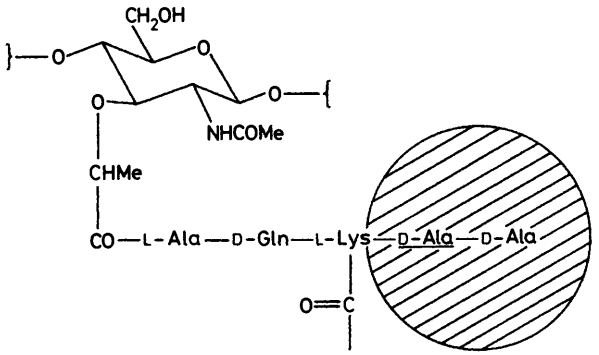


Figure 9 *A portion of the cell wall peptidoglycan in* Staphylococcus aureus

however, that *N*-methyl-leucine and asparagine in vancomycin (Figure 7) are replaced by substituted (and cross-linked) phenylglycine units in ristocetin A (Figure 8). The *p*-hydroxyphenylglycine units which are incorporated into these molecules have been shown[35,36] to be derived from tyrosine, and the *m*-dihydroxylated phenylglycine units to be derived from acetate.

The interesting question arises as to how these molecules act as antibiotics. A number of years ago, it was shown that they inhibit the biosynthesis of the bacterial cell wall in gram-positive bacteria. In the absence of antibiotic, a key step in completing a strong two-dimensional structure for the cell wall involves a trans-peptidase enzyme. The function of this enzyme is to attach the *N*-terminus of a $(Gly)_5$ residue to the carbonyl carbon of the D-Ala residue of (25) (indicated by underlining). During this process, the *C*-terminal D-Ala is displaced,[37] and the new peptide bond between D-Ala and Gly completes the cross-linking of two parallel strands of the cell-wall polysaccharide [Figure 9, which shows one -D-Ala-$(Gly)_5$-cross-link (centre) already formed, and a -D-Ala-D-Ala residue (right) prior to cross-link formation to an adjacent strand (to the further right, and not shown)].

[35] S. J. Hammond, M. P. Williamson, D. H. Williams, L. D. Boeck, and G. G. Marconi, *J. Chem. Soc., Chem. Commun.*, 1982, 344.
[36] S. J. Hammond, D. H. Williams, and R. V. Nielsen, *J. Chem. Soc., Chem. Commun.*, 1983, 116.
[37] K. Izaki, M. Matsuhashi, and J. L. Strominger, *J. Biol. Chem.*, 1968, **243**, 3180.



(25)

It was shown[38-40] that cell-wall peptide precursors terminating in -D-Ala-D-Ala accumulate when bacterial cell growth is inhibited by vancomycin and ristocetin. In particular, Perkins and co-workers[41-43] showed that cell-wall precursor peptides terminating in -D-Ala-D-Ala form strong 1:1 complexes with both vancomycin and ristocetin. Thus, if the hatched area in (25) represents vancomycin or ristocetin, it may reasonably be proposed that they exert their antibiotic action by binding, as indicted in (25), to cell wall precursors terminating in D-Ala-D-Ala. It is the molecular basis for this interaction that we should now be able to study.

Since ^{1}H n.m.r. parameters (*e.g.* chemical shift, NOE) are sensitive to intermolecular association, the interaction between the antibiotics and a simple cell-wall analogue, Ac-D-Ala-D-Ala, might usefully be proved by the n.m.r. method. The first studies which utilized this concept were carried out by Brown *et al.*[44,45] However, at the time of this work, the structures of the antibiotics were not known, and the NOEDS technology not yet available. Yet the concept represented an important advance in the area, and the useful discovery was made that the methyl resonance of the *C*-terminal alanine residue was shifted appreciably upfield upon complex formation with vancomycin. This indicated that in the complex, this methyl group lay over the face of an aromatic ring of vancomycin.

Once the structures of the antibiotics were known, it was possible to prove the molecular basis for the antibiotic/cell-wall analogue interaction by using two powerful methods. First, the temperature dependence of amide NH-resonances. These are *ca.* 6—10 $\times$ 10^{-3} p.p.m./°C to high field in DMSO solution, when such an NH is exposed fully to the DMSO solvent. This temperature dependence arises

[38] P. E. Reynolds, *Symp. Soc. Gen. Microbiol.*, 1966, **16**, 47.
[39] D. C. Jordan, *Biochem. Biophys. Res. Commun.*, 1961, **52**, 403.
[40] C. H. Wallas and J. L. Strominger, *J. Biol. Chem.*, 1963, **238**, 2264.
[41] A. N. Chatterjee and H. R. Perkins, *Biochem. Biophys. Res. Commun.*, 1966, **24**, 489.
[42] M. Nieto and H. R. Perkins, *Biochem. J.*, 1971, **123**, 773.
[43] M. Nieto and H. R. Perkins, *Biochem. J.*, 1971, **123**, 789.
[44] J. P. Brown, J. Feeney, and A. S. V. Burgen, *Mol. Pharm.*, 1975, **11**, 119.
[45] J. P. Brown, L. Terenius, J. Feeney, and A. S. V. Burgen, *Mol. Pharm.*, 1975, **11**, 126.

due to the breaking of $\gtrless S{=}O\cdots H{-}N\lessgtr$ hydrogen bonds with increasing temperature. In contrast, if an amide NH is involved in an intermolecular hydrogen bond, it is not exposed to solvent, and a much smaller temperature dependence [say $(0—2) \times 10^{-3}$ p.p.m./°C] results. Using this criterion, and the downfield shifts of amide NH resonances upon intermolecular hydrogen bond formation, it was possible to determine those amide NH-groups of antibiotic and cell-wall analogue which were involved in complex formation.[46,47] Such criteria led, in conjunction with other evidence such as model building (CPK models), to proposals for the interactions shown in Figures 10 and 11.[46,48] In both these figures, exploded views of the complexes are shown; hydrogen bonds, formed between carbonyl groups of one component and amide NH-groups of the second component, are indicated by broken lines which join the two groups. These are (from left to right in both figures, with the cell-wall analogue group given first), (i) acetyl carbonyl to the *C*-terminal (seventh residue) NH of the antibiotic, (ii) NH of the *C*-terminal D-alanine to the carbonyl of the trioxygenated phenylglycine unit (fourth residue) of the antibiotic, (iii) carboxyl oxygen of the *C*-terminal D-alanine to the NH-group of the fourth residue of the antibiotic, (iv) carboxyl oxygen of the *C*-terminal D-alanine to the NH-group of the third residue of the antibiotic and (v) carboxyl oxygen of the *C*-terminal D-alanine to the NH-group of the second residue of the antibiotic. The residue numbers of the antibiotic (numbered 1 to 7 from the *N*-terminus) are most readily identified by looking at Figures 10 and 11 in conjunction with Figures 8 and 7, respectively. Note that the proposed interactions allow a hydrophobic interaction of each of the alanine methyl groups with an aromatic ring of the antibiotic: the methyl group of the *C*-terminal alanine with the trioxygenated aromatic ring of residue 4; and the methyl group of the *N*-terminal alanine with the biphenyl moiety.

The second criterion (intermolecular NOE) can be used to check the validity of the above models. When the two components of either complex are brought together, it is seen that certain protons of the cell-wall analogue are in close proximity to protons of the antibiotics. These proximities are confirmed by irradiation, in turn, of these protons, and observation of the required NOEs. This use of intermolecular NOEs allows, in principle, the direct mapping of the receptor binding site of a drug. Unfortunately, the method is limited to small receptors, because the receptor has to give an analysable proton spectrum with reasonably sharp lines. We have, however, been able to apply the technique to determine the binding site, in solution, of a DNA tetranucleoside triphosphate for the antibiotic actinomycin D,[49] and the method appears to have great potential for future work.

[46] J. Kalman and D. H. Williams, *J. Am. Chem. Soc.*, 1980, **102**, 906.
[47] D. H. Williams and D. W. Butcher, *J. Am. Chem. Soc.*, 1981, **103**, 5697.
[48] D. H. Williams, M. P. Williamson, D. W. Butcher, and S. J. Hammond, *J. Am. Chem. Soc.*, 1983, **105**, 1332.
[49] D. G. Reid, S. A. Salisbury, and D. H. Williams, *Biochemistry*, 1983, **22**, 1377.

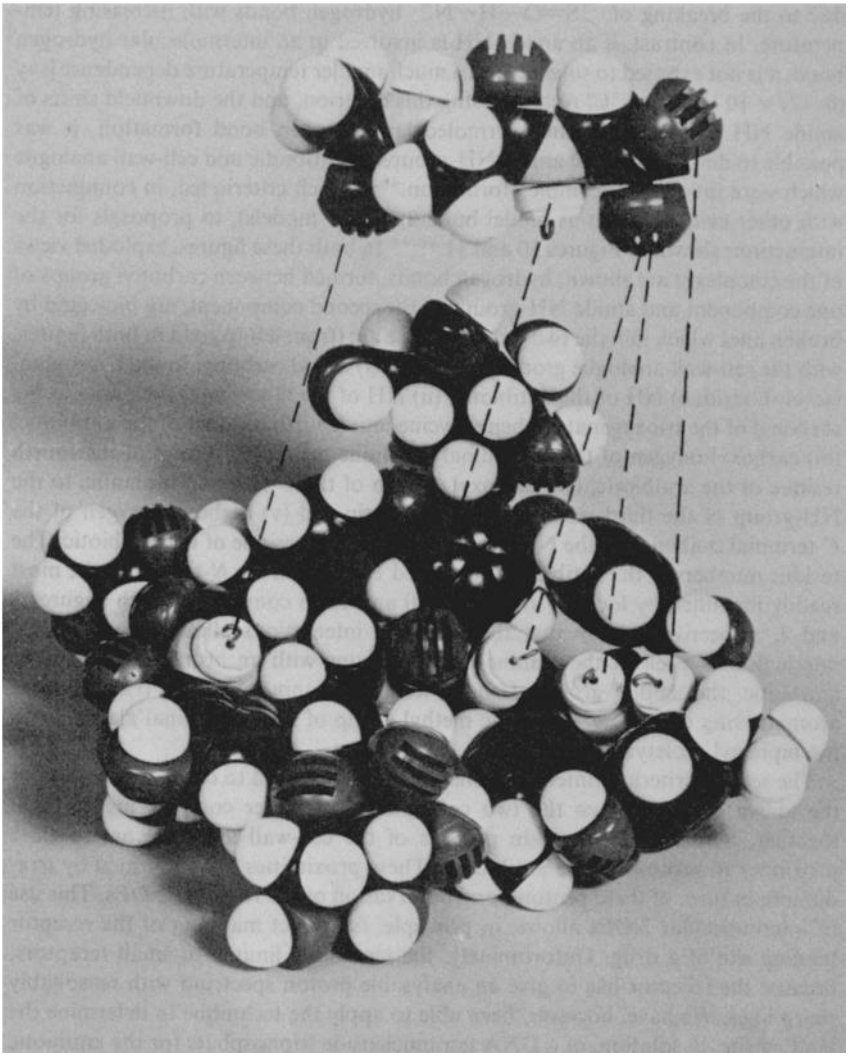


Figure 10 *Proposed hydrogen-bonding interactions between the cell-wall peptide analogue Ac-D-Ala-D-Ala (above) and ristocetin A (below, for simplicity, only the glucose unit of the tetrasaccharide is shown)*

6 Conclusion

During the last 20 years, mass spectrometry and nuclear magnetic resonance have evolved to become extremely powerful methods for the elucidation of complex organic structures. In particular, the recent advent of FAB mass spectrometry and

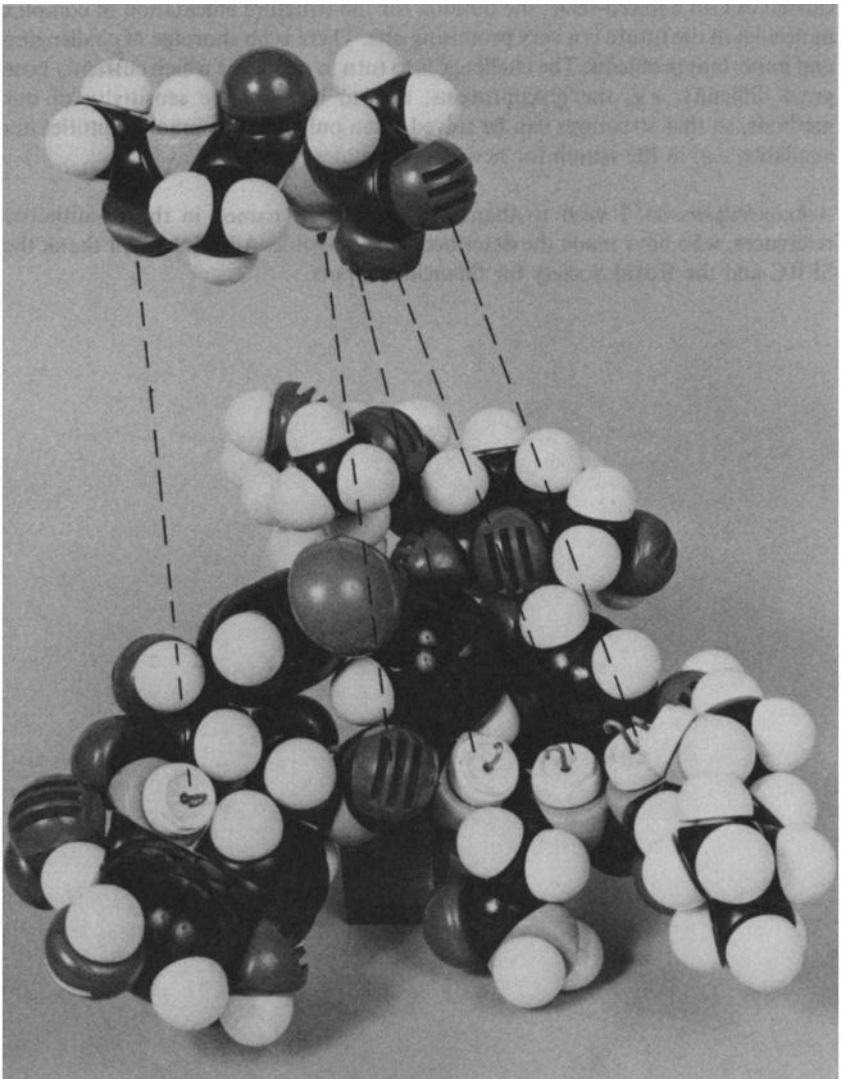


Figure 11 *Proposed hydrogen-bonding interactions between the cell-wall peptide analogue Ac-D-Ala-D-Ala (above) and vancomycin*

nuclear Overhauser effect difference spectroscopy is of importance; and it appears that two-dimensional n.m.r. spectroscopy[22] is another important advance in extracting structural information from the spectra of complex molecules.

It is concluded that if the above techniques are combined with the powerful chromatographic methods of GC and HPLC, and with chemical derivatizations

carried out on a micro-scale, the outlook for the structure elucidation of complex molecules in the future is a very promising one. There is no shortage of challenging and important problems. The challenge is to turn to problems which currently pose great difficulty, *e.g.* the glycoproteins; and to increase the sensitivity of our methods, so that structures can be solved when only nmol to pmol quantities are available, *e.g.* in the search for new neuropeptides.

Acknowledgements. I wish to thank my colleagues, named in the co-authored references, who have made the described work possible. Additionally, I thank the SERC and the Royal Society for financial support.