PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

# Applications of Mass Spectrometry in Structural Studies of Bioactive Compounds

K. Nakanishi and J. L. Occolowitz

*Phil. Trans. R. Soc. Lond. A* 1979 **293**, 3-11
doi: 10.1098/rsta.1979.0075

*Phil. Trans. R. Soc. Lond.* A **293**, 3–11 (1979)
*Printed in Great Britain*



# Applications of mass spectrometry in structural studies of bioactive compounds

By K. Nakanishi† and J. L. Occolowitz‡
† *Department of Chemistry,*
*Columbia University, New York, N.Y.* 10027, *U.S.A.*
‡ *Lilly Research Laboratories,*
*Indianapolis, Indiana* 46206, *U.S.A.*

The paper describes the role played by mass spectrometry in structural studies of molluscicidal saponins from a plant; the adducts formed between 7,12-dimethylbenz[a]anthracene and nucleic acids upon tissue culture with rat liver; and the G2 factor, the first natural mitotic hormone that regulates cell proliferation.

## 1. Molluscicidal saponins from dogwood (Hostettman *et al.* 1979)

In connection with our systematic isolation and structural studies (Kubo & Nakanishi 1977; Nakanishi 1977*c*) on biologically active compounds such as insect antifeedants, antimicrobial and molluscicidal agents and plant growth regulators from plants, it was noticed that the crude methanol extract from the common dogwood, *Cornus florida* L. (Cornaceae), contained a potent molluscicide. Snails are known to be hosts to a variety of parasitic nematodes which include schistosomes responsible for one of the most widespread tropical diseases, schistosomiasis. Accordingly, it may be possible that molluscicidal tests, which can be carried out readily in the laboratory with the use of the American species *Biomphalaria glabratus* (Nakanishi & Kulo 1977), would lead to agents capable of controlling the dangerous schistosomes either by killing the schistosome-transmitting snail or by direct action.

The methanol extract of the dried bark (200 g) of the plant was passed through a Sephadex LH 20 column and the biologically active fraction was submitted to droplet counter-current (d.c.c.) chromatography (Ogihara *et al.* 1976) with the solvent system indicated in figure 1. Neither preparative scale t.l.c. nor open column chromatography was able to separate the three fractions efficiently. The d.c.c. method used in the present case employed the upper layer of a $CHCl_3$ – MeOH – $H_2O$ system (7:13:8) as the stationary phase and the lower layer as the mobile phase. The apparatus was used in the descending mode at a flow rate of 10–20 ml/h. As shown in figure 1, after *ca.* 8 h a clear base-line separation of the three components was achieved in a total of only *ca.* 110 ml of solvent.

The structures of the three saponins Cf-1 to Cf-3 as determined by m.s. and $^{13}$C-n.m.r. are shown in figure 2. The interpretation of f.d.–m.s. data is exemplified in figure 3 for Cf-2. Acid hydrolysis of Cf-1 to Cf-3 yielded the same aglycone sarsapogenin ($M^+ = (C_{27}H_{44}O_3)^+$; 416) by $^1$H-n.m.r., $^{13}$C-n.m.r. and i.r. T.l.c. of the sugars obtained from the saponin hydrolysates showed the presence of galactose in Cf-1, xylose and galactose in Cf-2 and glucose and galactose in Cf-3. The sequence of sugar residues in Cf-2 can be established from the f.d.–m.s. spectrum (figure 3), which was obtained directly on an underivatized sample. The molecular mass (710) is clear from the triplet of peaks at $m/z$ 749, 733 and 711. The pairs at 601 and

579 show that fragment 132 is lost which corresponds to the pentose (xylose) whereas the peak at 417 corresponds to the further loss of the hexose (galactose). The ready cleavage of the inter-glycosidic linkage is presumably triggered by addition of $Na^+$ or $K^+$ to the glycosidic oxygen atoms (Schulten *et al.* 1978). The sugar sequence in Cf-2 is hence xylose–galactose–sarsapogenin. With Cf-3, f.d.–m.s. alone could not solve the sugar sequence because both glucose and galactose are hexoses. However, this was circumvented by partial hydrolysis of Cf-3 to Cf-1 which is galactose–sarsapogenin.

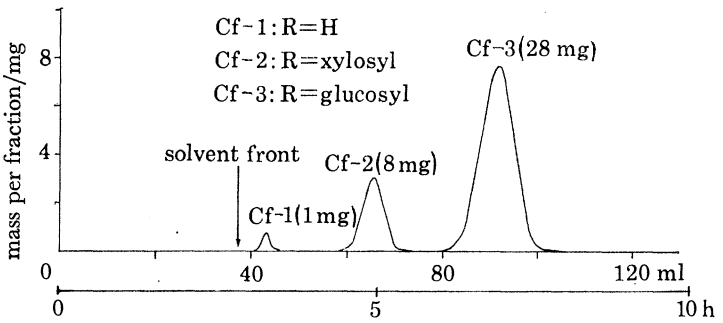
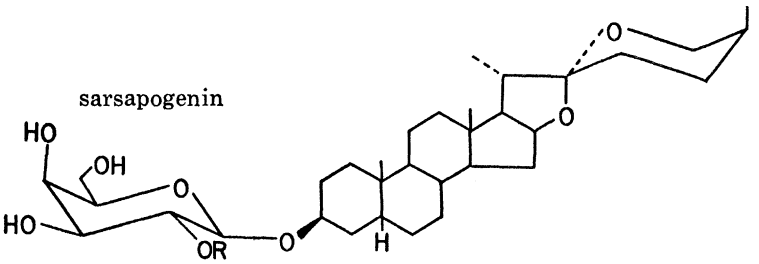



FIGURE 1. Droplet counter-current chromatography of molluscicidal fraction from dogwood bark.




| | m.i.c. (parts/$10^6$) |
|---|---|
| Cf-2 | 6 |
| Cf-3 | 12 |

FIGURE 2. Structures of Cf-1, Cf-2 and Cf-3 and minimum inhibitory concentration (m.i.c., parts/$10^6$) of molluscicidal activity against *B. glabratus* after 24 h. Cf-1 was inactive.

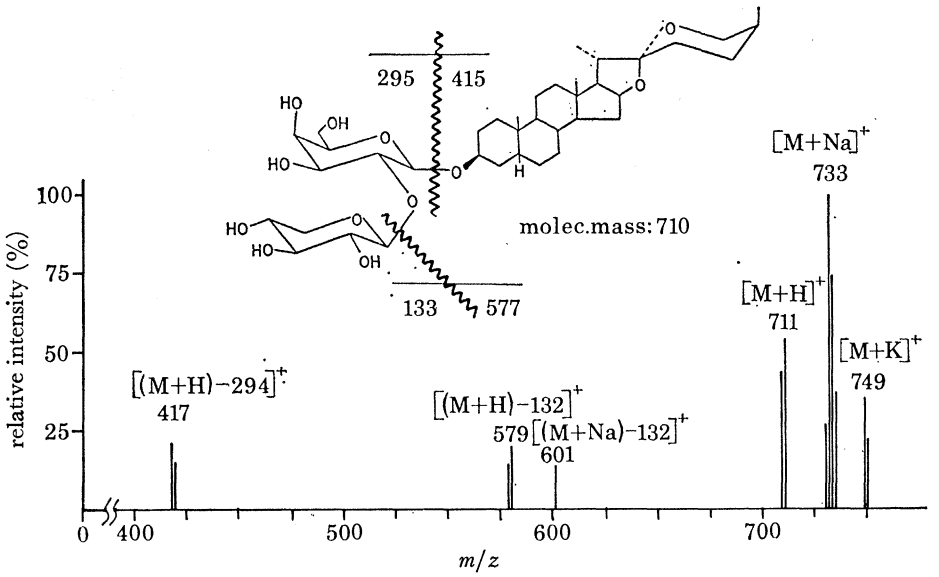



FIGURE 3. Field-desorption m.s. (f.d.–m.s.) of Cf-2 Varian-MAT Model 731, carbon dendrite emitter, resolution 1000, emitter current 21 mA.

The technique of plasma-desorption m.s. (f.d.–m.s.), developed recently by R. D. Macfarlane and coworkers (Macfarlane & Torgerdon 1976; Kasai *et al.* 1976), and which utilizes the fission energy of the radioactive element $^{252}Cf$ for generation of cations and anions, also successfully gave the molecular masses of underivatized saponins as shown in figure 4 for saponin Cf-3. The triplet of peaks at 786 $(M+2Na-H)^+$, 780 $(M+K)^+$, and 763 $(M+Na)^+$ are diagnostic for the molecular mass of this saponin.

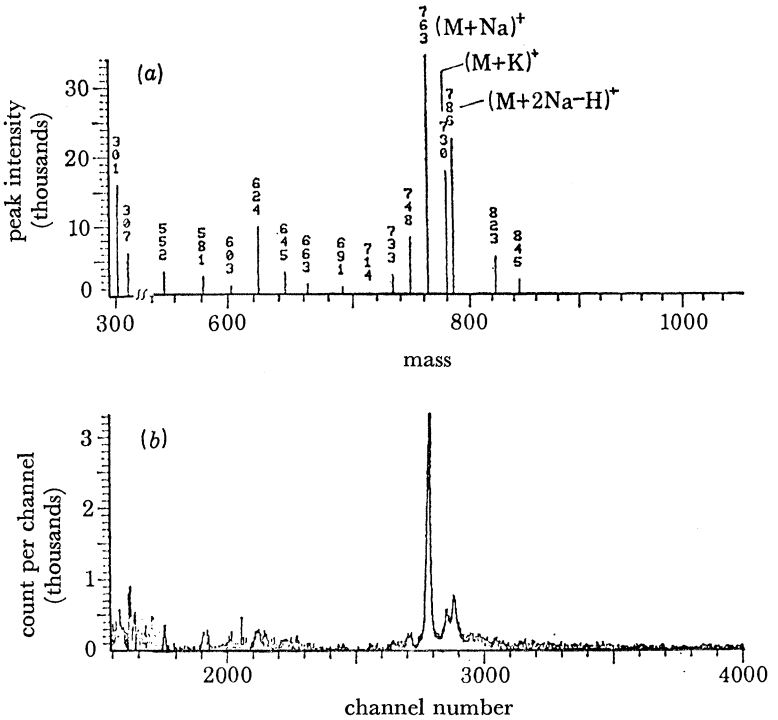



FIGURE 4. Plasma-desorption m.s. (f.d.–m.s.) of positive ions of Cf-3 in the molecular mass range 300–1000. Original data and its reduced form (i.e. conventional m.s. format).

The interglycosidic linkages were determined by $^1H$-n.m.r. (α or β linkage) and by $^{13}C$-n.m.r., which takes into account the chemical shifts of the anomeric and 2′-carbons of the galactose unit in Cf-2 and Cf-3: when compared with a 2′-unsubstituted galactopyranoside, the anomeric carbon and 2′-carbon are, respectively, shifted to higher field (*ca.* 3 parts/$10^6$) and lower fields (*ca.* 9 parts/$10^6$) (see also Tori *et al.* 1977; Konishi *et al.* 1978). Cf-2 is a new saponin but Cf-3 was described by Kawasaki & Yamauchi in 1963. Other plants contain molluscicidal factors, some of which are much more potent than the two saponins Cf-2 and Cf-3, and these are being investigated by techniques similar to those described here.

## 2. STRUCTURE OF A 7,12-DIMETHYLBENZ[a]ANTHRACENE (DMBA) ADDUCT G*-II FORMED FROM RAT LIVER TISSUE CULTURE (NAKANISHI *ET AL.* 1979)

Benzo[a]pyrene (BaP) and DMBA are two of the most potently carcinogenic polyaromatic hydrocarbons (PAH). Determination of the mode of its binding to nucleic acids in various tissues thus constitutes the first step in understanding the carcinogenic mechanism of such hydrocarbons on a molecular basis. In recent years the full structures of several adducts formed

by tissue culture and *in vivo* have been fully elucidated for BaP (Jeffrey *et al.* 1977) and DMBA (Kasai *et al.* 1977).

The general strategy for these microstructural studies is to incubate the tissue with randomly tritiated PAH, hydrolyse the isolated RNA and DNA fractions, purify the $^3$H-PAH–mononucleoside adduct by a series of chromatographic techniques and finally purify the $^3$H adducts by high-pressure liquid chromatography (h.p.l.c.). On the other hand, a non-radioactive derivative of the PAH which is suspected to be the ultimate carcinogen (i.e. the immediate precursor that leads to binding *in vivo*) is prepared synthetically and is incubated with various mononucleosides or homopolynucleotides under various conditions until products identical to those resulting from tissue culture are obtained; the identity is checked by the h.p.l.c. retention times of the hot $^3$H peaks and cold model peaks. Although structural studies are carried out with the cold adducts, even with these the available quantity is sub-milligram and therefore all microspectroscopic techniques have to be employed in full.

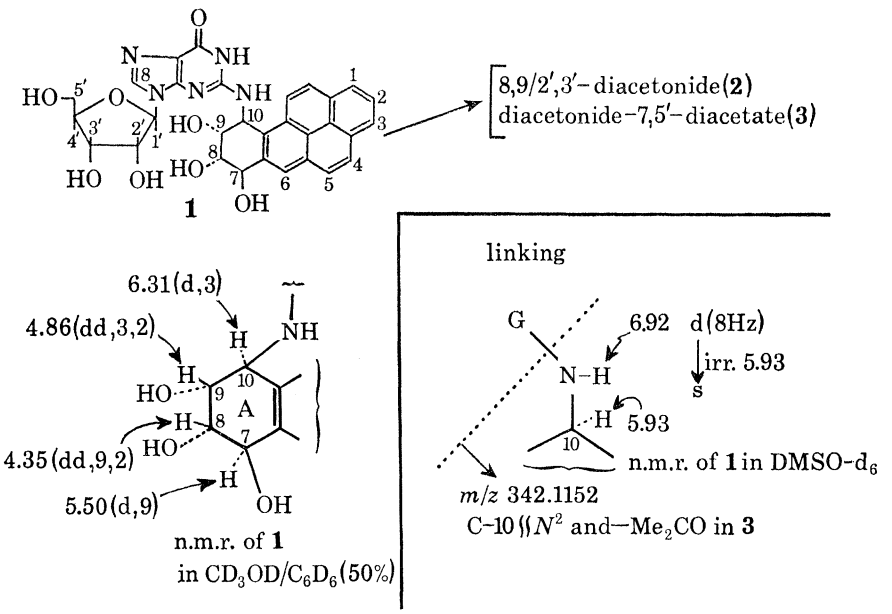



FIGURE 5. The full structure of major adduct resulting from BaP and nucleic acid (tissue culture).

The first PAH–nucleic acid to be fully identified is shown in figure 5 (Weinstein *et al.* 1976; Jeffrey *et al.* 1976*b*; Nakanishi *et al.* 1977). This guanosine adduct is the major product formed in human bronchial explants and rat liver tissue cultures (Jeffrey *et al.* 1977) and also by rat skin *in vivo* (Koreeda *et al.* 1978). The cold material used for structure studies was prepared by incubating racemic BaP-7β,8α-diol-9α,10α-oxide with polyguanylic acid. The substitution pattern on the reduced PAH ring was determined by micro-$^1$H-n.m.r. studies employing a combination of partial relaxed Fourier transform techniques, solvent shift and warming of the n.m.r. probe (Jeffrey *et al.* 1976*b*; Nakanishi 1976, 1977*a*). The point of attachment of the PAH moiety to the nucleoside, cannot however, be determined by $^1$H-n.m.r. because PAH contains hardly any protons. Fortunately, because the linkage was through $N^2$, positive evidence was available from the high-resolution m.s. as shown in the inset of figure 5: the peak at $m/z$ 342.1152 (1.9%, $C_{22}H_{16}NO_3$ = 342.1129) containing only one nitrogen atom and no additional carbon attached to the PAH fragment could only have arisen if the guanosine moiety

were linked via the exoskeletal $N^2$. Once this information was available it was possible to substantiate the conclusion by further $^1$H-n.m.r. studies. Thus the 6.92 parts/$10^6$ doublet (8 Hz) assignable to the —NH group (figure 5) became a singlet upon irradiation of the 10-H signal at 5.93 parts/$10^6$ (spectrum measured in DMSO-$d_6$); furthermore, the peak at 6.92 parts/$10^6$ disappeared upon addition of $D_2O$.

In most cases the PAH-nucleic acid adducts are fortunately through the exoskeletal amino groups of the guanosine (Moore *et al.* 1977) or adenosine (Straub *et al.* 1977), and therefore detection of the m.s. peak, which contains one nitrogen still connected to the PAH fragment, is possible.

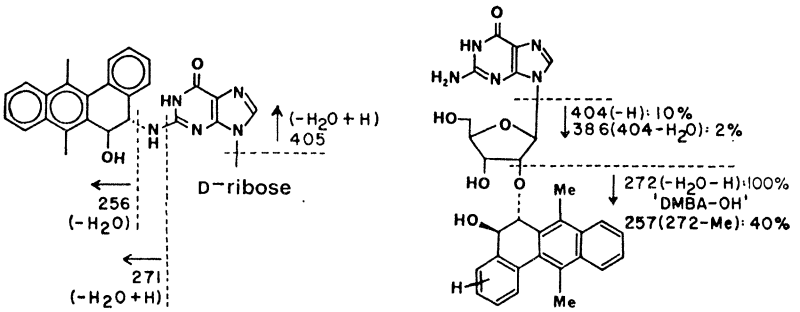



FIGURE 6. Left: general structures of adducts formed between DMBA-5,6-oxide and guanosine (Jeffrey *et al.* 1976*a*). Right: one of the adducts, G*-1b, resulting from incubation of DMBA with rat liver cells (in G*-Ia the substituents at C-5 and C-6 of the DMBA moiety are interchanged) (Kasai *et al.* 1977).

For DMBA, the four in-vitro products resulting from reaction of guanosine with DMBA-5,6-oxide under neutral conditions possessed the general structure shown on the left of figure 6 (the other three are position and configurational isomers of it). All peaks indicated on the left of figure 6 are based on high-resolution m.s. However, since none of the four coincided with tissue culture products, the in-vitro reactions were repeated at pH 9.5 when three products, G*-Ia, Ib and II, corresponding to tissue culture products, were obtained. As shown in figure 6, the PAH group is connected to the ribose 2′-OH in G*-1b (and in G*-Ia) (Kasai *et al.* 1977; Nakanishi 1977*b*). Only the peaks at $m/z$ 272 and 257 were measurable by high-resolution m.s. because of the limited sample amount; however, these two peaks are not diagnostic because the 272 peak simply corresponds to 7,12-dimethylbenzanthranol while the 257 peak corresponds to a demethylated benzanthranol. The important $m/z$ 404 and 386 peaks were only measured by low-resolution m.s. and hence were of limited value to pinpoint this unique structure; for example, a peak at $m/z$ 405 is present in III the compound on the left of figure 6: when the PAH portion is not bound to the exoskeletal nitrogen it is not easy to determine the point of linkage on the nucleic acid base. (Aflatoxin binds to $N^7$ of guanosine. Because of the positive charge on $N^7$, the chemical shift of the 8-H singlet is shifted downfield (Essigmann *et al.* 1977).)

We have therefore developed two new microspectroscopic methods for determining the substitution pattern on the base: one utilizes the strongly coupled c.d. curves to determine the dissociation constants while the other utilizes Fourier transform i.r.; both methods only require a few micrograms (Kasai *et al.* 1977; Nakanishi 1977*b*; Kasai *et al.* 1979). Thus the c.d./p$K$ method showed that G*-II had two apparent dissociation constants of less than 1.0 and 10.4. G*-II could be substituted on neither the exonuclear $N^2$ nor the ribose 2′-OH since all six products resulting from *trans*-opening of the oxirane had been identified earlier (Kasai *et al.*

1977; Jeffrey *et al.* 1976*a*) (figure 6). The possibility of a *cis* cleavage of the 5,6-oxide, although unlikely, could not be positively discarded. Overall evidence seemed to suggest substitution of C-8 of the guanosine but this was more or less based on a series of negative evidence. High-resolution m.s. of G*-II (figure 7) was not diagnostic because the only clear peaks at $m/z$ 272, 256 and 241 were derived from DMBA. However, it was gratifying when the e.i.–m.s. at 250 °C of G*-II peracetate (prepared from a few micrograms of G*-II) showed in addition a peak at 299 (obs. 299.1307; calc. for $C_{21}H_{17}NO$ 299.1310) which establishes that G*-II is linked through the guanosine C-8. Similarly, the high-resolution e.i.–m.s. of G*-IV, an in-vitro product with properties closely related to G*-II, showed a weak peak at 297 ($C_{21}H_{17}NO$)

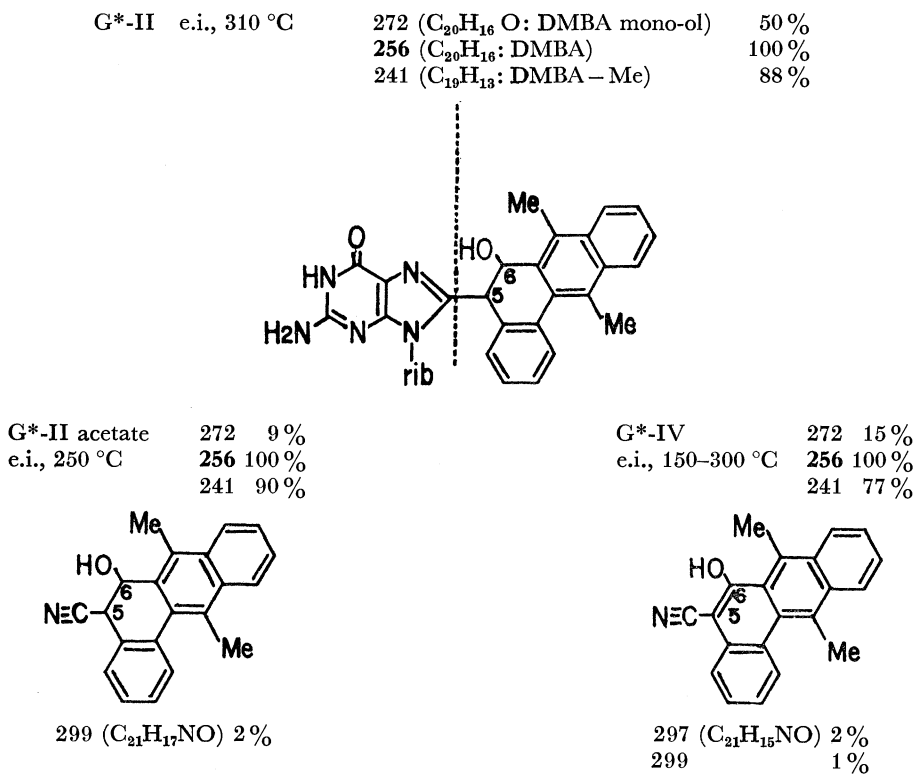



FIGURE 7. High-resolution m.s. of G*-II, G*-II acetate and G*-IV, with the use of a Varian MAT 731.

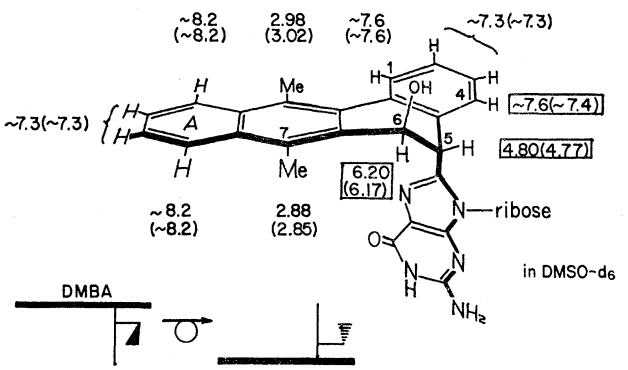



FIGURE 8. Full structure of G*-II (G*-III) formed in rat liver RNA (p.m.r.; 'major' products, antipodal c.d.).

as well as at 299, the former corresponding to the loss of two hydrogens from 299 (and hence fully aromatic).

The point of attachment thus being established, evidence from $^1$H-n.m.r., c.d., etc., leads to the structure shown in figure 8 for G*-II. However, it should be mentioned that products G*-Ia, Ib and II only constitute less than 10% of the total [$^3$H]DMBA adducts.

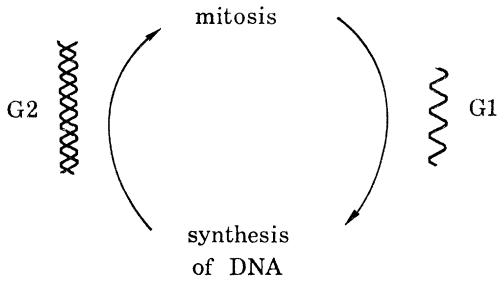



FIGURE 9. Mitotic cycle of pea cells.

TABLE 1. ISOLATION OF THE G2 FACTOR FROM *PISUM SATIVUM*

3200 cotyledons from seedlings 3 days old in 160 flasks

2 day incubation; Millipore filter; Dowex 50W–X4; 20% EtOH; LH–20;

10 g

Sephadex G–10, $H_2O$ (265 nm fraction (twice)

45 mg

$SiO_2$ t.l.c., $R_f$ 0.25, $CHCl_3$–MeOH–$H_2O$ (65:65:10)

10 mg

cellulose t.l.c., $R_f$ 0.40, MeCN–iso–PrOH–$H_2O$ (4:3:3)

2 mg

G–15, $H_2O$

50 µg

Process repeated five times: *ca.* 250 µg.

## 3. THE G2 FACTOR, THE FIRST NATURAL MITOTIC HORMONE THAT REGULATES CELL PROLIFERATION (LYNN *ET AL.* 1979; EVANS *ET AL.* 1979)

The mitotic cycle of garden peas (*Pisum sativum*) which regulates cellular processes to proliferate or not to proliferate is depicted in figure 9. Thus after 'M' (mitosis) in which the cells have divided, the cells go through a 'G1' stage (gap 1) before doubling their DNA content at 'S' (synthesis); the DNA synthesis is then followed by 'G2' stage (gap 2) in which the cells contain double-stranded DNA before passing on to M. When excised roots of peas are aseptically transferred to White's medium containing sucrose at 20 g/l, the meristematic cells are in the M, G1, S and G2 stages and a continuous growth of roots takes place. However, when the roots are put in White's medium not containing sucrose or carbon source, no growth of roots is seen after 24 h and the meristematic cells are now 80% in G1 and 20% in G2.

Recently, Evans & Van't Hof (1974) have characterized some physiological parameters of a G2 factor present in cotyledons that induces cell arrest in the G2 stage in both roots and shoots after seed germination; the term 'arrest' describes meristematic cells that have the potential to divide but are not actively progressing in the mitotic cycle.

The pure G2 factor was isolated from a total of *ca.* 16000 cotyledons according to the

procedure outlined in table 1, the final yield being about 250 μg. The factor thus obtained was extremely polar and insoluble in most organic solvents except DMSO and MeOH. Numerous attempts to measure the m.s. (c.i., f.d. and e.i.) were all unsuccessful except for a sample from one particular 50 μg batch (figure 10); knowing the structure now, we think that this particular batch was a mixture of zwitterion and its protonated form, the latter giving the satisfactory m.s. results (quite routinely!).

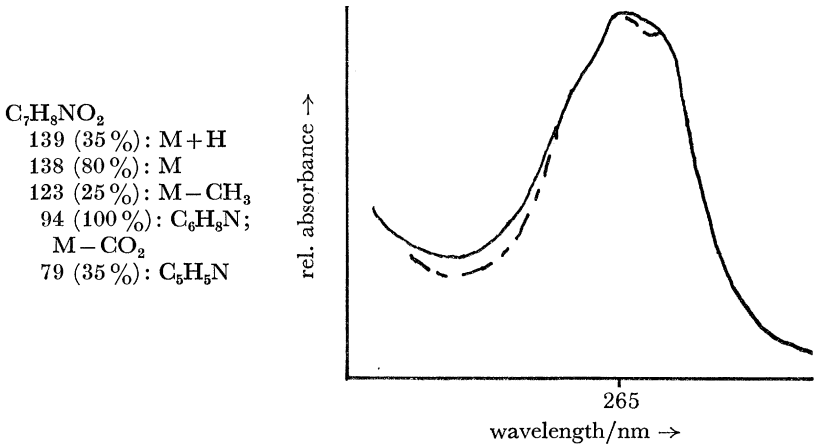



FIGURE 10. High-resolution e.i.–m.s. data (Varian MAT 731; 190 °C; 70 eV) and u.v. spectrum of G2 factor in $H_2O$ (—, pH 7; ——, pH 1; $\lambda_{max}$ = 265 nm; $\epsilon$ = 4070 $cm^{-1}$ $mol^{-1}$.

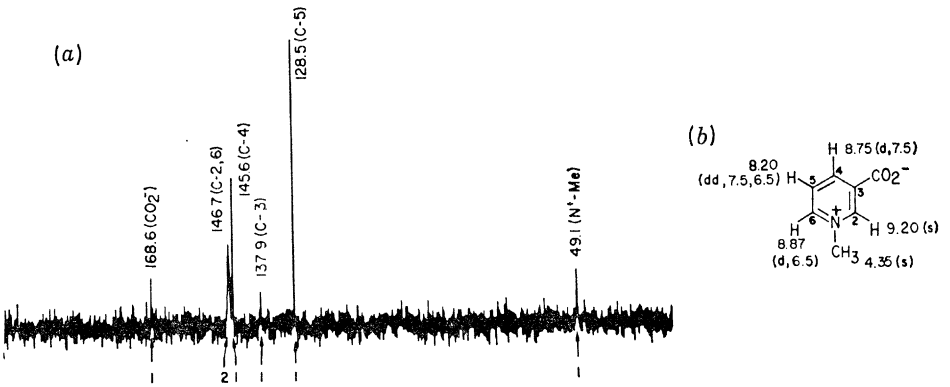



FIGURE 11. $^{13}$C-n.m.r. data of the G2 factor (trigonelline). F.t.–c.m.r. was carried out at 200 MHz (279 400 scans over 9 days, repetition rate 3.0 s) on 255 μg of sample. Integrated areas are shown beneath the scan.

The u.v. spectrum in water remained constant above pH 4, with $\lambda_{max}$ at 265 nm and a shoulder at 271 nm, but the 271 nm shoulder became more pronounced at pH 2. This u.v. behaviour strongly suggested that the G2 factor was zwitterionic and together with the m.s. and F.t.i.r. results, an *N*-methylnicotinic acid structure became the most plausible. A 9-day $^{13}$C-n.m.r. study, which had been running during these measurements, and a $^1$H-n.m.r. spectrum corroborated the structure which was finally confirmed by comparison with authentic material. The structure of the G2 factor thus proved to be trigonelline (*Merck Index* 1976) (figure 11), which was first isolated in 1895 from *Trigonella foenum-graecum* and subsequently from numerous other sources. However, it has not been associated with any hormonal activity before. Trigonelline is the first natural hormone to be isolated from plant or animal sources that regulates cell proliferation by cell arrest in the G1/G2 stage (1:1 mixture, see figure 9).

Identification of the G2 factor is only the first step leading to numerous detailed mechanistic investigations pertaining to natural hormonal regulation of the mitotic cycle and cellular differentiation in complex tissues.

The people at the Department of Chemistry, Columbia University, who undertook the studies described here are Dr Kurt Hostettmann and Dr Maryse Hostettmann (§1), Dr H. Kasai and Dr H. Komura (§2), Dr D. G. Lynn (§3) and Mr Iwao Miura (n.m.r.). In addition to the colleagues from other Institutes quoted in the respective references, we are greatly indebted to Professor R. D. Macfarlane, Texas A & M University (p.d.–m.s.), and Dr G. Jordan, Lederle Laboratories (F.t.i.r. of the G2 factor).

K. Nakanishi is also grateful to Professor A. W. Johnson and Professor J. H. Beynon for the invitation to give a talk at this meeting.

The studies were supported by NIH Grants CA 11572 and AI 10187.

## References (Nakanishi & Occolowitz)

Essigmann, J. M., Croy, R. G., Nadzan, A. M., Busby, Jr, W. F., Reinhold, V. N., Büchi, G. & Wogan, G. N. 1977 *Proc. natn. Acad. Sci. U.S.A.* **74**, 1870.

Evans, L. S. & Van't Hof, J. 1974 *J. exp. Cell Res.* **87**, 259.

Evans, L. S., Almeida, S., Lynn, D. G. & Nakanishi, K. 1979 *Science, N.Y.* **203**, 1022.

Hostettman, K., Hostettmann-Kaldas, M. & Nakanishi, K. 1979 *Helv. chim. Acta.* (In the press.)

Jeffrey, A. M., Blobstein, S. H., Weinstein, I. B., Beland, F. A., Harvey, R. G., Kasai, H. & Nakanishi, K. 1976*a* *Proc. natn. Acad. Sci. U.S.A.* **73**, 2311.

Jeffrey, A. M., Jennette, K. W., Blobstein, S. H., Weinstein, I. B., Beland, F. A., Harvey, R. G., Kasai, H., Miura, I. & Nakanishi, K. 1976*b* *J. Am. chem. Soc.* **98**, 5714.

Jeffrey, A. M., Weinstein, I. B., Jennette, K. W., Grzekowiak, K., Nakanishi, K., Harvey, R. G., Autrup, H. & Harris, C. 1977 *Nature, Lond.* **269**, 348, and references therein.

Kasai, H., Nakanishi, K., Macfarlane, R. D., Torgerson, D. F., Ohashi, Z., McCloskey, J. A., Gross, H. J. & Nishimura, S. 1976 *J. Am. chem. Soc.* **98**, 5044.

Kasai, H., Nakanishi, K., Frenkel, K. & Grunberger, D. 1977 *J. Am. chem. Soc.* **99**, 8500.

Kasai, H., Nakanishi, K. & Traiman, S. 1979 *J. chem. Soc. chem. Commun.* (In the press.)

Kawasaki, T. & Yamauchi, T. 1963 *Chem. pharm. Bull., Tokyo* **11**, 1221.

Konishi, T., Tada, A., Shoji, J., Kasai, R. & Tanaka, O. 1978 *Chem. pharm. Bull., Tokyo* **26**, 668.

Koreeda, M., Moore, P. D., Wislocki, P. G., Levin, W., Conney, A. H., Yagi, H. & Jerina, D. M. 1978 *Science, N.Y.* **199**, 778.

Kubo, I. & Nakanishi, K. 1977 *Am. chem. Soc. Symp. Ser.* **62**, 165.

Lynn, D. G., Nakanishi, K., Patt, S. L., Occolowitz, J. L., Almeida, S. & Evans, L. S. 1979 (Manuscript submitted.)

Macfarlane, R. D. & Torgerson, D. F. 1976 *Science, N.Y.* **191**, 920.

*Merck Index* 1976 9th edn, p. 1243. Rakway, N.J.: Merck.

Moore, P. D., Koreeda, M., Wislocki, P. G., Levin, W., Conney, A. H., Yagi, H. & Jerina, D. M. 1977 In *Drug metabolism concepts* (*Am. chem. Soc. Symp. Ser.* no. 44) (ed. D. M. Jerina), p. 127.

Nakanishi, K. 1976 In *Synthesis, structure and chemistry of tRNA and their components*, p. 447. Polish Academy of Sciences.

Nakanishi, K. 1977*a* In *Bio-organic chemistry*, vol. 3 (ed. E. E. van Tamden), p. 153. Academic Press.

Nakanishi, K. 1977*b* *Pure appl. Chem.* (In the press.)

Nakanishi, K. 1977*c* *Scr. varia pont. Acad. Sci.* **41**, 186.

Nakanishi, K. & Kubo, I. 1977 *Israel J. Chem.* **16**, 28.

Nakanishi, K., Kasai, H., Cho, H., Harvey, R. G., Jeffrey, A. M., Jennette, K. W. & Weinstein, I. B. 1977 *J. Am. chem. Soc.* **99**, 258.

Nakanishi, K., Komura, H., Miura, I., Kasai, H., Frenkel, K. & Grunberger, D. 1979 (Submitted.)

Ogihara, Y., Inoue, O., Otsuka, H., Kawai, K. I., Tanimura, T. & Shibata, S. 1976 *J. Chromatogr.* **128**, 218.

Schulten, H. R., Komori, T., Nohara, T., Higuchi, R. & Kawasaki, T. 1978 *Tetrahedron* **34**, 1003.

Straub, K. M., Meehan, T., Burlinghame, A. L. & Calvin, M. 1977 *Proc. natn. Acad. Sci. U.S.A.* **74**, 5285.

Tori, K., Seo, S., Yoshimura, Y., Arita, H. & Tomita, Y. 1977 *Tetrahedron Lett.*, p. 179.

Weinstein, I. B., Jeffrey, A. J., Jennette, K. W., Blobstein, S. H., Harvey, R. G., Harris, C., Autrup, H., Kasai, H. & Nakanishi, K. 1976 *Science, N.Y.* **193**, 592.



PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES