Pure & Appl. Chem., Vol. 50, pp. 159-169. Pergamon Press, 1978. Printed in Great Britain.

MASS SPECTROSCOPIC INVESTIGATION OF NUCLEIC ACID DEGRADATION PRODUCTS

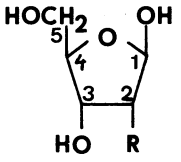
Herbert Budzikiewicz

Institut für Organische Chemie der Universität zu Köln,
Greinstraße 4, 5000 Köln 41, Bundesrepublik Deutschland.

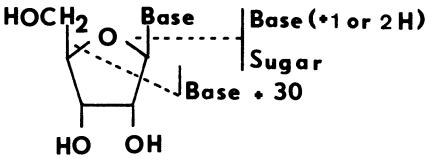
Abstract - The possibilities of structure elucidation of nucleic acid degradation products and of sequencing of oligonucleotides are discussed.

There are three classes of natural products, the sequencing of which has been attempted by mass spectrometry, viz. oligosaccharides, peptides and nucleotides. While oligosaccharides offer some special problems due to structural variability, peptides and nucleotides are built according to the same basic principle: a backbone of recurrent structural units which is substituted at regular intervals by characteristic groups. In case fragmentation of the backbone occurs in a specific way (as at the amide bonds in peptides) sequential information may be obtained from a mass spectrum. There are, however, two further premises for the successful application of mass spectrometry: it must be possible to distinguish the two ends of the strand from each other and it is necessary to obtain ions of the entire molecule or at least of parts of it still carrying sequential information.

The backbone of nucleotides is an alternating sequence of sugar (ribose, 1, R = OH, and desoxyribose, 1, R = H) and phosphoric acid residues which is substituted N-glycosidically by purine and pyrimidine bases (Fig. 1).



1



2

The mass spectrometric behavior of pyrimidine and purine bases has been investigated *in extenso* (Ref. 1), but is of minor importance in this context, since due to the stability of the aromatic nucleus, fragment formation in nucleosides and nucleotides takes place preferentially in the sugar/phosphate units. Identification of a base in bigger units has, therefore, to come from mass and elemental composition of pertinent fragments. That nucleosides (i.e. base + sugar) are amenable to EI mass spectrometry (with the possible exception of guanosine which shows severe pyrolytic degradation) was shown first by *Biemann* (2). They exhibit a rather straight forward fragmentation pattern (Refs. 2, 3) (2), though the abundance of the typical fragment ions varies (Ref. 1). CI, upon suitable selection of the auxilliary gas, may enhance pseudo-M^+ (M + 1) and somewhat simplify the spectrum (Ref. 4). FI (Ref. 5) and FD (Refs. 6, 7) spectra are dominated by M^+ (occasionally $[M - H_2O]^+$) and show clearly the sugar and base fragments (Fig. 2). Mass spectra of nucleosides are especially important for the detection of unusual bases (Refs. 1, 8-10) and for biogenetic studies by isotope labelling (Ref. 11).

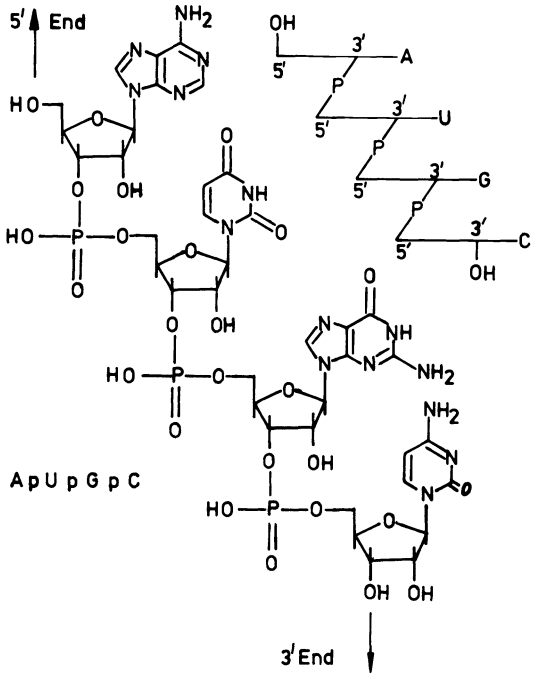


Fig. 1. Primary structure of ribonucleotides.

Addition of a phosphate unit (i.e. formation of a mononucleotide) renders these compounds too involatile for direct EI mass spectrometry. Various types of derivatives (TMS, methyl ethers, acetates, acetonates, phenyl boronic acid esters) have been suggested (Ref. 1) to enhance volatility. FD (Refs. 6, 7, 12) allows the measurement of free nucleotides giving $[M + H]^+$ and ions characteristic of the base and the sugar moiety (Fig. 3). Although not of much use for structural work - such information is extracted much more readily from the corresponding nucleosides which are also easier to handle - they may serve as models to discuss problems encountered with higher units:

- Derivatisation. Trimethyl silylation is readily achieved (per-TMS derivatives allow even analysis by GC/MS; CI (i-butane) enhances the abundance of M^+ (Ref. 13)), but since each TMS group adds 72 amu to the molecular weight the limitation is obvious (Ref. 14). Methylation, depending upon the method used usually results in mixtures (Refs. 7, 13).

Differentiation between 3'- and 5'-phosphates. FD spectra of the two isomers are within the range of variability (*v. infra*) indistinguishable (Figs. 3c and 3d). At the best, it may be observed that loss of H_2O occurs at lower temperatures for 3'-phosphates due to cyclophosphate formation (Ref.7). It has been reported by *McCloskey* (14) that a differentiation is possible with EI spectra of per-TMS derivatives. Comparing the spectra of 3'- and 5'-adenosine monophosphoric acid he found that m/e 169 (part of the sugar moiety) and $[M - PO_2(OTMS)_2]^+$ are both characteristically more abundant in the 5'-isomer. Examination of the other three pairs showed (Ref. 13) that higher abundance of m/e 169 for the 5'-isomer could be observed also for GMP and UMP, but not for CMP, and that $[M - PO_2(OTMS)_2]^+$ was of about the same abundance for the GMP and CMP pairs and far more abundant for U-3'-MP. However one can rely on m/e 501 $[M - base]^+$ which, in all four cases, is of much higher intensity for the 3'-isomer (Table 1) (Ref. 13). A differentiation between 2'- and 3'-isomers seems not to be possible for any of the nucleoside phosphoric acids, owing to isomerisation (Refs. 13, 14).

- Problems in FD work (Ref. 12). FD spectra of nucleotides change with the anode temperature. Starting from $[M + H]^+$ as sole ion in the spectrum (Fig. 3a) of A-3'-MP with increasing temperature fragments appear (m/e 136, $[adenine + H]^+$ in Fig. 3b), increase in variety (loss of 1 or 2 H_2O from $[M + H]^+$ and from $[A + H]^+$, m/e 250 and 232, $[M - adenine - H_2O + H]^+$, m/e 195 and $H_4PO_4^+$, m/e 99 in Fig. 3c). With higher temperature uncharacteristic ions appear. Hence a suitable temperature has to be selected for the

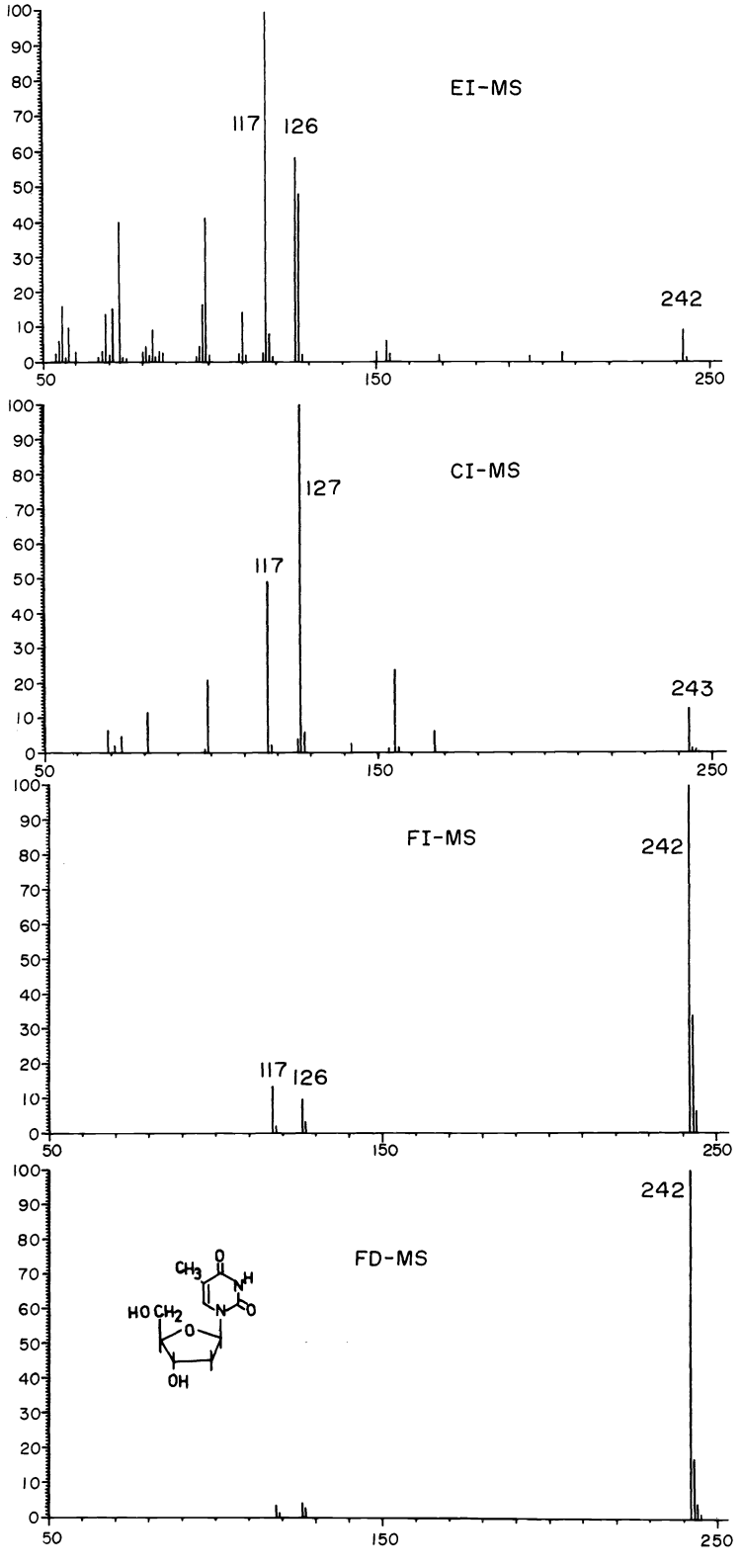


Fig. 2. EI, CI(CH_4), FI and FD spectra of thymidine

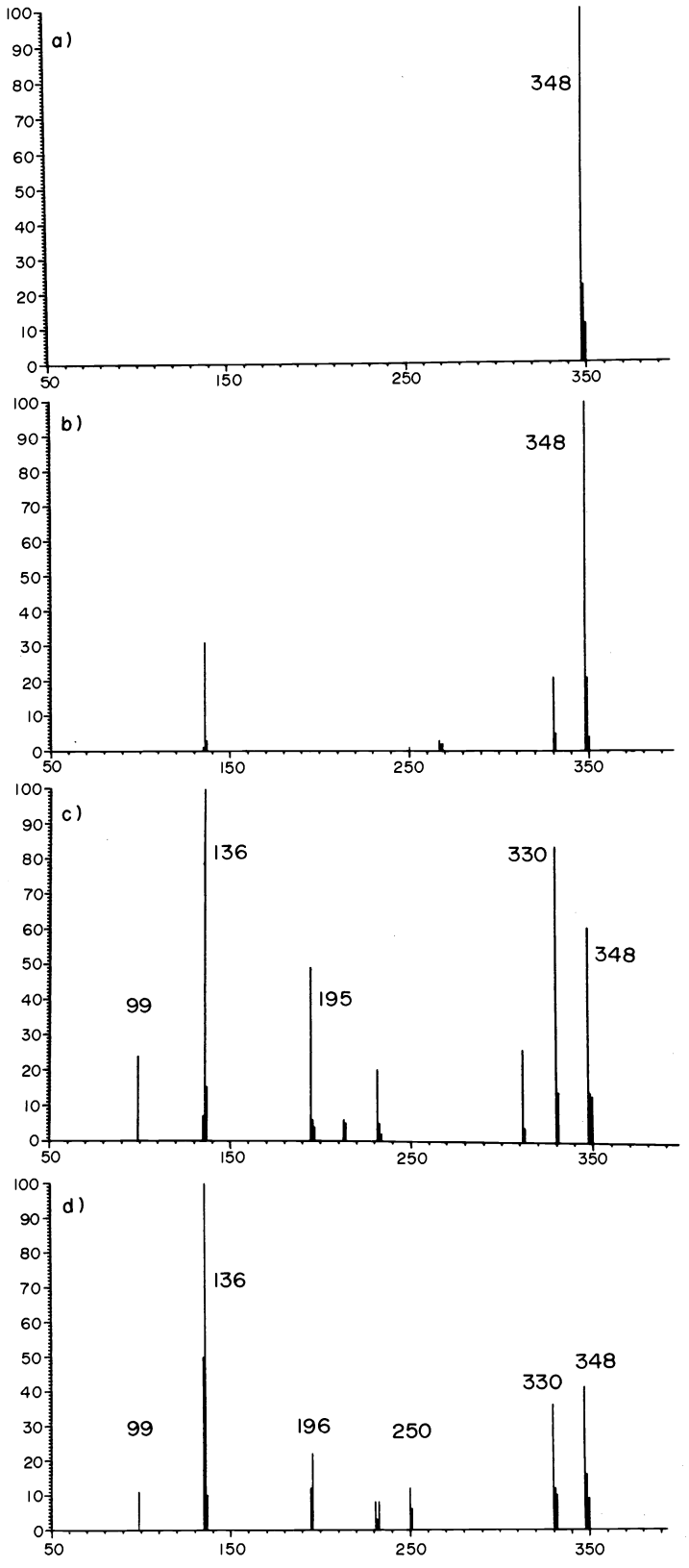


Fig. 3. a) - c) FD spectra of A-3'-MP, 14, 18 and 19 mA emitter current, resp., d) FD spectrum of A-5'-MP, 18 mA emitter current; all spectra obtained by electron multiplier recording.

TABLE 1. Characteristic ions for nucleoside phosphoric acid per-trimethylsilyl ethers

	% rel. intensity							
	A		G		U		C	
	3'	5'	3'	5'	3'	5'	3'	5'
m/e 169	25	72	9	28	48	79	23	22
$[M - PO_2(OTMS)_2]^+$	3	5	1	2	11	-	2	-
m/e 501	9	-	14	-	7	0,5	5	0,2

information required (molecular weight, structural information). Also the degree of reproducibility of spectra encountered with EI work cannot be expected with FD: Emitter quality (for the question: activated or non activated wires cf. Refs. 15, 16), sample distribution on the emitter, speed of heating and traces of cations (*v. infra*) etc. greatly influence the relative abundance of the various ions. At least with higher emitter temperature thermal decomposition to a certain extent cannot always be avoided. Since in most cases thermal degradation before ionisation and fragmentation of the ionised molecule parallel each other these two processes yield ions differing only by one or a few hydrogen atoms. Thus peak groups rather than single lines are produced. Frequently it is not possible to decide whether a given ion stems from the intact ionised molecule, ionisation of pyrolytic fragments or decomposition of clusters (*v. infra*). Since the relative rates of these competing processes are very dependent upon the experimental conditions, a further factor reflecting upon the reproducibility of spectra has to be taken into account.

- Cationisation and cluster formation. Free nucleotides are desorbed from the emitter in a protonated form (Refs. 6, 7, 16); cationisation with alkali ions apparently does not occur (Ref. 16). Alkali salts when subjected to FD (Refs. 7, 12) also form protonated molecular ions and in addition they foster cluster formation. Thus, in the spectrum of C-5'-$MPNa_2$ (Fig. 4) ions are observed which contain two cytosine moieties, e.g. m/e 224 $[2\ \text{cytosine}]^+$, 235,5 $[4\ \text{cytosine} + Na]^{++}$, 266,5 $[(\text{cytidine} + Na)_2 + H]^{++}$. The otherwise (Ref. 17) desirable Li salts of nucleotides yield spectra which can hardly be analysed: G-2':3'-MPLi (Fig. 5)does not give any ions in the vicinity of M^+ but instead a peak group at m/e 539 $[(\text{G-2':3'-MP})_2 - \text{guanine}]^+$ and unidentifyable ions up to beyond m/e 1000. Pyrolytic decomposition of such clusters may yield ions (*v. supra*) whose genesis cannot be explained otherwise.

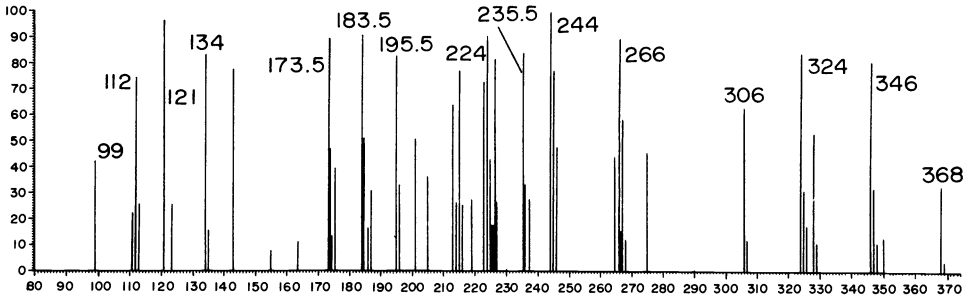


Fig. 4. FD spectrum of C-5'-$MPNa_2$, 10 - 22 mA emitter current, photo plate recording

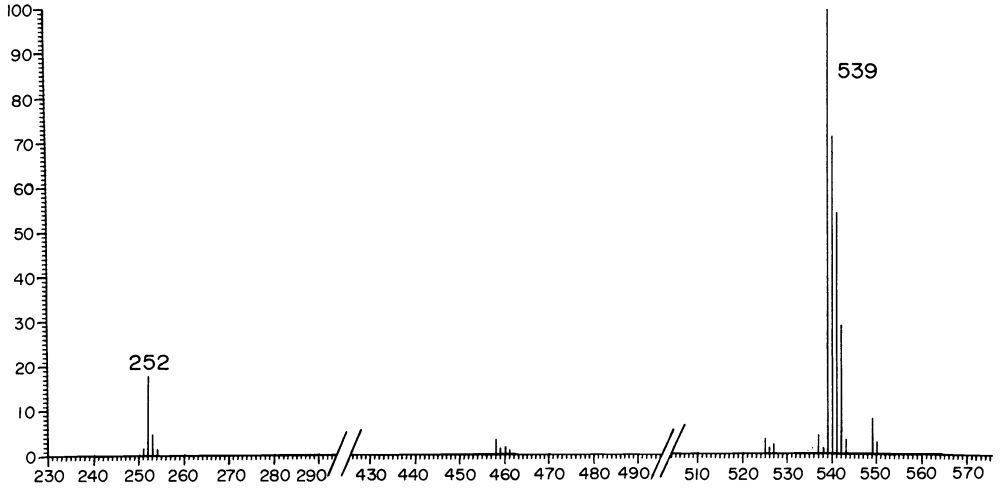
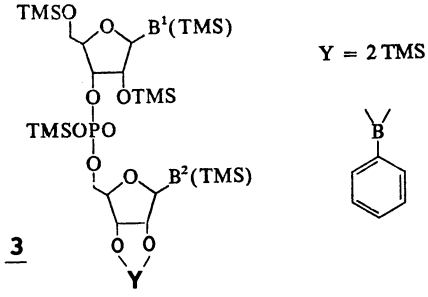


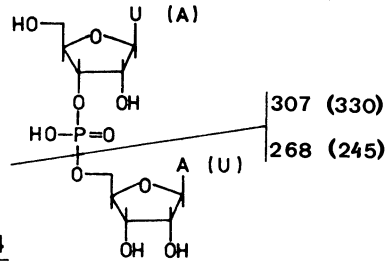
Fig. 5. FD spectrum of G-2':3'-MPLi, 18 mA emitter current, electron multiplier registration

The differentiation between 3'- and 5'-nucleoside phosphoric acids is the first step towards sequencing, which really commences with a clearcut distinguishing between isomeric pairs of mixed dinucleoside phosphates. For the isomeric pairs of per-TMS-ethers (Refs. 1, 18) and per-trifluoroacetates differences in the abundance of certain characteristic ions have been observed (Ref. 19). Possible ambiguities can be circumvented by asymmetric derivatisation (Ref. 20) (3) which is, however, applicable only to the ribo series. In any case increase of molecular weight (e.g., GpG-per-TMS 1276) hardly allows an extension of this method beyond dinucleoside phosphates.



3

Y = 2 TMS



4

Dinucleoside phosphates (both free acids and their salts) readily give FD spectra comprising $[M + 1]^+$ and a series of characteristic fragments (Refs. 7, 16, 21, 22, 32). In the ribo series - as for example the pair ApU and UpA, Fig. 6 and Table 2 - fragmentation between the 5'-oxygen and P yields a cyclophosphate ion for the 3'-substituted nucleoside and the nucleoside ion for the 5'-substituted one (4). The situation is not at all straight forward. For one thing cyclophosphate ions may also be formed from the 5'-substituted nucleoside (cf. m/e 330 in Fig. 6b). On the other hand the cyclophosphate ion may be completely absent, as in a published ApC spectrum (m/e 330) (Ref. 21). Experimental conditions, especially emitter temperature seem to exercise a big influence here. Thus, in a different ApC spectrum (Ref. 22) "low abundance" is reported for m/e 330. A similar situation prevails with UpGNa (Ref. 7): the G-cyclophosphate ion (m/e 345) is observed with 5% relative intensity with 18 mA emitter heating, but absent with 19 and 20 mA; the characteristic U-cyclophosphate (m/e 307) appears with 19 mA, but is neither observed with 18 nor with 20 mA, which demonstrates that one should not rely upon a single recording (Ref. 12). Nevertheless, the presence of a cyclophosphate ion seems to identify the 3'-substituted nucleoside. When two cyclophosphate ions are observed, the one of higher abundance is associated with 3'-substitution.

Relatively little is known about the behavior of representatives of the desoxy series. Obviously, 2',3'-cyclophosphate ions cannot be formed. Accordingly, in the published FD spectra of both TpT (Ref. 21) and TpTNa (Ref. 7) m/e 305 is lacking. But while the former yields an ion at m/e 323

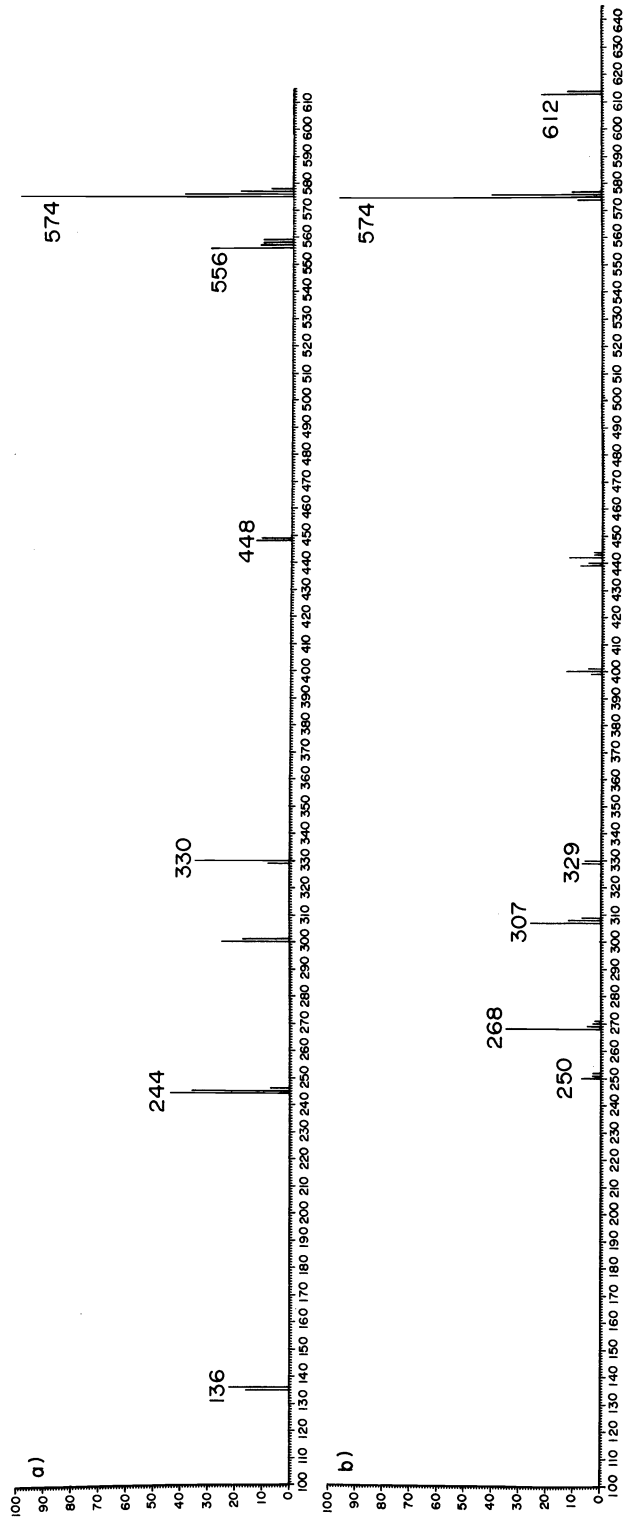


Fig. 6. FD spectra of a) ApU, 19 mA emitter current, b) UpA, 14 mA emitter current; both obtained with electron multiplier recording.

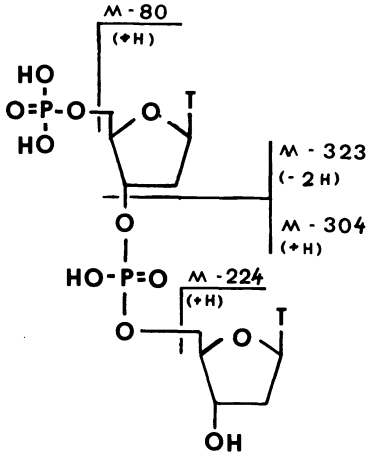
TABLE 2. Main fragments of ApU and UpA (cf. Fig. 5)

m/e		type of ion
UpA	ApU	
	136	$[adenine + H]^+$
	244	$[uridine]^+$
250		$[adenosine + H - H_2O]^+$
268		$[adenosine + H]^+$
307		$[UpA - adenosine]^+$
330	330	$[UpA \text{ or } ApU - uridine]^+$
	556	$[UpA + H - H_2O]^+$
574	574	$[UpA \text{ or } ApU + H]^+$
612		$[UpA + K]^+$

($[T\text{-}MP + H]^+$), the latter exhibits m/e 242 (T^+). In addition, both compounds show $[M + H]^+$ (m/e 547 and 569, respectively), $H_4PO_4^+$ (m/e 99) and $[thymine]^+$ (m/e 126). To what extent this different behavior is the result of experimental parameters, or of cluster formation in the case of the Na salt cannot be judged from the data available.

The high thermal lability of the desoxy series becomes evident with the isomeric pair TpdA and dApT (Ref. 23): the appearance of the spectra changes quickly with the temperature and the spectra are crowded with ions stemming from cluster formation, phosphate transfer (i.e., aggregation of phosphate units resulting in ion series 80 amu apart), etc. Amongst the structurally characteristic ions of both isomers are $[M + H]^+$, $[dA + H - 1 \text{ or } 2\ H_2O]^+$, $[T + H]^+$ and $[T\text{-}MP + H]^+$. Apparently, it is the more stable nucleotide which is formed (perhaps thermally?). Sequential information can, therefore, not be obtained.

As expected $pTpTNa_3$ (in the presence of NEt_3) is even more difficult to handle (Ref. 23): at first $[pTpT(NEt_3)_3 + H]^+$ appears and disappears again; then $[pTpTNa + H]^+$ and $[pTpT + H]^+$ and ions due to fragment formation come up, followed by species whose genesis can only be explained by recombination of desoxyribose, phosphoric acid and thymine residues from some kind of "pool". Structurally relevant ions are indicated in 5. Whether sequentially informative ions can be obtained by specific derivatisation is presently under investigation. In any case it should be noted that complete FD spectra of true dinucleotides (and not merely of dinucleoside phosphates) can be obtained even in the labile desoxy series.



5

An interesting approach for the distinguishing of isomeric di-desoxyribonucleoside phosphates has been suggested by *Wiebers* and *Shapiro* (24): the EI spectra of the pyrolysis products obtained by heating the samples on a direct probe in the ion source differ for the isomeric pairs in a characteristic and reproducible way. Thus via a simple computer program identification is possible.

Higher units than dinucleoside phosphates - derivatised or underivatised - are no longer directly accessible to EI analysis. Several indirect approaches have, however, been suggested, *viz.*:

- Mass spectrometric analysis of pyrolysates by heating DNA on a direct probe of a mass spectromer results in ions identifying the four common bases but no sequential information is obtained directly (Refs. 19, 25). The data published on pattern recognition studies with such pyrolysates are still too incomplete (Ref. 26) to allow an assessment of their value for units beyond dinucleoside phosphates (*v. supra*) (Ref. 24). Analysis of Curie point pyrolysis products both by low voltage EI and by high resolution FI and CA mass spectrometry revealed that by this method not the bases, but essentially degradation products from the carbohydrate moieties are obtained, allowing the differentiation between DNA and RNA (Refs. 27, 28).

- A combined chemical degradation and mass spectrometric analysis has been suggested for the sequencing of trinucleotides using the periodate oxidation-β-elimination method. Thus the 3'-terminus may be obtained acetylated, the 5'-terminus as a nucleoside and the central base in free form; the mixture is then subjected to mass spectrometry. For tetranucleotides the method becomes more complex. The degradation procedure is obviously restricted to the ribo series (Ref. 29).

- A combination of pyrolysis mass spectrometry and degradation is based (Ref. 19) on the observation that the "mass spectra" of per-trifluoroacetylated oligo-desoxyribonucleotides will identify the two termini and give information about the bases present in the central units, although without indication of their sequence. This may be circumvented by limited enzymatic degradation starting from the 3'-end yielding a mixture of products of decreasing chain length which can be separated by chromatography, per-trifluoroacetylated and analysed as indicated, to establish the 5'-terminus and sequentially the 3'-termini, from which the original chain can be reconstructed:

oligonucleotide	5'-terminus	3'-terminus
ApBpCpDpEpF	A	F
↓		
ApBpCpDpE	A	E
⋮		⋮

- Since the main difficulties in the mass spectrometric sequencing of oligonucleotides stem from the sugar-phosphate chain, transfer of the bases to a less complex matrix has been suggested (Ref. 30). No practical results are available yet.

Some direct information may again be obtained from FD investigation. Pyrolysis of DNA on a special emitter directly in the ion source and evaluation of the spectra thus obtained by high resolution measurements allows the identification of the common bases, of fragments comprising nucleoside and nucleotide portions of the molecule, and of doubly charged dinucleotide ions (Ref. 31). The last group seems to offer information about neighboring nucleosidic units and may thus be of importance for obtaining sequential information from oligonucleotides. One should, however, keep in mind that via cluster formation (*v. supra*) ions may be formed which contain the structural units of a dinucleotide which is not originally present in the molecule.

In an FD spectrum (Fig. 7) (Ref. 7) of UpApG-2':3'-MP$(NH_4)_3$ ions are found corresponding to the protonated bases (m/e 113, 136 and 152), and to the three cyclophosphate fragments $[U\text{-}2':3'\text{-}MP + H]^+$, $A\text{-}2':3'\text{-}MP^+$ and $G\text{-}2':3'\text{-}MP^+$ (both unprotonated), and to aggregates containing one additional phosphoric acid (+ 80 amu, *v. supra*). One is tempted to conclude from this spectrum that for UpApG sequencing should be possible (U giving a protonated cyclophosphate ion, A an unprotonated one and G a nucleosidic ion, *v. supra*), but the care that has to be taken is demonstrated by the spectrum of ApU-2':3'-MP(NEt_3) (Ref. 7) where both cyclophosphate ions appear in the protonated form.

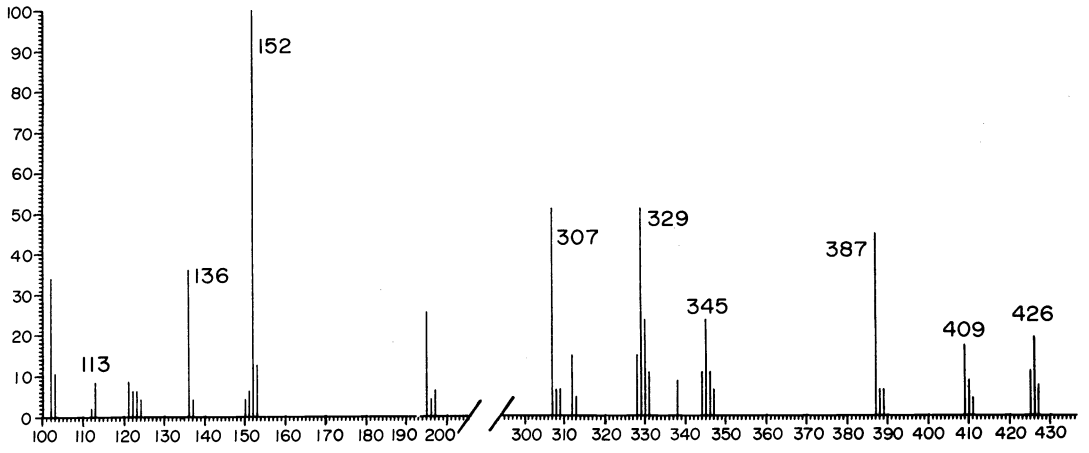


Fig. 7. FD spectrum of UpApG-2':3'-MP$(NH_4)_3$, 19 mA emitter current, electron multiplier recording

The data available by now are far too incomplete to allow any generalisations. One may draw the following conclusions: FD seems the most promising method for further investigations, especially if the experimental possibilities are exploited, *viz.* selection of the emitter (Refs. 15, 16), heating speed (Ref. 21) controlling partial pyrolysis, registration (and summation) of spectra obtained at different temperatures (Ref. 12), selection of the cation (Refs. 12, 16) and derivatisation with or without isotope labelling (Refs. 7, 13). On the other hand, the many experimental parameters which may influence the appearance of an FD spectrum of such thermolabile compounds, which are prone to all kinds of reactions owing to their many functional groups should also be kept in mind when conclusions are drawn from isolated data. Much additional work remains to be done, but one can say today that at least trinucleotides are within the realm of our technical possibilities.

Now the question arises: is all this research on the sequencing of nucleic acids by mass spectrometry of purely academic interest? Mass spectroscopic sequencing of peptides has its well established place in the armamentarium for structure elucidation since units comprising up to ten amino acids and beyond can be handled. For oligonucleotides the limit for a direct analysis will probably be much lower. It has also to be taken into account that sequencing of both ribo and desoxyribonucleotides can readily be achieved by biochemical methods which require only minute amounts of material. Yet, there are three areas, where mass spectrometry can compete successfully, *viz.*, a) if unusual bases are present (originally or after transformation by chemical reagents or by radiation) which are not readily recognised by the standard methods; b) for control purposes in synthetic or other chemical work (e.g. Ref. 13), and c) for the structure elucidation of small units carrying characteristic genetic information isolated from RNA or DNA which cannot be handled by the standard methods. And it should not be forgotten that much information has been accumulated during the mass spectrometric work in this area which is of importance for the investigation of other highly nonvolatile and labile compounds.

Finally I wish to thank *Dr. M. Linscheid* and *Dipl.-Chem. G. Feistner* from this laboratory, whose work was intimately connected with the results reported here.

Abbreviations used: a) methods: CA collision activation, CI chemical ionisation, EI electron impact ionisation, FD field desorption, FI field ionisation; amu absolute mass units; b) compounds: A adenosine, C cytidine, G guanosine, T thymidine, U uridine, dA desoxyadenosine etc.; A-3'-MP adenosine-3'-monophosphoric acid etc., A-2':3'-MP adenosine-2':3'-cyclophosphate etc., A-3'-$MPNa_2$ di-Na salt of A-3'-MP etc.; ApU adenylyl(3'-5')-uridine etc.; DNA desoxyribonucleic acid, RNA ribonucleic acid; TMS trimethylsilyl.

REFERENCES

1. For a compilation see: C.Hignite in "Biochemical Applications of Mass Spectrometry" (G.Waller, ed.), Wiley-Interscience, New York, p.429ff,1972.
2. K.Biemann and J.A.McCloskey, Jr., J. Amer. Chem. Soc. 84, 2005-2006 (1962).
3. J.G.Liehr, D.L.v.Minden, S.E.Hattox and J.A.McCloskey, Biomed. Mass Spectrom. 1, 281-285 (1974).
4. M.S.Wilson and J.A.McCloskey, J. Amer. Chem. Soc. 97, 3436-3444 (1975).
5. P.Brown, G.R.Pettit and R.K.Robins, Org. Mass Spectrom. 2, 521-531 (1969).
6. H.-R.Schulten and H.D.Beckey, Org. Mass Spectrom. 7, 861-867 (1973).
7. M.Linscheid, Dissertation, Univ. Köln, 1975.
8. J.A.McCloskey, L.B.Barnes, P.F.Crain, K.J.Lyman and S.H.Bishop, Biomed. Mass Spectrom. 2, 90-96 (1975).
9. A.M.Lawson, W.A.Koenig, L.C.Smith, N.R.Earle and J.A.McCloskey, Adv. in Mass Spectrom. 5, 753-756 (1971).
10. T.S.Ro-Choi, Y.C.Choi, D.Henning, J.McCloskey and H.Busch, J. Biol. Chem. 250, 3921-3928 (1975).
11. R.Caprioli and D.Rittenberg, Proc. Natl. Acad. Sci. 60, 1379-1382 (1968).
12. H.Budzikiewicz and M.Linscheid, Biomed. Mass Spectrom. 4, 103-106 (1977).
13. G.Feistner, Diplomarbeit, Univ. Köln, 1977.
14. A.M.Lawson, R.N.Stillwell, M.M.Tacker, K.Tsuboyama and J.A.McCloskey, J. Amer. Chem. Soc. 93, 1014-1023 (1971).
15. F.W.Röllgen, U.Griessmann, H.J.Heinen and S.J.Reddy, Int. J. Mass Spectrom. Ion Phys. 24, 235-238 (1977).
16. H.J.Veith, in press.
17. H.J.Veith, Angew. Chemie 88, 762-763 (1976).
18. D.F.Hunt, C.E.Hignite and K.Biemann, Biochem. Biophys. Res. Comm. 33, 378-383 (1968).
19. J.L.Wiebers, Anal. Biochem. 51, 542-556 (1973).
20. J.J.Dolhun and J.L.Wiebers, J. Amer. Chem. Soc. 91, 7755-7756 (1969).
21. H.-R.Schulten and H.M.Schiebel, Z. Anal. Chem. 280, 139-144 (1976).
22. H.-R.Schulten and H.M.Schiebel, Nucleic Acid Res. 3, 2027-2031 (1976).
23. Unpublished results.
24. J.L.Wiebers and J.A.Shapiro, Biochem. 16, 1044-1050 (1977).
25. G.A.Chamock and J.L.Loo, Anal. Biochem. 37, 81-84 (1970).
26. D.R.Burgard, S.P.Perone and J.L.Wiebers, Biochem. 16, 1051-1057 (1977).
27. M.A.Posthumus, N.M.M.Nibbering, A.J.H.Boerboom and H.-R.Schulten, Biomed. Mass Spectrom. 1, 352-357 (1974).
28. K.Levsen and H.-R.Schulten, Biomed. Mass Spectrom. 3, 137-139 (1976).
29. J.L.Wiebers, Adv. in Mass Spectrom. 5, 757-760 (1971).
30. K.Taraz, private communication.
31. H.-R.Schulten, H.D.Beckey, A.J.H.Boerboom and H.L.C.Meuzelaar, Anal. Chem. 45, 2358-2362 (1973).
32. H.Budzikiewicz and M.Linscheid, Adv. in Mass Spectrom. 7, in press.