Collagen Cross-links

A MASS-SPECTROMETRIC AND ^1H - AND ^{13}C -NUCLEAR-MAGNETIC-RESONANCE STUDY

By ERIC HUNT and HOWARD R. MORRIS

University of Cambridge Chemical Laboratory, Lensfield Road, Cambridge CB2 1EW, U.K.

(Received 11 July 1973)

1. The structure of a tetrafunctional cross-link (compound C) isolated from borohydride-reduced cow skin collagen has been determined by the combined use of mass spectrometry and ^1H and ^{13}C n.m.r. spectroscopy.
2. A new technique, of potentially wide applicability, is described for the mass-spectrometric determination of functional groups.
3. Detailed study of protonation shifts in the ^{13}C spectrum has allowed an unambiguous assignment of the exact structural relationship of the component parts of the molecule.
4. New interpretations of existing data on this molecule are given.

Significant progress has been made in recent years in the isolation and characterization of compounds derived from collagen and believed to act as intermolecular cross-links in the stabilization of collagen fibres. Several examples of difunctional cross-links have been reported (Paz *et al.*, 1969; Bailey *et al.*, 1969; Bailey & Peach, 1968; Tanzer *et al.*, 1970), and recently Tanzer *et al.* (1973) have proposed a structure for a new tetrafunctional cross-link isolated from cow skin collagen.

The importance of spectroscopic techniques in the structure elucidation of these compounds has increased with their complexity, and it is now essential for the biochemist to be aware of their potential when studying biologically important molecules. We now report the results of a study involving mass spectrometry and ^1H and ^{13}C n.m.r. spectroscopy on a potential cross-link isolated by Bailey *et al.* (1970) and described by these authors as 'compound C'. Our structure is entirely based upon spectroscopic data; the only biochemical evidence available (apart from its isolation) was the possible involvement of allysine aldol in the formation of compound C (A. J. Bailey, personal communication; Bornstein & Piez, 1966; Kang *et al.*, 1970).

Our independent study has led to the full structure elucidation of compound C which we find agrees with the structure proposed for the 'post-histidine' compound described by Tanzer *et al.* (1973). The work described in the present paper has differed from that of the latter authors in several important respects. Far stronger evidence for the exact structural features of the cross-link and in particular the position of attachment of the imidazole portion, is provided from our ^{13}C n.m.r. studies, which were greatly aided by the use of off-resonance decoupling and protonation shifts. This contrasts with the emphasis placed by Tanzer *et al.* (1973) on the ^1H n.m.r. spectrum, and we find

ourselves in disagreement with several of the interpretations made by these authors. In our opinion the complex methine and methylene region in the ^1H spectrum cannot be of direct use in the structure assignment of the major portion of the molecule. Our choice of derivatives for the mass-spectrometric study included a new mass shift technique, which has provided a rapid and simple method for the identification of structurally significant signals, and immediate recognition of the number of derivatized functional groups not only in the molecular ion but in each fragment ion in the spectrum. This new method greatly aids the interpretation of the mass spectrum, and comparison with the data of Tanzer *et al.* (1973) illustrates this point (see the Results section and the Discussion section).

The study presented here provides considerably stronger evidence for the structure proposed by Tanzer *et al.* (1973), and in addition the methods described should prove useful in spectroscopic studies on other biologically important molecules.

Experimental

Preliminary studies indicated that acetylpermethyl derivatives did not produce good mass spectra, but that fluoroacetyl esters were satisfactory.

The compound was esterified by using a pH 1 solution prepared by bubbling dry HCl gas into methanol- $[\text{P}_2\text{H}_4]$ methanol (1:1, v/v). The ester was isolated by evaporation of reagents after 12 h at room temperature.

Trifluoro- and difluoro-acetylations were carried out on the ester by using a 1:1 mixture of the respective anhydride and acid, at room temperature for 3 h followed by evaporation of reagents under vacuum. Difluoroacetic anhydride was prepared by refluxing difluoroacetic acid with P_2O_5 for 3 h, followed by distillation from P_2O_5 (b.p. 125°C).

H-²H-exchange experiment: compound C (0.5 mg) was allowed to stand in ²H₂O (2 ml), pH 5.5, at 20°C for 3 days, followed by freeze-drying. Derivative formation was then carried out as above except for the use of deuterio trifluoroacetic acid in the acylation step. The resulting mass spectrum showed incorporation of ²H into imidazole-containing fragments (see the Results section).

All low-resolution mass spectra were recorded on a G.E.C.-A.E.I. MS902 instrument operating at 6.6 kV accelerating voltage, and a resolving power of 1400. High-resolution mass measurements were made on the same instrument, with perfluorokerosene as reference, operating at 10000 resolving power, on-line to a PDP8 (DS10) data acquisition system. Atomic compositions were calculated for masses within ± 5 p.p.m. of the measured mass.

Proton noise-decoupled pulsed Fourier transform ¹³C n.m.r. spectra were obtained at 25.2 MHz for ²H₂O solutions of compound C (0.1 M and 0.15 M), 1-methylhistidine (0.8 M), 3-methylhistidine (0.1 M) and δ -hydroxylysine (0.2 M) by using a Varian Associates XL-100 Fourier transform n.m.r. system. The pH values of the solutions were adjusted by using NaO²H-²H₂O (2:3, w/w) and ²HCl-²H₂O (1:4, w/w) solutions. Chemical shifts (δ) are reported in p.p.m. downfield from the resonance of an external tetramethylsilane reference. ¹H n.m.r. spectra of compound C in ²H₂O solution (0.1 M; pH 5.5) were obtained at 100 MHz on a Varian Associates HA-100 instrument.

Results

¹H n.m.r.

The proton n.m.r. spectrum of compound C in ²H₂O (for a description of this spectrum see Tanzer *et al.*, 1973) showed two low-field equi-intense singlets at δ 7.01 and 7.75.

When the compound was left in ²H₂O, pH 5, 20°C, and the spectrum was re-run, the lowest field signal decreased in intensity, and disappeared completely within 3 days; all other signals in the spectrum remained unchanged. Slow H-²H exchange has been observed for the C-2 imidazole proton of histidine by Meadows *et al.* (1968). We thus assign the signals at δ 7.01 and 7.75 to the C-4(5) and C-2 protons of a C-5(4) substituted imidazole function. The occurrence of a triplet ($J = 5.5$ Hz) at δ 3.91 (α -CH, histidine; McDonald & Phillips, 1969), which integrates for approximately one proton (taking the signal at δ 7.01 as one proton), is strongly suggestive of the presence of a histidine moiety in compound C, in agreement with the mass-spectral data.

The doublet ($J = 6$ Hz) of triplets ($J = 7.5$ Hz) centred at δ 4.15 integrates for one proton, and we assign this to a methine group attached to an

imidazole N atom (cf. methyl protons of methyl-histidines at δ approx. 3.7). The coupling to this proton shows that the methine group is attached to one methylene group and one methine group, indicating the structural unit formula (I):

The complex pattern between δ 3.44 and 3.80 integrates for approximately four protons. The protons giving rise to this pattern can be assigned to methine groups of α -amino carboxylic acid functions and/or to methylene groups and/or methine groups attached directly to oxygen.

The remainder of the spectrum was structurally uninformative because of considerable overlapping of the signals. These may have been resolvable at higher frequencies.

Mass spectrometry

The mass spectrum of the trifluoroacetylmethyl ester of compound C is shown in Fig. 1. The use of the new mass shift technique (see the Discussion section) allows a clear assignment of the number of ester groups (derived from carboxylic acids) in each fragment in the one spectrum. The 1:4:6:4:1 pattern at m/e 1205 (²H₀), 1208 (²H₃), 1211 (²H₆), 1214 (²H₉) and 1217 (²H₁₂), the molecular ions, shows the definite presence of 4 carboxyl groups in compound C. The assignment of m/e 1205 as a (non-deuterated) molecular ion is corroborated by typical losses of F⁺ at m/e 1186 (1:4:6:4:1), ⁺OCH₃ at m/e 1174 (1:3:3:1) and ⁺CO₂CH₃ at m/e 1146 (1:3:3:1).

The mass spectrum of the difluoroacetylmethyl ester shows molecular ions at m/e 1097, 1100, 1103, 1106 and 1109 (1:4:6:4:1) which are 108 mass units lower than those in Fig. 1. Since the difluoroacetyl group is 18 mass units lighter than the trifluoroacetyl one, compound C clearly contains 6 acylable groups. One of these is a hydroxyl function as evidenced by the facile loss of CF₃CO₂H to give m/e 1091 and associated ions in Fig. 1. The absence of multiple losses of CF₃CO₂H suggests that the remaining acylable groups are primary or secondary amino functions.

Since the molecular ion (²H₀) in Fig. 1 occurs at an uneven mass number, compound C must contain an odd integral number of N atoms.

Fig. 1. Low-resolution mass spectrum (above m/e 200) of the trifluoroacetyl methyl ester derivative of compound C
For details see the text.

Summarizing the ^1H n.m.r. and mass-spectrometric data at this point, a strong possibility for the functional groups present in compound C is:

- (a) $4 \times \text{NH}_2\text{—CH—CO}_2\text{H}$ } giving 5 protons in the
 (b) $1 \times \text{>CHOH}$ } 83.4 $\rightarrow$ 4.0 region

- (d) $1 \times$ acylable N (1° or 2°)

Signals at m/e 265, 268 and 266, 269 in Fig. 1 are assigned to a histidine moiety (cleavage with H transfer at the C-10–N-11 bond of Fig. 2) by the following evidence.

- The 1:1 doublet indicates 1 CO_2CH_3 group.
- The corresponding signals in the spectrum of the difluoro derivative at m/e 247, 250 and 248, 251 show the presence of 1 acylable group.
- Accurate mass measurement gives the composition $\text{C}_9\text{H}_{10}\text{N}_3\text{O}_3\text{F}_3$ for m/e 265.
- Characteristically for histidine the intensities of the doublets vary with source pressure.

- Incorporation of ^2H in the H– ^2H exchange experiment (see the Experimental section) is consistent with the presence of imidazole.

The singlet signal at m/e 206 and its difluoro analogue at m/e 188 are derived by loss of CO_2CH_3 from the histidine moiety, and again accurate mass measurement substantiates this.

The signals at m/e 503, 506 and 509 and m/e 504, 507 and 510 are assigned to a larger histidine-containing fragment (corresponding to cleavage at the C-9–C-10 bond in Fig. 2) by the following evidence.

- The 1:2:1 triplet indicates 2 CO_2CH_3 groups.
- The corresponding signals in the spectrum of the difluoro derivative at m/e 467, 470 and 473 show the presence of 2 acylable groups.
- Characteristically for histidine, the intensities of the triplets vary with source pressure.

- Incorporation of ^2H in the H– ^2H exchange experiment is consistent with the presence of imidazole.

(v) Several possible atomic compositions were compatible with the accurate mass measurement for m/e 503, but considering the number of heteroatoms necessary to satisfy points (i)–(iv) we are left with only one possibility, $\text{C}_{18}\text{H}_{21}\text{N}_4\text{O}_6\text{F}_6$.

The signals at m/e 363 and 366 are assigned to a fragment derived from a hydroxylysine moiety by the following evidence.

(i) The 1:1 doublet shows the presence of only 1 CO_2CH_3 group.

(ii) The corresponding signals in the spectrum of the difluoro derivative show the presence of 2 acylable groups.

(iii) Accurate mass measurement gives an atomic composition of $\text{C}_{12}\text{H}_{13}\text{N}_2\text{O}_4\text{F}_6$ for m/e 363.

(iv) The doublet is the only signal not affected during the $\text{H}-^2\text{H}$ exchange experiment, showing the absence of imidazole.

Subtracting $2 \times \text{CF}_3\text{CO}$ and $1 \times \text{CO}_2\text{CH}_3$ from the above atomic composition leaves only N as a heteroatom, showing that both acylable functions are amines. The data suggests a fragment derived by cleavage at the C-8-C-9 bond (Fig. 2) and loss of $\text{CF}_3\text{CO}_2\text{H}$ as expected from a trifluoroacetate.

The signals at m/e 728, 731, 734 and 737 are also assigned to cleavage at the C-8-C-9 bond, but with charge retention on the histidine-containing fragment, by the following evidence.

(i) The 1:3:3:1 quartet indicates 3 CO_2CH_3 groups.

(ii) The corresponding signals in the spectrum of the difluoro derivative are present (although not intense) at m/e 674, 677, 680 and 683 indicating the presence of 3 acylable groups.

(iii) Incorporation of ^2H in the $\text{H}-^2\text{H}$ exchange experiment shows the presence of imidazole in the fragment.

(iv) Considering the derivation of m/e 363 described earlier (i.e. from m/e 477 by loss of $\text{CF}_3\text{CO}_2\text{H}$) we see that summing the two fragments $477+728$ gives the molecular weight of derivatized compound C, i.e. 1205. This strongly suggests that both fragments are formed by cleavage of the same bond (C-8-C-9) with charge retention on either part of the molecule.

(v) Comparison with m/e 503 shows the presence of 1 more carboxyl and 1 more acylable group suggesting an extra $\text{CF}_3\text{CO}-\text{NH}-\text{CH}-\text{CO}_2\text{CH}_3$ unit. This leaves only 41 mass units to be accounted for in the difference between m/e 728 and m/e 503 at this stage.

(vi) Accurate mass measurement of m/e 728 gives several possible atomic compositions, including $\text{C}_{26}\text{H}_{31}\text{N}_5\text{O}_9\text{F}_9$, which is compatible with other data on m/e 728.

Fig. 2. Proposed structure of compound C

Scheme 1. Partial structure of compound C indicated by mass spectrometric and ^1H n.m.r. evidence

Numbers refer to the masses of principal fragment ions. For details see the text.

A number of other cleavages, particularly at high mass, are consistent with simple losses from various fragments, and with a few exceptions are not particularly structurally informative.

Thus far the ^1H n.m.r. and mass-spectrometric evidence is consistent with the structure shown for compound C in Scheme 1, where X is derived from an α -amino acid, probably lysine.

^{13}C n.m.r.

The ^{13}C n.m.r. spectrum of compound C is shown in Fig. 3. Single-frequency proton decoupling experiments (Archer *et al.*, 1970) at frequencies corresponding to δ 7.91 and 4.0, at pH 5.0, and to δ 3.5 and 3.0, at pH 3.5, in the proton n.m.r. spectrum made it possible to assign the ^{13}C signals to methylene, methine and fully substituted carbons (no methyl C atoms found). These data, together with the mass-spectral and proton n.m.r. studies, allowed us to propose the structure shown in Fig. 2 for compound C (with the exception that at this point in the analysis the position of attachment of C-10 at the imidazole ring was undetermined).

The calculated chemical shifts for the C atoms in the structure shown are compared with the observed ^{13}C shifts in Table 1. The calculated shifts were based on the observed shifts for lysine hydrochloride (Johnson & Jankowski, 1972) (C-8 to C-27; C-10 to C-23), δ -hydroxylysine hydrochloride (C-1 to C-6) and 3-methylhistidine hydrochloride (C-10 to C-18), by using the parameters of Grant & Paul (1964) for alkyl substitution and the parameters of Horsley *et al.* (1970) for amino substitution. For a change of primary to secondary amino group we have used the following parameters based on ^{13}C

Table 1. Observed and calculated ^{13}C chemical shift for compound C in $^2\text{H}_2\text{O}$ (pH 5.0)

The letters in parentheses refer to the observed multiplicity in off-resonance decoupling experiments; s, singlet; d, doublet; t, triplet.

Observed (p.p.m.)	Assignment	Calculated (p.p.m.)
175.1 (s)	C-1	175.4
174.9 (s)	C-27	175.4
173.9 (s)	C-23	175.4
	C-18	173.3
138.3 (d)	C-12	139.5
136.0 (s)	C-14	135.3
117.6 (d)	C-15	120.4
67.05 (d)	C-5	66.6
59.7 (d)	C-10	58.7*, 60.0†
	C-17	56.15
55.7 (d)	C-2	55.6
55.5 (d)	C-26	55.5
55.3 (d)	C-22	55.2
54.0 (t)	C-6	54.3
48.9 (t)	C-8	48.5
41.3 (d)	C-9	42.2
32.05 (t)	C-19	29.6†
30.95 (t)	C-4	31.1
	C-21	27.6
29.5 (t)	C-16	30.0
27.7 (t)	C-3	27.8
	C-25	25.1
24.5 (t)	C-24	22.4
22.2 (t)	C-20	20.4

* Based on 3-methylhistidine.

† Based on lysine hydrochloride.

Fig. 3. ^{13}C n.m.r. spectrum of compound C

Proton noise is decoupled. For further details see the text.

shift data for a variety of primary and secondary amines (Johnson & Jankowski, 1972; Stothers, 1972): at the α -C atom a downfield shift of 8.2 p.p.m.; at the β -C atom an upfield shift of 3.4 p.p.m.; and at the γ -, δ -C atoms etc. no change (small shifts are expected, but there is at present insufficient data available for reliable calculations at these positions) The calculated shifts and the observed shifts for compound C are generally in good agreement.

Of central importance to the assignment of signals in the ^{13}C n.m.r. spectrum was the measurement of protonation shifts i.e. changes in chemical shift with pH. A comparison of the protonation shifts for compound C with those for the methylhistidines unequivocally demonstrates that in compound C C-10 is attached to N-11 rather than to N-13. The protonation shifts for compound C, 1-methylhistidine,

and 3-methylhistidine are shown in Figs. 4, 5(a) and 5(b). As can be seen from the data for the methylhistidines, the protonation shifts for the C atoms of the imidazole ring are diagnostic for the position of *N*-alkylation. The protonation shifts of the corresponding C atoms in compound C are in line with those of 3-methylhistidine, thus showing that C-10 is attached to N-11, as shown in Fig. 2. Further, the C atom of the β -methylene group of 1-methylhistidine generally appears at higher field than the corresponding C atom of 3-methylhistidine, a fact which undoubtedly reflects the greater steric crowding of this group in 1-methylhistidine (Woelfenden & Grant, 1966; Grant & Cheney, 1967). If, in compound C, C-10 were attached to N-13, then the steric crowding of the β -methylene group of the histidine unit would be expected to be such that the C-16 resonance

Fig. 4. Variation in ^{13}C chemical shifts with pH for compound C

(a) and (b) are for different C atoms and different values of δ . For details see the text. For C atom numbering see Fig. 2.

Fig. 5. Variation in ^{13}C chemical shifts for (a) 1-methylhistidine, (b) 3-methylhistidine and (c) δ -hydroxylysine

For details see the text. Numbering of atoms of the methylhistidines is in accordance with the usual biochemical nomenclature, i.e. the imidazole N atoms are numbered 1 and 3, the amino acid side chain (C-6, C-7 and C-8) is attached to C-5, C-9 is the C atom of the N-methyl group. Greek letters are used in accordance with the chemical and biochemical nomenclature to label the positions in δ -hydroxylysine, i.e. C- α is the C atom bonded to the carboxyl and amino group, C- β is the next C atom along the chain, etc.

would appear at even higher field than in 1-methylhistidine. In Fig. 4(b) C-16 was readily assigned from its characteristic protonation-shift curve, which shows an inflexion at approx. pH 8.5, owing to protonation of the C-17 amino group, and at approx. pH 4, owing to protonation of the imidazole function. The chemical shifts shown by C-16 on this curve agree closely with the corresponding values for 3-methylhistidine, again pointing to C-10 being attached to N-11.

The curious S-shaped protonation-shift curve shown by the methyl C atom of 1-methylhistidine appears to be characteristic for a C atom in this environment, since it is also seen for C-10 in Fig. 4(b). (3-Methylhistidine may also show this effect, but, because of the considerable amount of time required to obtain spectra from our relatively small sample of this compound, there is at present insufficient data to completely define its protonation-shift curve). We are unable at present to give a full interpretation of this observation, but the fact that the S-shaped curve for 1-methylhistidine is a reversal of that for compound C is consistent with the observed reversal of α -C protonation shifts in passing from α -methylene to α -methine in simple primary amines (Morishima *et al.*, 1973).

The protonation shifts for the other C atoms of compound C could be rationalized in terms of the known effects of protonation on the ^{13}C chemical shifts of amines and amino acids (Horsley *et al.*, 1970; Gurd *et al.*, 1971; Morishima *et al.*, 1973). Thus the chemical shifts of the carboxyl C atoms move to higher field on protonation of the α -amino groups and of the carboxylate anions. At two pH values (11.5 and 6.8) four separate signals were observed for the four carboxyl groups of compound C. The protonation-shift curves for C-5, C-6, C-8 and C-9 all show a single inflexion at approx. pH 8.5 associated with protonation of the secondary amino N. Characteristically, the C atoms (C-5 and C-9) which are β to N-7 show larger protonation shifts than the C atoms α to N-7.

The protonation-shift curve for C-2, C-17, C-22 and C-26 shows an inflexion at approx. pH 9 associated with the protonation of the α -amino groups and the start of an inflexion at low pH where the carboxylate anions are becoming protonated. We assign four C atoms to this curve since four α -amino carboxylic acid functions are known to be present from the mass-spectral and n.m.r. data. However, at most, only three separate signals (separation 0.2–0.6 p.p.m.) could be found in this

region of the spectrum (pH 10.5, 5.0, 3.5). The protonation-shift curve for C-21 shows an inflexion at approx. pH 9 owing to protonation of its β -amino group. This curve is paralleled by the curve for C-3 and C-25 at approx. 3 p.p.m. higher field. The separate signals (separation 0.1–0.2 p.p.m.) for C-3 and C-25 were observed at pH 9 and pH 1.

C-19 shows a shift of 1.8 p.p.m. to higher field upon protonation of the imidazole function. The signal owing to C-4 was never clearly seen, but from a consideration of the protonation-shift curve for the γ -C atom of δ -hydroxylysine (Fig. 5c), the protonation-shift curve for C-4 is expected initially (i.e. at high pH) to follow the curve for C-16 and then to cross over at approx. pH 9 to follow the curve for C-21 to lower pH values. C-24 shows a small overall protonation shift of 1.4 p.p.m. which is presumably associated with protonations of its two γ -amino groups. C-20, which has only one γ -amino group, shows an overall protonation shift of 0.7 p.p.m.

At certain pH values the signals assigned to C-5, C-8, C-9, C-19, C-20 and C-24 appear as two closely spaced equi-intense singlets. This observation undoubtedly reflects the fact that compound C is a mixture of two diastereomers in agreement with A. J. Bailey's observation (personal communication) that on chromatography compound C was partially resolved into two compounds of apparently identical chemical properties. Similar chromatographic behaviour has also been reported by Tanzer *et al.* (1973) for their histidinohydroxymesodesmosine.

Discussion

Bailey *et al.* (1970) first described the isolation of compound C and indicated that it was derived from an acid-labile collagen cross-link that could be stabilized by borohydride reduction. The spectroscopic studies on compound C described in the present paper provide strong evidence that its structure is that shown in Fig. 2, and establish its identity with the structure recently proposed for the 'post-histidine' compound isolated from cow skin collagen (Tanzer *et al.*, 1973).

In spite of our general agreement with Tanzer *et al.* (1973) on the structure of this possible cross-link, we disagree with these authors over the interpretation of certain spectral data. This, together with several new facets of our work, warrants further discussion, as they may be of some importance to other workers engaged in structural studies on compounds of this type.

Since low-resolution mass spectrometry provides only integral masses for molecular and fragment ions, it is highly desirable to have some information in addition to the fragmentation pattern of one simple derivative, in order to draw structurally relevant conclusions from the mass spectrum of an unknown

compound. The mass shift technique (Biemann, 1962) provides one way of obtaining additional information, and can be particularly useful in studies of complex molecules for determining the number of reactive groups. The conventional technique entails the preparation of two separate sample derivatives, by using two reagents related in a homologous manner (e.g. methanol and ethanol) or occasionally isotopically (e.g. methanol and [$^2\text{H}_3$]methanol). The shift in mass of the molecular ion or fragment ions is determined by comparing the spectra of the two derivatives, and hence the number of groups blocked by that particular reagent is determined. We have found that this method suffers from the following disadvantages.

(a) Homologous derivatives may differ significantly in their principal mass-spectral fragmentations, leading to difficulties or possibly error in interpretation.

(b) A potential source of more serious error lies in the possibility of differing reactivity of functional groups towards homologous reagents. For example, a carboxyl site may differ significantly in its reactivity towards esterification with homologous alcohols, thus leading to serious difficulties in interpretation.

(c) Whether the reagents are homologously or isotopically related, the conventional technique requires the formation of two separate derivatives, and comparison of the two different mass spectra. This can pose several problems particularly when only very small samples are available for examination. Further, if the sample derivatives are not run under identical conditions of background, temperature, etc. in the mass spectrometer the result may be an inability to spot structurally relevant signals.

All of the above disadvantages are removed by the application of a novel approach which we term the 'equi-isotopic mass shift' technique. With this method the compound under study is esterified with methanol- $[\text{P}^2\text{H}_4]$ -methanol (1:1, v/v) containing dry HCl. Assuming the original presence of n carboxyl groups, an assembly of molecules is generated containing $3n$, $3(n-1)$, $3(n-2)$, ..., 0 ^2H atoms. For example, with 2 esterifiable groups, molecules containing 6, 3 and 0 ^2H atoms are obtained. The relative abundance of each of these species is governed by the coefficients (representing probabilities) of a simple binomial expansion, and can readily be determined without calculation from Pascal's triangle. So for $n = 4$, the relative abundances are 1:4:6:4:1 for $^2\text{H}_0$, $^2\text{H}_3$, $^2\text{H}_6$, $^2\text{H}_9$ and $^2\text{H}_{12}$ species respectively. In the mass spectrum therefore the presence of 4 esterifiable groups is simply revealed by a series of ions, 3 mass units apart forming a characteristic 1:4:6:4:1 quintet. This was the pattern observed in the molecular-ion region of the mass spectrum (Fig. 1) of the compound C derivative, thus unambiguously showing that compound C contains 4 carboxyl groups. Further, the number of esterified groups in

each fragment ion is immediately apparent from the patterns observed in the one spectrum. Structurally useful signals can be recognized immediately and differentiated from background signals by the presence of a meaningful pattern. Conversely, we can say that m/e 980 (Fig. 6 of Tanzer *et al.*, 1973) cannot wholly be attributed to the structurally significant fragment suggested by these workers since the equivalent signal in our spectrum (Fig. 1) at m/e 979 is not part of a simple 1:3:3:1 quartet. Our data suggest that a major contributor to m/e 979 is a fragment containing 4 carboxyl groups, formed by successive losses of fluorine-containing fragments, and that it is not a structurally relevant signal. Similarly, we assign the 1:3:3:1 quartet at m/e 1077, 1080, 1083 and 1086 to the loss of CF_3 from $M^+-CO_2CH_3$.

Finally, it is important to realize that the equi-isotopic shift technique is not limited simply to the determination of carboxyl groups. For example, methyl iodide- $[^2H_3]$ methyl iodide (1:1, v/v) can be used for determining hydroxyl groups by methyl ether formation; acetic anhydride- $[^2H_6]$ acetic anhydride (1:1, v/v) can be used to give the number of acetylatable functions, and so on. Since only secondary isotope effects can be expected for both chemical reaction and mass-spectral fragmentation, the equi-isotopic mass shift technique does not suffer from the disadvantages of the homologous mass shift technique discussed above [points (a) and (b)]. The new method also offers the advantage of simplicity and rapidity both in technical operation and interpretation of data, and this facet may well prove useful in the routine mass-spectrometric screening of micro amounts of biologically derived compounds from both organic and mineral sources.

One problem which arose in our mass-spectrometric study of compound C was the choice of acyl group. We had hoped to use the equi-isotopic shift technique with acetic anhydride- $[^2H_6]$ acetic anhydride, but the acetyl derivative was thermally unstable in the mass spectrometer. We were thus forced to use the conventional mass shift technique to determine the number of acylable groups in compound C.

Our approach to the problem of choice of derivative compared with that of Tanzer *et al.* (1973) emphasizes several important points. We found that the trifluoroacetyl derivative of compound C methyl ester gave an abundant molecular ion, as was observed for the 'post-histidine' compound (Tanzer *et al.*, 1973). The latter authors chose the pentafluoropropionyl group for their second derivative. Because of the increased molecular weight (50 mass units/acyl group) and increased opportunities for fragmentation compared with those of the trifluoroacetyl compound, it is hardly surprising that this second derivative gave no observable molecular ion.

These workers were forced to speculate as to the molecular weight of the second derivative by extrapolation from the highest mass ion in the spectrum. This work cannot be considered as proof of the number of acylable groups in 'post histidine'. Our own choice of derivative was the difluoroacetyl compound C methyl ester, which has a lower molecular weight (18 mass units/acyl group) than the trifluoroacetyl derivative and contains the same number of covalent bonds. This compound gave a strong molecular-ion region (m/e 1097 2H_0) in the mass spectrum, which clearly indicated the presence of 6 acylable groups in compound C.

Low-resolution mass spectrometry in recent years has been increasingly replaced by automated high-resolution studies for structural elucidation work. Although high-resolution mass spectrometry can be a very powerful tool for the unique definition of atomic composition for small (m/e 200–500) molecular or fragment ions containing small numbers of heteroatoms, the method does have serious limitations when applied to high-molecular-weight species containing large numbers of heteroatoms. In these cases literally hundreds of possible atomic compositions are obtained which will fit the 'accurate' mass. For example, with the exact mass measurement of 1341.273 quoted by Tanzer *et al.* (1973) from the spectrum of their pentafluoropropionyl derivative of 'post-histidine' methyl ester, and setting upper limits of 64C, 7N, 13O and 25F atoms and a search interval of ± 10 p.p.m., we obtained a total of 183 possible atomic compositions for this ion. If higher heteroatom and C atom limits had been set, and this would normally be the case when working with an unknown compound, then a considerably larger number of possibilities would be computed. Setting no limit on C atoms in the above example gave almost 1000 possible atomic compositions for the measured mass. The majority of these possibilities could be eliminated by recourse to other data, but an appreciable number still remain as real possibilities. Such mass measurements can therefore be regarded as only indirect evidence for an atomic composition which has been inferred from other data or from intuition. In our studies of compound C we have limited the use of high-resolution mass spectrometry to the lower-molecular-weight fragments. Even for some of these (m/e 503 and 728) a considerable number of possible atomic compositions were computed, which could only be limited to one by using other data (mass shift, H- 2H exchange, etc.).

The proton n.m.r. spectrum of compound C proved of limited usefulness in providing structural information. Those structural features that could be deduced from the spectrum have been described above. After we had arrived at a full structure for compound C (Fig. 2), a re-examination of its proton

n.m.r. spectrum showed that all its features were entirely consistent with this structure. At this point we again find ourselves in disagreement with Tanzer *et al.* (1973). These authors have assigned the multiplet at $\delta 4.2$ to the C-5 and C-10 methine protons. However, in our spectrum, and presumably in theirs, this multiplet integrates for a single proton, and we have assigned this to the proton attached to C-10. The C-5 proton we would expect to resonate as a complex multiplet at approx. $\delta 3.8$, which is the position which we have observed for the signal from the δ -methine proton in δ hydroxylysine hydrochloride. In compound C, this resonance is obscured by other signals.

Because the proton n.m.r. spectrum of compound C is complicated by the overlapping of all but three or four signals we turned to an examination of the ^{13}C n.m.r. spectrum. ^{13}C resonances are usually found over a range of approx. 200 p.p.m., compared with the approx. 10 p.p.m. range for proton resonances, and under conditions of complete proton decoupling the overlapping of signals in the ^{13}C n.m.r. spectrum is less likely. This was the case for compound C and the only signals that could not be individually observed were those owing to C-4, and to C-2, C-17, C-22 and C-26, which were at best resolved into only three signals.

On changing the pH of solutions containing compounds carrying basic or acid functions the ^{13}C chemical shifts have been observed to vary in characteristic manners (Horsley *et al.*, 1970; Gurd *et al.*, 1971; Morishima *et al.*, 1973). This greatly facilitates the assignments in the ^{13}C spectra of some compounds and we have used this technique in our study of compound C. This method was particularly useful for the assignment of the site of attachment of C-10 at the imidazole ring. Thus, although comparison of the imidazole C atom chemical shifts for the methylhistidines and compound C at just one pH value may have suggested that C-10 was attached to N-11, a comparison of the protonation-shift curves for these compounds unambiguously demonstrates that C-10 is attached to N-11. We would suggest that in comparative n.m.r. (^1H or ^{13}C) studies, which are aimed at deciding between two or more possible isomers for an unknown compound, the protonation-shift curves should always be determined rather than attempting to draw conclusions from measurements at a single pH value (this will, of course, only apply to compounds which contain acidic or basic groups in the vicinity of the sites of potential isomerism). One reason for this is evident from our study on compound C. The pK_a for the imidazolium group of the methylhistidines differs significantly from that for the imidazolium group of compound C; thus for this and similar cases, comparison of chemical shifts for solutions at a pH between these pK_a values

could, in principle, lead to erroneous conclusions being drawn regarding the structure under investigation. Fortunately, this is not the case for the ^{13}C n.m.r. spectrum of compound C, or apparently its ^1H n.m.r. spectrum (see Tanzer *et al.*, 1973), but this does not invalidate the general point we have just made.

The ^{13}C protonation shifts for compound C provide good characterization of those C atoms that are affected by protonation of the secondary amino and imidazole groups. The assignments within the closely spaced set of signals between 20 and 35 p.p.m. might be regarded, with the possible exception of C-16, as being tentative. We feel, however, that those assignments which we have made in this region of the spectrum do provide a good fit between the experimental data and the proposed structure. Since the C atom resonances in this 20 to 35 p.p.m. region are all involved in structural units that are well characterized by the mass spectrometry and the single-frequency proton decoupling experiments in the ^{13}C spectrum, the small doubts associated with assignments in this region do not present a handicap to the structural study as a whole.

In conclusion we believe that our study of compound C may provide a useful model for structural elucidation work on other compounds of this type. Neither the mass-spectrometric nor the ^{13}C n.m.r. studies, if taken in isolation, would provide enough data to arrive at the complete structure shown in Fig. 2; however, taken together these two methods have provided very strong evidence for this structure for compound C. The significance of this potential cross-link in the chemistry of collagen is discussed elsewhere (Robins & Bailey, 1973).

The authors wish to thank the Science Research Council and the Medical Research Council for financial assistance, Dr. D. H. Williams for his interest in this work and Dr. A. J. Bailey for samples of compound C.

References

- Archer, R. A., Cooper, R. D. G., Demarco, P. V. & Johnson, L. F. (1970) *Chem. Commun.* 1291–1293
- Bailey, A. J. & Peach, C. M. (1968) *Biochem. Biophys. Res. Commun.* **33**, 812–819
- Bailey, A. J., Fowler, L. J. & Peach, C. M. (1969) *Biochem. Biophys. Res. Commun.* **35**, 663–671
- Bailey, A. J., Peach, C. M. & Fowler, L. J. (1970) *Biochem. J.* **117**, 819–831
- Biemann, K. (1962) *Mass Spectrometry: Organic Chemical Applications* p. 187, McGraw-Hill, New York
- Bornstein, P. & Piez, K. A. (1966) *Biochemistry* **5**, 3460–3473
- Grant, D. M. & Cheney, B. V. (1967) *J. Amer. Chem. Soc.* **89**, 5315–5318

- Grant, D. M. & Paul, E. G. (1964) *J. Amer. Chem. Soc.* **86**, 2984–2990
- Gurd, F. R. N., Lawson, P. J., Cochran, D. W. & Wenkert, E. (1971) *J. Biol. Chem.* **246**, 3725–3630
- Horsley, W., Sternlicht, H. & Cohen, J. S. (1970) *J. Amer. Chem. Soc.* **92**, 680–686
- Johnson, L. F. & Jankowski, W. C. (1972) *Carbon-13 NMR Spectra*, Wiley, New York
- Kang, A. H., Faris, B. & Franzblau, C. (1970) *Biochem. Biophys. Res. Commun.* **39**, 175–182
- McDonald, C. C. & Phillips, W. D. (1969) *J. Amer. Chem. Soc.* **91**, 1513–1521
- Meadows, D. H., Jardetzky, O., Epand, R. M., Ruterjams, H. & Scheraga, H. A. (1968) *Proc. Nat. Acad. Sci. U.S.* **60**, 766–772
- Morishima, I., Yoshikawa, K., Okada, K., Yonezawa, T. & Goto, K. (1973) *J. Amer. Chem. Soc.* **95**, 165–171
- Paz, M. A., Lent, R. W., Faris, B., Franzblau, C., Blumenfeld, O. O. & Gallop, P. M. (1969) *Biochem. Biophys. Res. Commun.* **34**, 221–229
- Robins, S. P. & Bailey, A. J. (1973) *Biochem. J.* **135**, 657–665
- Stothers, J. B. (1972) *Carbon-13 NMR Spectroscopy*, p. 152, Academic Press, New York
- Tanzer, M. L., Mechanic, G. & Gallop, P. M. (1970) *Biochim. Biophys. Acta* **207**, 548–552
- Tanzer, M. L., Housley, T., Berube, L., Fairweather, R., Franzblau, C. & Gallop, P. M. (1973) *J. Biol. Chem.* **248**, 393–402
- Woolfenden, W. R. & Grant, D. M. (1966) *J. Amer. Chem. Soc.* **88**, 1496–1502