Structure Elucidation of the Glycopeptide Antibiotic Complex A40926

Jonathan P. Waltho and Dudley H. Williams*
University Chemical Laboratory, Lensfield Road, Cambridge, CB2 1EW
Enrico Selva and Pietro Ferrari
Lepetit Research Center, via per Rescalda 34, 21040 Gerenzamo (VA), Italy

The structure of the two factors of the novel glycopeptide antibiotic A40926 have been determined using a combination of FAB-MS, GC-MS and ^{1}H n.m.r. studies. A40926 Is a member of the vancomycin family of antibiotics and is structurally related to teicoplanin, A35512B and aridicin.

A40926 Is a complex of glycopeptides produced by the *Actinomadura* strain ATCC 39726, from which factors A and B are the major recoverable species.[1] The A40926 complex is purified from the fermentation broth by precipitation at acidic pH followed by affinity chromatography on Sepharose-D-alanyl-D-alanine.[2] Affinity to peptides terminating in D-alanyl-D-alanine is characteristic of the vancomycin family of antibiotics.[3] The single pure factors are obtained by reverse phase chromatography of the complex.[1]

A40926 Is of particular clinical interest because, in addition to having *in vitro* activity against Gram positive aerobic and anaerobic pathogens almost identical with that of teicoplanin, it possesses activity against *Neisseria gonorrhoeae*, a Gram negative coccus which is responsible for gonorrhea.[1] It is the first member of the vancomycin family to show such activity.

The structures of the two factors of A40926 have been determined using a combination of ^{1}H n.m.r. spectroscopy, fast atom bombardment mass spectrometry (FAB-MS), chemical degradation and gas chromatography-mass spectrometry (GC-MS). One and two dimensional ^{1}H n.m.r. techniques were employed including 2-D correlated spectroscopy (COSY) and nuclear Overhauser effect (n.O.e.) spectroscopy (NOESY). The strategy used was a modification of that employed previously in the structure elucidation of teicoplanin[4] in this laboratory. The vancomycin family have a good deal of structural homology, especially in the peptide portion, which is of great assistance in structural studies. Consequently, considerable reference and comparison will be made to the structures of vancomycin,[5] ristocetin,[6] and teicoplanin.[4,7,8]

Results and Discussion

Analysis by paper chromatography of the products of hydrolysis of both factors in 2M-H_2SO_4 revealed the presence of mannose and a second, unidentified, reducing sugar. Potentiometric data obtained in the mixed solvent methylcellosolve–water (4:1) indicated the presence of two free carboxyl groups in the factors.

The molecular weights of factors A and B were determined to be 1 716 and 1 730 daltons respectively by FAB-MS. The isotope pattern of each molecular ion was characteristic of a ristocetin-type peptide with two chlorine substituents.[9]

Preliminary ^{1}H n.m.r. studies indicated that, in addition to an aglycone, a long chain fatty acyl group and probably two sugars (by integration) were present. The aglycone appeared to be similar to those found in ristocetin and teicoplanin. Consequently, the fragment ions observed by FAB-MS may be assigned as shown in Table 1 with both the mannose and the second sugar (which will be referred to as the 'acid sugar' as it contains one of the carboxyl groups, see later) being terminal. The molecular weight of the aglycone portion is thus 1 211 daltons.

Table 1. Molecular ions and fragment ions observed in the positive ion FAB-MS spectra of A40926 factors A and B

Factor A	Factor B	Mass loss from MH^+	Ion
1 717	1 731	0	MH^+
1 683	1 697	34	Cl replaced by H in MH^+
	1 687	44	$MH^+ - CO_2$
1 555	1 569	162	MH^+ – mannose
1 374		343	MH^+ – acid sugar A[a]
	1 374	357	MH^+ – acid sugar B[a]

[a] Acid sugar A = acid sugar of factor A; acid sugar B = acid sugar of factor B.

It can be concluded from the FAB-MS results that the molecular weight difference between factors A and B (14 daltons) exists in the acid sugar. This is because the $[MH^+$–acid sugar$]^+$ ions of both factors have the same molecular weight.

The arguments defining the structure of the aglycone and the sugars will be detailed for one factor, factor B, and the differences occurring in factor A discussed later. The numbering system used is depicted in Figure 1, and is that employed previously for teicoplanin.[4] In this text, the numbering system will be used to describe a single proton except where stated otherwise.

Structure of the Aglycone.—The aglycone portion of A40926 was characterised by ^{1}H n.m.r. The ^{1}H n.m.r. spectrum of A40926-B in $(CD_3)_2SO$ possessed broad lines at room temperature which sharpened upon heating of the solution. However, during heating, new resonances were discernible above *ca.*40 °C, indicative of decomposition. To avoid this problem, two solvent systems were used, in which both factors of A40926 gave sharp ^{1}H n.m.r. resonances at room temperature without decomposition. These were 1:1 $(CD_3)_2SO$–CD_2Cl_2 + 0.2% CF_3CO_2D and 5:2 D_2O–CD_3CN. A spectrum of factor B in the former solvent is reproduced in Figure 2.

Double quantum filtered phase sensitive COSY (DQFCOSY) spectra of A40926-B in solvents A and B allowed the assignment of seven aromatic spin systems, including one at anomalously high field which is characteristic of ring 4 of the vancomycin family.[3] The aromatic systems consisted of one 1,4-disubstituted, three 1,2,4-trisubstituted, and three 1,2,3,5-tetrasubstituted rings. Couplings close to the diagonal were not unambiguously discernible in the absence of the double quantum filter. Other immediately observable spin systems included five NH-αCH, one NH-αCH-βCH_2, one anomeric-CH, and one anomeric-CH-NH systems.

Assignment of the protons of the peptide backbone and their relationship to the aromatic rings was achieved by using n.O.e.s, obtained from a phase sensitive NOESY experiment, in combination with the above scalar coupling information. An *N*-

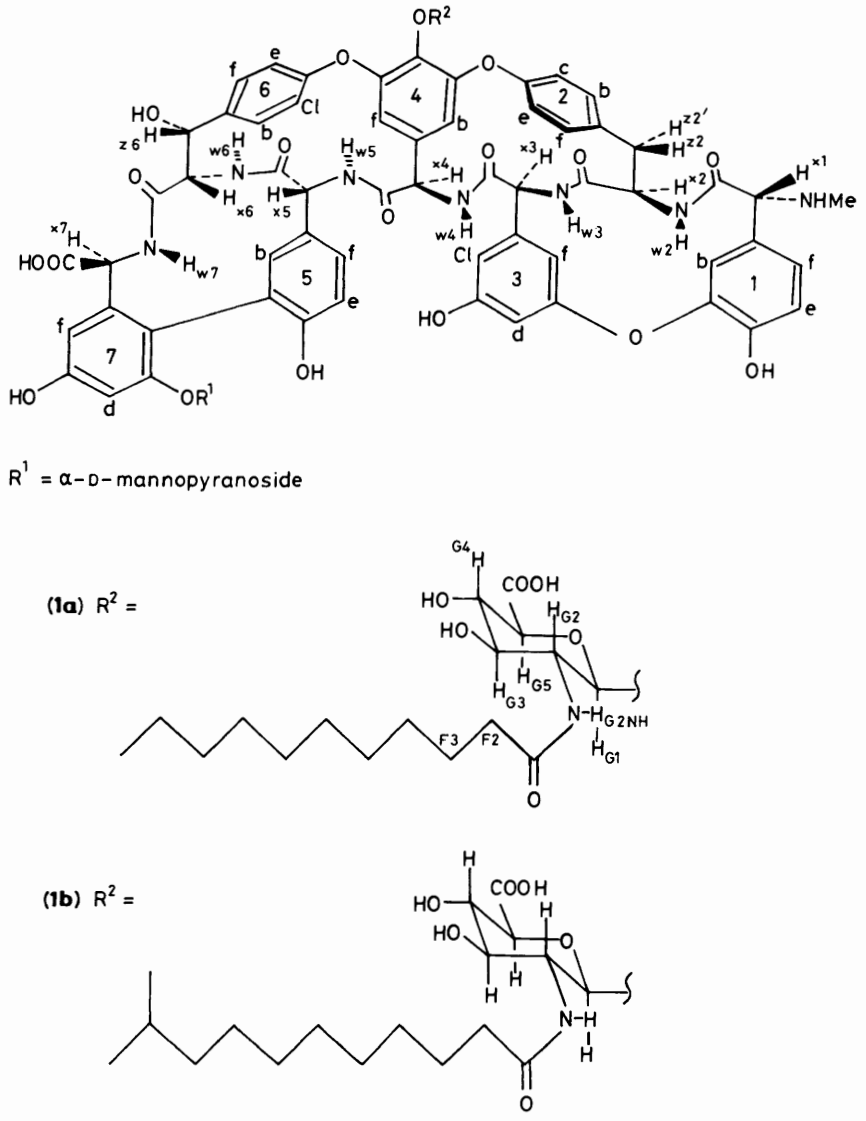

Figure 1. Structure of A40926, showing the proton nomenclature: factor A (**1a**) and factor B (**1b**)

methyl group on the *N*-terminus, identified by its characteristic downfield shift on the addition of acid, was a convenient starting point from which to base the assignments. It has an n.O.e. to 1f from which ring 1 can be assigned. The ring 3 assignments can be found from the ring 1 assignments using the n.O.e. between 1b and 3f. The *N*-methyl group also shows an n.O.e. to x1, which in turn shows an n.O.e. to w2. Scalar coupling leads through x2 to z2 and z2′ (which were misassigned in teicoplanin*) which have n.O.e.s to 2f and 2b respectively. The ring 4 assignments can be made following the ring 2 assignments using the n.O.e. between 2c and 4b.

The n.O.e. between 4f and x5 leads into the 'n.O.e. nest' (involving protons x5, 5b, x6, z6, 6b, and w7) which is characteristic of the carboxy terminus of the vancomycin family (see *e.g.* ref. 5). Hence rings 5 and 6 can be assigned. Ring 7 can be assigned using a slowly building n.O.e. between w7 and 7f. Thus the identity of all the aromatic spin systems may be derived from the methyl group of the *N*-terminus. N.O.E.s between w2, w3, and w4 complete the assignment of the NH-αCH systems. Further n.O.e.s support the above assignment, which is listed in Table 2.

The only one remaining ambiguity about the two dimensional structure of the aglycone portion was the position of the chlorine substituent of ring 3. There are two pieces of evidence to support the structure shown in Figure 1.

Firstly, the proton x3 in A40926-B is *ca.* 0.7 p.p.m. downfield of the analogous protons in ristocetin and teicoplanin. It is therefore of almost identical chemical shift with x3 in A35512-B[10] (δ 5.94 p.p.m.) which has been shown by degradation studies to have a chlorine in the 3b position.[11]

Secondly, model n.O.e. studies support the same conclusion. Proton 3f is *meta*-coupled to a single proton, *i.e.* a proton at position 3b or 3d (the other position being occupied by a chlorine atom). Examination of molecular models shows that a proton at position 3d has no near carbon-bound proton neighbours and so a fast building n.O.e. would not be expected. In contrast, a proton at position 3b should have a large n.O.e. to x3. Teicoplanin A3-2 (T-A3-2) possesses a proton at both positions 3b and 3d,[4,7] and so is a good model from which to distinguish between the two possible isomers of ring 3 in A40926-B. The ^{1}H n.m.r. spectrum of T-A3-2 in 5:2 D_2O–

* Previously misassigned in reference 4; J. P. Waltho and D. H. Williams, unpublished results.

Table 2. Chemical shifts and coupling constants of the carbon-bound protons of the aglycone portion of A40926 factor B[a] compared with analogous data for teicoplanin (T-A3-2),[4] ristocetin A,[6] and A35512B[10]

	A40926		T-A-3-2		Ristocetin A		A35512B	
Proton[b,c]	δ	*J*	δ	*J*	δ	*J*	δ	*J*
x1	5.37	s	4.6	s	4.83	s	5.50	
x2	4.84	ddd	4.97	ddd	5.09	dd	5.10	
x3	6.09	10.4	5.34	10.4	5.25	10	5.94	10.5
x4	5.49	8.2	5.60	8.2	5.65	8.2	5.56	8
x5	4.41	5.2	4.30	5	4.73	6.5	4.59	5
x6	4.08	11.5	4.13	12	4.38	12	4.31	11
x7	4.50	5.7	4.46	6.1	4.55	5	4.39	5.5
z2	2.88	13.0	2.84	13.5[d]	nc	nc	nc	nc
z2	3.32	dd	3.31	dd[d]	nc	nc	nc	nc
z6	5.06	s	5.30	s	5.17	s	5.10	s
1b	6.67	s	6.67	s	6.59	s	6.63	
1e	6.96	8.3	7.11	8.3	7.18	8	7.14	
1f	7.12	8.3	6.92	8.3	7.03	8	7.26	
2b	7.07	8.1	7.26	nc	7.12	8	7.10	
2c	7.03	8.1	nc	nc	7.29	8	7.14	
2e	6.93	8.4	7.18	8	7.25	8	7.07	
2f	7.74	8.4	7.60	8	7.86	8	7.96	8
3d	6.71	s	6.39	s	nc	nc	6.66	
3f	6.39	s	6.29	s	6.42	s	6.54	
4b	5.76	s	5.57	s	5.85	s	5.84	
4f	5.04	s	5.10	s	5.38	s	5.18	
5b	7.06	s	7.05	s	7.26	2	7.26	
5e	6.71	o	6.65	8.4	6.77	8	6.72	8
5f	6.71	o	6.70	8.4	6.84	8,2	6.74	8
6b	7.64	s	7.84	s	7.55	nc	7.50	
6e	7.19	8.5	7.21	8.4	7.20	8	7.15	
6f	7.39	8.5	7.27	8.1	7.41	8	7.41	
7d	6.75	s	6.36	s	6.85	2	6.44	
7f	6.44	s	6.29	s	6.32	2	6.06	
NMe	2.45	s	nc	nc	nc	nc	nc	nc

[a] Spectra were recorded under the same conditions as in Figure 2; s = singlet or $J < 3$ Hz, (d)dd = (doublet of) doublet of doublets, o = overlapping (second order) spin system, nc = not comparable. [b] The nomenclature is that indicated in Figure 1 and was previously introduced for teicoplanin.[4] [c] δ in p.p.m.; *J* in Hz. [d] See footnote on page 2104, left-hand column.

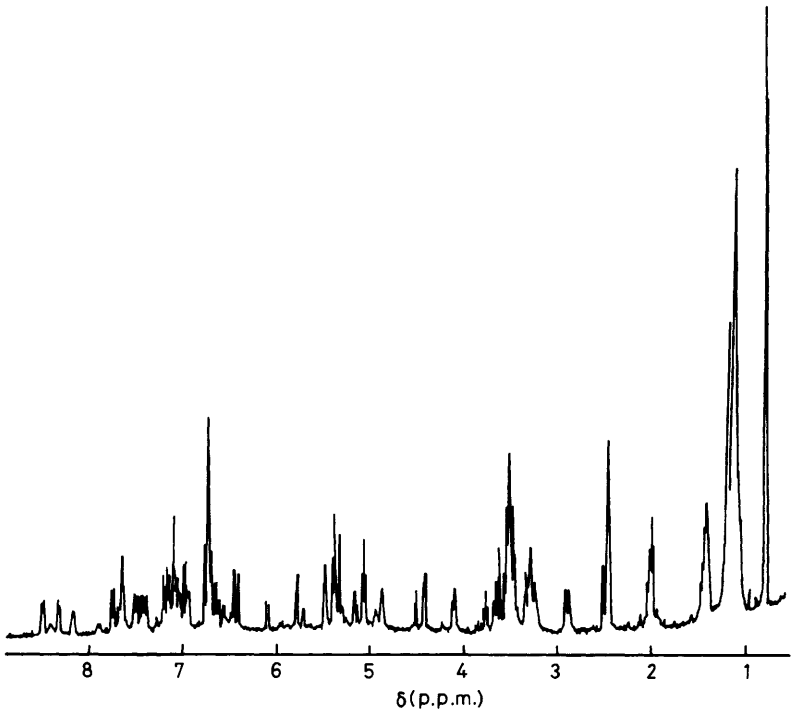

Figure 2. 400-MHz ^{1}H n.m.r. spectrum of A40926 factor B at 290 K in 1:1 $(CD_3)_2SO$–CD_2Cl_2 + 0.2% CF_3CO_2D

CD_3CN was assigned by COSY and NOESY experiments. As expected, the proton at 3b showed a strong n.O.e. to x3, whereas the proton at 3d showed no n.O.e.s. The absence of an n.O.e. in A40926-B under identical conditions from the proton *meta*-coupled to 3f is evidence that the chlorine atom occupies the 3b position and ring 3 has protons at positions 3d and 3f, as observed in A35512-B.[10]

It may be argued that the second point above is dependent on A40926-B adopting a conformation very similar to that of T-A3-2. Considering n.O.e. data from throughout the molecule, this indeed appears to be the case. However, irrespective of this, in A40926-B and T-A3-2, every phenyl glycine-derived residue with two protons *ortho* to the αCH linkage shows an n.O.e. between the αCH and one of the *ortho* protons. The proton 3f in A40926-B does *not* show an n.O.e. to x3 and so a proton at 3b would be expected to do so.

The pattern of n.O.e.s observed in A40926-B is very similar to those observed in teicoplanin[4,7,8] and ristocetin,[6] except where obvious structural differences occur. Indeed the similarity is such that, in combination with δ and $^3J_{HH}$ values (see Table 2), it can be concluded confidently that the stereochemistry of all the constituent amino acids is the same as that reported for teicoplanin and ristocetin.

In summary, the ^{1}H n.m.r. spectrum of A40926-B is consistent with a teicoplanin-type aglycone with an *N*-terminal methyl group and a chlorine at position 3b and not position 2c (see the aglycone portion of the structure reproduced in Figure 1).

Structure and Linkage Sites of the Sugars.—As mentioned above, each A40926 factor contains two sugars, mannose and an acid sugar, the latter being at this point uncharacterised. The non-exchangeable protons of both sugars may be assigned in the ^{1}H n.m.r. spectra of A40926-B, using the aforementioned DQFCOSY and NOESY spectra. The anomeric proton of mannose has been named 'M1', with the proton bonded to the adjacent carbon, 'M2', and so on around the ring to the geminal protons 'M6' and 'M6'. The nomencalture of the acid sugar is given in Figure 1.

In combination with the above FAB-MS results, the identity of one sugar unit as mannose may be confirmed by the values of the $^3J_{HH}$ coupling constants around the ring (*e.g.* M1-M2 and M2-M3 have $^3J_{HH} < 3$ Hz). A strong n.O.e. between the anomeric proton M1 and the aglycone proton at 7d (and not between the anomeric proton and 7f) indicates that mannose is glycosidically linked to the phenolic oxygen at position 7c, as observed in teicoplanin and ristocetin. The presence of an n.O.e. between the protons at M1 and M2, but not between those at M1 and M3, is evidence that mannose is present as the α anomer* (D stereochemistry is assumed, by analogy to teicoplanin[4,7] and ristocetin[6,12]).

The second sugar is made up of two parts, a pyranose ring and a fatty acyl group. The anomeric proton G1, assigned on account of its chemical shift (5.36 p.p.m.), possesses a 9.7 Hz coupling with G2, indicative of a *trans*-diaxial relationship between them. Proton G2 has a 7.2 Hz coupling with an amide NH (G2NH) and a *ca.* 8 Hz coupling with G3. Proton G4 is linked to G2, and proton G5 to G1 and G3 by fast building n.O.e.s, and so it is proposed that all the pyranose CH protons are axial. Hence, the pyranose has the same stereochemistry around the ring as glucose.

An n.O.e. between G2NH and F2 links the fatty acyl chain to the sugar amide nitrogen. The chemical shift of F1 (2.00 p.p.m.) is consistent with such an acylation. The precise nature of the fatty acyl chain was determined by GC-MS using the method of Barna *et al.*[4] Acid hydrolysis of A40926-B using a two-phase system (2M-HCl–n-hexane), followed by esterification of the resulting carboxylic acid with diazomethane and derivatisation with pyrrolidine, produced a single fatty acid *N*-pyrrolidide by GC. A Molecular ion ($M^{+\cdot}$) at *m/z* 253 indicated that the fatty acyl group was a derivative of dodecanoic acid. A fragment ion at $M - 15$, no ion at $M - 29$, but subsequent fragment ions at 14 a.m.u. intervals reducing in mass from $M - 43$, is characteristic of a linear long-chain fatty acid with an iso terminus.[13] Further evidence for an isoterminus to the fatty acyl group is present in the ^{1}H n.m.r. spectrum where there is a six proton high-field methyl doublet with a $^3J_{HH}$ coupling of 6.5 Hz.

The masses of the fragment ions observed by FAB-MS indicate that the molecular weight of the pyranose portion of the acid sugar is 16 a.m.u. greater than that of a glucosamine derivative. This is consistent with the second carboxy group (identified above using potentiometric titrations) constituting part of the pyranose system. Again assuming D-stereochemistry for the sugar, the ^{1}H n.m.r. data are characteristic of the β anomer of a 2-amino-2-deoxyglucuronic acid derivative, where the 2-amino group is acylated with isododecanoic acid.

The site of linkage of the acid sugar to the aglycone was more difficult to define by ^{1}H n.m.r. than that of mannose. No large n.O.e. was observed between the anomeric proton G1 and the aglycone. Such a situation would be expected only if the acid sugar were linked to the *para* phenolic oxygen of ring 4 (position 4d). Careful irradiation of G1 and the aromatic proton 2e showed that they possessed a mutual slow building n.O.e. The site of linkage of the acid sugar is therefore proposed to be at the position 4d.

In summary, the aglycone of A40926-B has two sugars glycosidically linked to it. One is α-D-mannopyranose linked *via* position 7c and the other is consistent with 2-deoxy-2-[(1-oxo-10-methylundecyl)amino]-β-D-glucopyranosiduronic acid linked *via* position 4d. The derived structure (**1b**) of A40926-B is reproduced in Figure 1.

Structure of Factor A.—Factor A of A40926 was analysed in an identical way with that for factor B above. From the FAB-MS results (Table 1) it was known that the difference between the factors (in terms of the molecular weight at least) lay within the acid sugar. ^{1}H N.m.r. studies confirmed that the aglycone portions, and the linkage of mannose, in both factors are identical. Analysis of the fatty acid group *N*-pyrrolidide by GC-MS confirmed that the fatty acid group was 14 a.m.u less in molecular weight in factor A than factor B. Moreover, the fragmentation of the molecular ion in EI-GCMS was characteristic of a non-branched fatty acyl chain.[13] That is, the mass spectrum showed a molecular ion ($M^{+\cdot}$), a lower intensity ion at $M - 15$, and high intensity ions at $M - 29$, $M - 43$, $M - 57$, *etc.* Again the molecular weight of the pyranose ring of the acid sugar was indicative of the presence of a free carboxy group. ^{1}H N.m.r. confirmed that all the pyranose CH protons are axial and that the linkage to the aglycone is *via* position 4d. It is therefore proposed that the acid sugar in A40926-A is 2-deoxy-2-[(1-oxoundecyl)amino]-β-D-glucopyranosiduronic acid. Hence, the proposed structure for A40926-A is (**1a**), shown in Figure 1.

The proposed structures for the factors of A40926 are similar, though not identical, to the aridicin family of glycopeptide antibiotics reported recently.[14]

Experimental

FAB-MS mass spectra were recorded on a Kratos MS-50 instrument fitted with a standard FAB source and a high-field magnet. The sample (*ca.* 10 nmol) was dispersed in a few microlitres of a matrix (typically 1:1 α-thioglycerol–diglycerol) and bombarded with a 6—9 keV beam of Xe atoms. GC-MS

* NOESY spectra were run on teicoplanin A3-2 and ristocetin A under identical solvent conditions with those used for A40926.

was performed using a column of SE-54 and a programmed temperature rise (100—280 °C) in a Finnigan 4000 instrument. Mass spectra were recorded at 1-s intervals using EI (70 eV) ionisation.

^{1}H N.m.r. spectra were recorded on Bruker AM500 and WM400 spectrometers equipped with Aspect 3000 and 2000 computers respectively. Solutions of A40926 were approximately 15mM in 1:1 $(CD_3)_2SO$–CD_2Cl_2 with 0.2% (v/v) CF_3CO_2D or 5:2 D_2O–CD_3CN. Before dissolution in the last-named solvent system, samples were dissolved in D_2O and lyophilised to replace the exchangeable protons. Quadrature detection was used throughout with spectral widths typically 5 000 or 4 000 Hz, depending on the spectrometer used. Smaller spectral-width COSY and NOESY experiments were run to reduce the sampling time whilst retaining resolution. The built-in electronic filter was set to the desired spectral width to minimise 'folded' peaks in f_2. Spectra needed to be analysed with some care but 'folded' crosspeaks could, if necessary, be identified by re-recording the spectrum with a small change in the spectral width.

COSY, DQFCOSY and NOESY spectra were run in the phase sensitive mode using time-proportional-phase-incrementation in f_1. The carrier was placed in the centre of the spectrum. In general for DQFCOSY and NOESY experiments covering *ca.* 10 p.p.m., 1 024 or 512 2K spectra were recorded, each consisting of 16 or 32 scans. These figures were correspondingly smaller for the smaller spectral-width experiments. The relaxation delay was approximately twice the longest t_1 value and each 2D spectrum took about 12 hours to record. Data sets were multiplied by Lorentzian–gaussian functions and zero filled once in f_1 before Fourier transformation.

Identification of the Fatty Acid Group.—A40926-B (5 mg) was hydrolysed in aqueous HCl (2M; 10 ml) under n-hexane at 80 °C overnight. The cooled mixture was separated and the hexane layer was washed with H_2O (5 ml) and dried. Excess of ethereal diazomethane was added (5 min) and then removed under a slow stream of dry N_2. The methyl esters were analysed by GC and GC-MS. Factor A was treated similarly.

Solutions of diazomethane in diethyl ether were prepared from Diazald–KOH.

A sample of the methyl ester (20 μg) was heated in pyrrolidine–AcOH (10:1; 0.5 ml) for 40 min at 100 °C in a sealed vessel. The cooled mixture was extracted with n-hexane (2 × 2 ml) and the combined extracts were washed with dilute HCl (0.1M; 2 × 500 μl) and H_2O (500 μl), dried and analysed by GC and GC-MS.

Acknowledgements

We gratefully acknowledge Dr. E. Houghton and P. Teale of Racecourse Securities Laboratory, Newmarket, for access to GC-MS facilities and assistance in obtaining data, and the S.E.R.C. and Gruppo Lepetit for financial support.

References

1 G. Cassani, E. Selva, L. Gastaldo, P. Ferrari, B. Goldstein, R. Pallanza, M. Berti, V. Arioli, and F. Parenti, 26th ICAAC, Abstract No. 943, 1986.
2 A. Corti and G. Cassani, *Appl. Biochemistry Biotechnology*, 1985, **11**, 101.
3 For a recent review see: J. C. J. Barna and D. H. Williams, *Ann. Rev. Microbiol.*, 1984, **38**, 339.
4 J. C. J. Barna, D. H. Williams, D. J. M. Stone, T.-W. C. Leung, and D. M. Doddrell, *J. Am. Chem. Soc.*, 1984, **106**, 4895.
5 M. P. Williamson and D. H. Williams, *J. Am. Chem. Soc.*, 1981, **103**, 6580.
6 J. R. Kalman and D. H. Williams, *J. Am. Chem. Soc.*, 1980, **102**, 897.
7 A. H. Hunt, R. M. Molloy, J. L. Occolowitz, G. G. Marconi, and M. Debono, *J. Am. Chem. Soc.*, 1984, **106**, 4891.
8 P. Ferrari, J. Kettenring, G. G. Gallo, and B. Cavalleri, *J. Antibiot.*, 1987, in press.
9 G. D. Roberts, S. A. Carr, S. Rottschaefer, and P. W. Jeffs, *J. Antibiot.*, 1985, **38**, 713.
10 A. H. Hunt, *J. Am. Chem. Soc.*, 1983, **105**, 4463.
11 C. M. Harris and T. M. Harris, *Tetrahedron, 1983*, **39**, 1661.
12 N. N. Lomakina, L. I. Muravyeva, and M. Puskas, *Antibiotiki*, 1968, **13**, 867.
13 B. A. Anderson and R. T. Holman, *Lipids*, 1975, **10**, 716.
14 (*a*) P. W. Jeffs, L. Mueller, C. DeBrosse, S. L. Heald, and R. Fisher, *J. Am. Chem. Soc.*, 1986, **108**, 3063; (*b*) R. D. Sitrin, G. W. Chan, F. Chapin, A. J. Giovenella, S. F. Grappel, P. W. Jeffs, L. Phillips, K. M. Snader, and L. J. Nesbit, *J. Antibiot.*, 1986, **39**, 68.

Received 30*th October* 1986; *Paper* 6/2106