STRUCTURE ELUCIDATION OF DINUCLEOTIDES BY MASS SPECTROMETRY[1]

D. F. Hunt, C. E. Hignite, and K. Biemann

Department of Chemistry

Massachusetts Institute of Technology

Cambridge, Massachusetts 02139

Received September 18, 1968

Mass spectrometry has proved to be a valuable technique for the determination of the structure of nucleic acid components. Pyrimidine (Rice et al., 1965) and purine bases (Rice et al., 1967) and many nucleosides (Biemann et al., 1966; Biemann et al., 1966; Hall et al., 1966; Biemann et al., 1967; Baczynskyj et al., 1968) can be sublimed directly into the electron beam of the mass spectrometer. Nucleotides and the more polar nucleosides (e.g., guanosine, cytosine) on the other hand, must first be converted to their more volatile trimethylsilyl (TMS) derivatives before mass spectra can be obtained (McCloskey et al., 1968).

Aside from the structure of nucleosides the most important chemical aspect of this field is the nucleotide sequence in nucleic acids. Dinucleotides are the smallest subunits carrying sequence information, and their mass spectrometric behavior was thus investigated. Their very low volatility requires the use of a derivative which decreases the polarity by elimination of the acidic proton and modification of most or all of the polar groups. The poly(trimethylsilyl) derivatives were found to be useful for this purpose and gave excellent mass spectra. These derivatives were prepared of thirteen dinucleotides (Table I) commonly found in RNA, and their mass spectra were obtained. A molecular

1. Work supported by grants from the National Institutes of Health (Research grant GM 05472 and training grant GM 01523).

ion (M) and an ion due to the loss of a methyl group (characteristic of trimethylsilyl derivatives) were observed for all compounds investigated. Since the four common bases, adenine, guanine, uracil, and cytosine, as well as most of the rare bases, have a unique elemental composition the molecular weight thus determined indicates the bases present in the dinucleotide.

Figure 1 shows the mass spectrum of per(trimethylsilyl)-ApU as an example. The fragments a_1, a_2, $b_1(+2H)$, $b_2(+2H)$, c_1 and c_2, to be discussed below for reasons which will soon become apparent, are schematically indi-

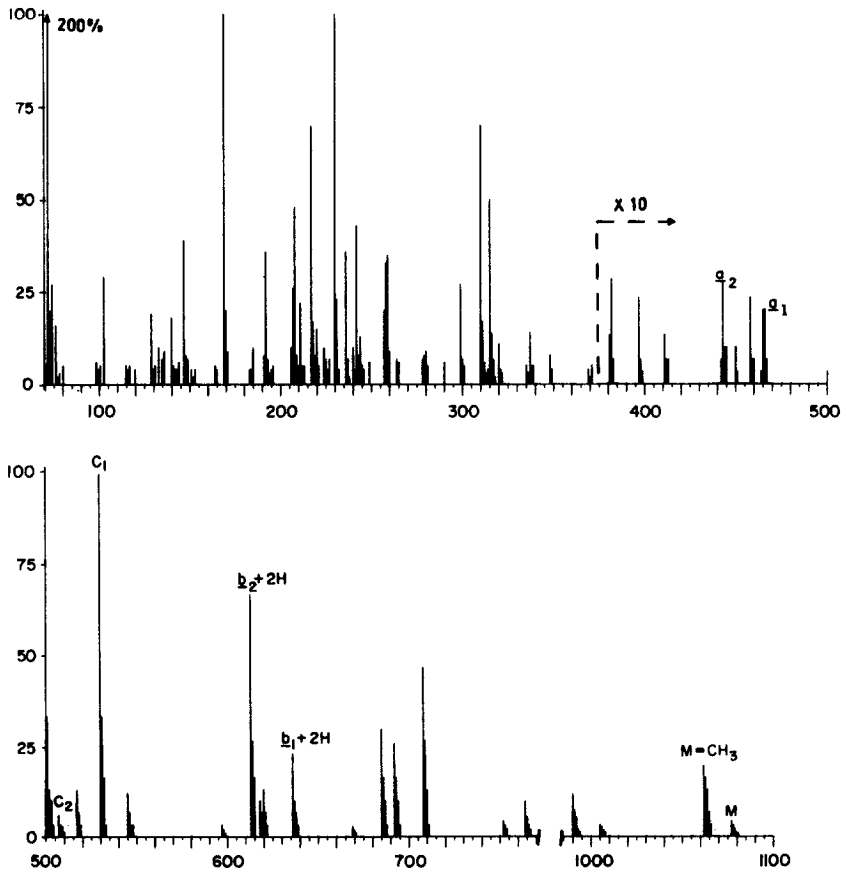
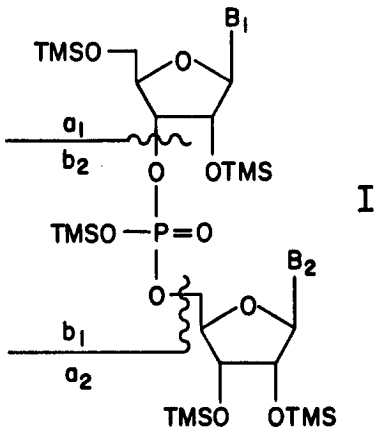


Figure 1. Mass Spectrum of hepta-(trimethylsilyl)-adenylyl-(3'-5')-uridine.

cated in structure I. The elemental composition of the fragment ions in the ApU derivative was confirmed by high resolution mass spectrometry. The subscripts 1 and 2 denote fragments containing the bases B_1 (the 5'-hydroxyl end) and B_2 (the 3'-hydroxyl end) of the dinucleotide respectively.



I

$$c_1, c_2 = b_1, b_2 + H - (TMSOH + CH_3)$$

Ions a_1 and a_2 are due to cleavage of the PO-C3' and PO-C5' bonds, respectively, with charge retention on the carbon atom. In keeping with the known behavior of trialkyl phosphates (McLafferty, 1956; Bafus et al., 1966) charge retention on the oxygen is always accompanied by transfer of two hydrogens to form a protonated phosphate ion, in this case $b_2(+2H)$ and $b_1(+2H)$. A third set of ions are related to the same process except that it involves only transfer of one hydrogen, loss of HOTMS, and a further loss of a methyl group from another TMS moiety (i.e., $c_1 = b_1 + H - (HOTMS + CH_3)$.

It will be noted that the ions a_1 in ApU and a_2 in UpA are isomeric and their mass alone does thus not distinguish these two dinucleotides. The mass spectra of such pairs showed, however, that the ions resulting from cleavage of the O-C5' bond are more abundant than those originating from the cleavage of an O-C3' bond. It is therefore the intensity ratio of these two peaks which indicates the sequence, because a_2 is more abundant than a_1. The same problem of isomerism exists for b_1 and b_2, and c_1 and c_2, and it was found that

Table I. Fragment Ions Used for Determining the Sequence of Bases in Silylated Dinucleotides

Compound	M^+	Fragment (m/e)		Relative Abundance*	Fragment (m/e)		Relative Abundance*	Fragment (m/e)		Relative Abundance*
		a_1	a_2		b_1(+2H)	b_2(+2H)		c_1	c_2	
A_pU	1077	466	443	23:30	636	613	25:75	530	507	100:5
U_pA	1077	443	466	20:100	613	636	20:20	507	530	42:4
A_pC	1076	466	442	15:47	636	612	12:15	530	506	100:1
C_pA	1076	442	466	14:100	612	636	1:10	506	530	5:8
G_pC	1164	554	442	8:21	724	612	6:1	618	506	100:1
C_pG	1164	442	554	15:25	612	724	6:65	506	618	100:7
A_pG	1188	466	554	65:50	636	724	20:33	530	618	100:25
G_pA	1188	554	466	3:100	724	636	1:8	618	530	10:1
U_pC	1053	443	442	100:80	613	612	35:25	507	506	75:12
C_pU	1053	442	443	70:75	612	613	8:8	506	507	100:50
G_pU**	1165	554	443	S:S	724	613	S:L	618	507	L:S
U_pG**	1165	443	554	S:S	613	724	S:L	507	618	L:S
A_pA	1100	466	466	40	636	636	40	530	530	100
G_pG**	1276	554	554	S	724	724	S	618	618	L
U_pU	1054	443	443	100	613	613	52	507	507	73
C_pC	1052	442	442	60	612	612	8	506	506	100

* The most abundant of the six ions = 100 units.

** The letters L and S (large and small) represent predicted values based on data obtained for the other thirteen dinucleotides.

the relationships $b_1 < b_2$ and $c_1 > c_2$ hold for the abundance ratio of at least two of these three pairs (Table I).

As Figure 1 shows there are many more ions formed from these derivatives upon electron impact. Their detailed discussion would go far beyond the scope of this paper, and it suffices to say that they fall generally into three groups: (1) ions resulting from further decompositions of the six fragments discussed above; (2) ions due to the base moieties as such or containing also part of the sugar (these do not carry any sequence information but can be used to further confirm the nature of the bases present in the dinucleotide); and (3) ions not containing the base moieties and thus appearing in all spectra at the same mass (i.e., 73, 169, 299, 501).

Because of the limited utility of dinucleotides in the sequencing of nucleic acids extension of the above work to the mass spectra of trimethylsilylated trinucleotides would be of even greater interest with the aim of developing a rapid sequencing method. The higher molecular weight of a completely trimethylsilylated trinucleotide (ca. 1700) and their lower volatility will cause some experimental difficulties, but we feel that it may be possible to overcome these. Work along these lines is in progress. The interpretation of such mass spectra in terms of nucleotide sequencing thus involves the testing of the experimental data for the presence and abundance of certain ions, the mass of which varies with that of the bases involved. Such processes lend themselves particularly well to computer techniques and programs we developed for the sequencing of peptides (Biemann et al., 1966) could be adapted for this purpose.

EXPERIMENTAL

General procedure for the Preparation of Silylated Dinucleotide Samples for Mass Spectrometry. The dinucleotide (500 µg) was placed in a 6 x 50 mm culture tube which was then sealed with a rubber cap. Pyridine (50 µl), N, O-bis-(trimethylsilyl)trifluoroacetamide (100 µl), and trimethylchlorosilane (20 µl) were introduced by means of a syringe. After three hours at room

temperature (or 30 min. at 60°C) 10-20 μl of the solution (corresponding to 25-50 μg of dinucleotide) was transferred to a capillary tube and evaporated at reduced pressure (0.1 mm). The capillary containing the sample was then introduced through the vacuum lock into the ion source of the mass spectrometer (CEC-21-104 or CEC-21-110). All of the derivatives examined were found to be volatile at 190-230°C (10^{-6} mm).

References

Baczynskyj, L., Biemann, K., and Hall, R. H., Science, 159, 1481 (1968).

Rafus, D. A., Gallegos, E. J., and Kiser, R. W., J. Phys. Chem., 70, 2614 (1966).

Biemann, K., Cone, C., Webster, B. R., and Arsenault, G. P., J. Am. Chem. Soc., 88, 5598 (1966).

Biemann, K. and Fennessey, P. V., Chimia, 21, 226 (1967).

Biemann, K. and McCloskey, J., J. Am. Chem. Soc., 84, 2005 (1962).

Biemann, K., Tsunakawa, S., Sonnenbichler, J., Feldmann, H., Dutting, D. and Zachau, H. G., Angew. Chemie (Intern. Ed.) 5, 590 (1966).

Eggers, S. H., Biedron, S. I., and Hawtrey, A. O., Tetrahedron Letters, 3271 (1966).

Hall, R. H., Robins, M. J., Stasiuk, L., and Thedford, R., J. Am. Chem. Soc., 88, 2614 (1966).

Hanessian, S., DeJongh, D. C., and McCloskey, J., Biochim. Biophys. Acta, 117, 480 (1966).

McCloskey, J., Lawson, A. M., Tsuboyama, K., Krueger, P. M., and Stillwell, R., J. Am. Chem. Soc., 90, 4182 (1968).

McLafferty, F. W., Anal. Chem., 28, 313 (1956).

Rice, J. M. and Dudek, G. O., J. Am. Chem. Soc., 89, 2719 (1967).

Rice, J. M., Dudek, G. O., and Barber, M., ibid., 87, 4569 (1965).