**Tetrahedron Letters No.45, pp. 4019-4025, 1965. Pergamon Press Ltd.
Printed in Great Britain.**

THE STRUCTURE OF THEAFLAVIN, A POLYPHENOL OF BLACK TEA

Y. Takino
Department of Agricultural Chemistry,
Tokyo University of Education, Tokyo, Japan

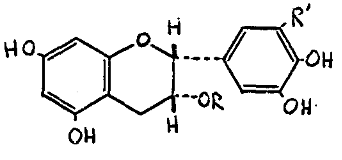
A. Ferretti, V. Flanagan, M. Gianturco[a]
Fundamental Research Department, The Coca-Cola Company,
Linden, N. J., U.S.A.

M. Vogel
School of Chemistry, Rutgers - The State University,
New Brunswick, N. J., U.S.A.

(Received 23 August 1965)

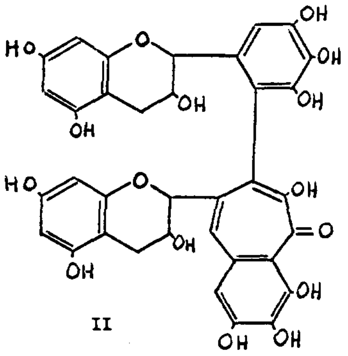
A crystalline orange-red polyphenol, which has been named theaflavin, has been known for some time to be present in black tea. This substance has been fairly extensively studied, because it is believed that the commercial value of black tea can be correlated with its content of theaflavin and of related substances[1].

In particular, Roberts suggested[1] that theaflavin originates from the coupling oxidation of epigallocatechin (Ia) and epigallocatechin gallate (Ib), and recognized that its molecule contains a benzotropolone nucleus. On the basis of the limited evidence available, the same author finally postulated that theaflavin might have structure II.



I

a: R = H; R' = OH
b: R = galloyl; R' = OH
c: R = H; R' = H



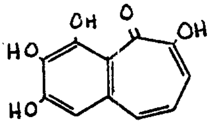
II

---

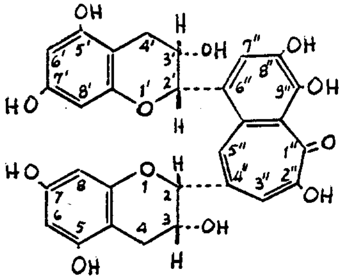
[a] To whom inquiries concerning this paper should be addressed.

These views were not accepted by Takino et al., who have recently suggested[2] that theaflavin derives from the coupling oxidation of epigallocatechin (Ia) and epicatechin (Ic), a process analogous to that proposed for the formation of purpurogallin (III) from pyrogallol[3,4]. Indeed, these authors have been able to obtain a crystalline substance identical with tea theaflavin by treatment of a mixture of Ia and Ic with the polyphenol oxidase of tea, and also by oxidation of the same mixture with potassium ferricyanide[3]. Takino et al. have finally proposed[5] for theaflavin the tentative structure IV.

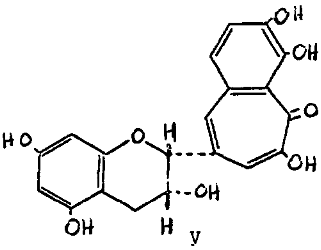
The postulation of this structure was based primarily upon the results of the synthetic experiments described above and upon those of some related oxidations, in addition to the following chemical and spectral evidence: a) the carbon-hydrogen analyses of theaflavin agree with the formula $C_{29}H_{24}O_{12}$, if it is assumed that its crystals contain 3.5 moles of water of crystallization; b) the carbon-hydrogen analyses of the crystalline acetate obtained from theaflavin agree with the formula $C_{47}H_{42}O_{21}$; c) the UV and IR spectra of theaflavin closely resemble those of categallin, a product formed by the coupling oxidation (ferricyanide or tea polyphenol oxidase) of a mixture of epigallocatechin (Ia) and catechol, and supposedly having structure V[6].



III



IV



V

More conclusive evidence for the correctness of the structure proposed by Takino et al. has now been obtained by n.m.r. and mass spectrometry. In particular, since the elemental analyses and the osmometrically determined molecular weights of theaflavin and its acetate cannot unequivocally establish the molecular formulas of these substances[b], it was thought desirable to determine the molecular weight of the acetate by mass spectrometry[d]. The value obtained (942) is in agreement with the formula $C_{47}H_{42}O_{21}$. Moreover, the acetyl number of the compound and the deuterium content of the deuterated analogue confirm the presence of nine acyl groups in the acetate; consequently, the molecular formula of theaflavin remains established as $C_{29}H_{24}O_{12}$[e].

---

b) The analytical results obtained in the course of the present investigation (Linden Laboratory) are as follows: Theaflavin (crystallized from MeOH and air dried). Found: C,55.92; H,4.98; no methoxyl present. Calcd. for $C_{29}H_{24}O_{12}\cdot 3.5\ H_2O$: C,55.50; H,4.94%.- Theaflavin (crystallized from MeOH and dried at room temperature under vacuum)[c]. Found: C,60.51, 60.70; H,4.46,4.63; methoxyl: 3.75. Calcd. for $C_{29}H_{24}O_{12}\cdot 0.7\ CH_3OH$: C,60.77; H,4.60%; methoxyl: 3.69. - Theaflavin acetate. Found: C,60.29, 60.18; H,4.66,4.51; O,35.45; acetyl number: 40.6. M.W. (osmometry):914 (acetone), 919($CHCl_3$). Calcd. for $C_{47}H_{42}O_{21}$: C,59.87; H,4.49: O,35.64; acetyl number: 41.0%; M.W.: 942.- Theaflavin deuteroacetate. Found: C,58.28; H,4.42; deuterium: 62.5 atom percent. Calcd. for $C_{47}H_{15}D_{27}O_{21}$: C,58.19; H,4.36%; deuterium: 64.28 atom percent.

c) The sample loses methanol, slowly and incompletely, when heated at 80° under high vacuum.

d) Theaflavin itself does not have sufficient thermal stability to yield a significant mass spectrum.

e) The mass spectrum of the acetate was kindly determined by Dr. H. M. Fales, National Institutes of Health, Bethesda, Maryland. The deuterated acetate could not be volatilized unchanged.

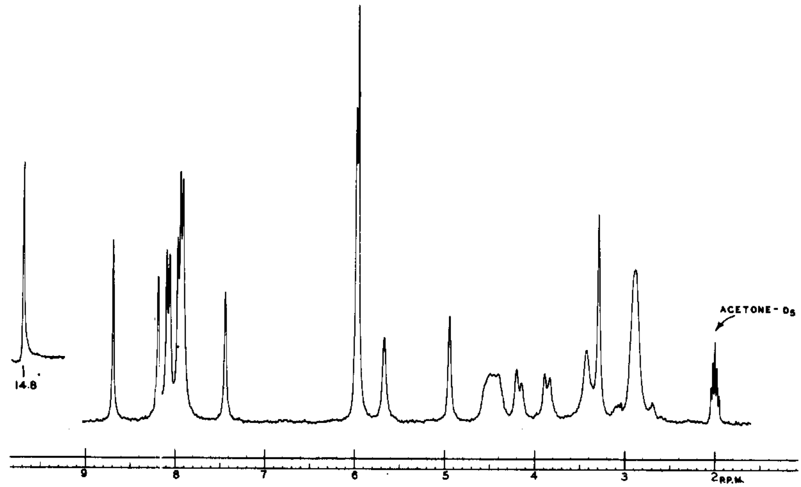


FIG. 1 - 100 Mc. N.M.R. spectrum of Theaflavin

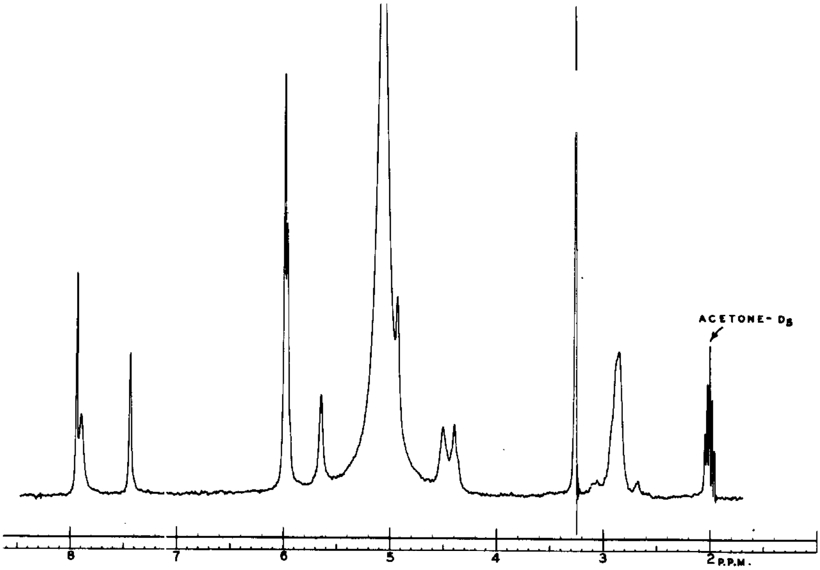


FIG. 2 - 100 Mc. N.M.R. spectrum of Theaflavin after addition of H_2O and HCl

The n.m.r. spectra of theaflavin (before and after addition of a drop of water and some gaseous HCl, Figs. 1 & 2) and that of theaflavin acetate appear to be completely consistent with structure III for the parent substance[f].

The absorptions at 14.84 and 8.70 p.p.m. are ascribed to the OH-protons at C-9" and C-2", respectively. The corresponding protons in purpurogallin (III) absorb at 15.00 and 8.68 p.p.m. The five remaining phenolic protons are singlets at 7.91, 7.97, 8.06, 8.10 and 8.19 p.p.m. The two alcoholic OH protons absorb at 3.86 and 4.17 p.p.m. and are split into doublets by the α-protons ($J \sim 6$ and 5 c.p.s. respectively). All the above lines average in with those of water in the spectrum of Fig. 2, due to rapid exchange.

The four protons at C-4 and C-4' appear as a broad absorption centered at 2.86 p.p.m.; the two protons at C-3 and C-3' also are broad, but absorb at 4.50 p.p.m. The two pairs of aromatic protons at C-6, C-8 and C-6', C-8' form two coincident AB patterns centered at 5.97 p.p.m. (J_{AB}=2 c.p.s.). The one remaining aromatic proton and the two tropolone-ring protons appear as singlets at 7.44, 7.90 and 7.94 p.p.m.; tentatively, we assign these to the C-3", C-7" and C-5" protons, respectively. (The 7.90 p.p.m. line is hidden by the absorptions of the phenolic protons until water

---

f) The spectra of theaflavin, its acetate and appropriate model compounds (catechin, epicatechin, purpurogallin and their completely acetylated derivatives) have been determined with a Varian A-60 spectrometer. The spectra of theaflavin shown in Figs. 1 & 2 were recorded with a Varian HA-100 spectrometer; for the use of this instrument we are indebted to Mr. Ross Pitcher, Varian Associates, Pittsburgh, Pennsylvania. Deuteroacetone was the solvent for the phenolic substances and $CDCl_3$ for the acetates. Approximately 20% (or saturated) solns. were used, with TMS as internal reference.

and acid are added).

Protons at C-2 and C-2' give rise to singlets at 4.95 and 5.67 p.p.m. The absorption at 5.67 p.p.m. is remarkably deshielded compared to that of the corresponding proton of epicatechin (4.90 p.p.m.). We conclude that the 5.67 p.p.m. resonance is due to the C-2' proton, which would be forced near the tropolone ring by steric effects; either a ring current in the tropolone ring or the anisotropy of the Δ 4" double bond would account for the deshielding effect on the C-2' proton.

The stereochemistry of the two pyran rings of theaflavin is most clearly indicated by the singlet nature of the C-2 and C-2' proton resonances. The corresponding proton in epicatechin is a singlet at 4.90 p.p.m.; coupling to the vicinal proton is small, consistent[7] with the expected dihedral angle near 60^{o}[8]. The corresponding proton in catechin appears as a doublet at 4.60 p.p.m. (J=7 c.p.s.) consistent with a dihedral angle near 180^{o} between it and the vicinal proton[g]. Therefore, the two pyran rings of theaflavin have the epicatechin stereochemistry.

The chroman substituent on the tropolone ring must be at the 4" position, and not at 3" or 5", or the two remaining ring protons would be strongly coupled and not appear as singlets. Also, there is a strong indication that the other chroman substituent is at the 6" position rather than, for instance, at the 7" position, or one could not account for the deshielding of the C-2' proton.

The spectrum of theaflavin crystallized from MeOH clearly shows the

---

g) Whalley (Ref. 8) concluded that catechin exists with the 3,4-dihydroxyphenyl and hydroxy substituents in diaxial positions. This cannot be the correct stereochemistry for catechin in acetone solution or a much smaller vicinal coupling would have been observed[7]. An analogous conclusion has been reached by Clark-Lewis and Jackman[9] with regard to the stereochemistry of catechin tetramethyl ether.

presence of solvent (methyl singlet at 3.28 and OH singlet at 3.42 p.p.m.). The integral of the MeOH absorption indicates somewhat less than one mole of MeOH per mole of theaflavin. The integrals of the n.m.r. absorptions of theaflavin agree well with those expected for structure IV and with the above assignments.

REFERENCES

1. E.A.H. Roberts in "The Chemistry of Flavonoid Compounds", (T.A. Geissman, Editor), 468, The MacMillan Co., New York (1962).
2. Y. Takino and H. Imagawa, Agr. Biol. Chem., 27, 319 (1963).
3. Y. Takino, H. Imagawa, H. Horikawa and A. Tanaka, ibid., 28, 64 (1964).
4. L. Horner and S. Göwecke, Ber., 94, 1267 (1961).
5. Y. Takino and H. Imagawa, Agr. Biol. Chem., 28, 255 (1964).
6. Y. Takino and H. Imagawa, ibid., 28, 125 (1964).
7. N.S. Bhacca and D.H. Williams, "Applications of NMR Spectroscopy in Organic Chemistry", pp. 49-52, Holden-Day, Inc., San Francisco, California (1964).
8. W.B. Whalley in "The Chemistry of Flavonoid Compounds", (T.A. Geissman, Editor), 444, The MacMillan Co., New York, (1962).
9. J.W. Clark-Lewis and L.N. Jackman, Proc. Chem. Soc., Lond., 165 (1961).