Tetrahedron, 1958, Vol. 3, pp. 168–174. Pergamon Press Ltd., London

SOME ASPECTS OF THE CHEMISTRY OF CIGARETTE SMOKE—I

G. R. Clemo
The Laboratory, Cherryburn, Mickley-on-Tyne

(*Received* 2 *March* 1958)

Abstract—Eight varieties of cigarettes have been "smoked", the smoke contents dissolved in methyl alcohol and separated into volatile in methyl alcohol, neutral, acidic, phenolic and basic fractions. The water soluble part of the dissolved mixture from seven of the varieties is acidic and basic from the eighth variety.

The acidic and phenolic fractions have been examined in some detail and their possible effect on health comented on.

In the course of a detailed study of the chemistry of cigarette smoke during recent years, numerous statistics have appeared, particularly in the United States of America and the United Kingdom, pointing to a definite connection between lung cancer and cigarette smoking in those countries. Thus the figures for the crude death rate from lung cancer per million in thirteen different countries for 1950–1951 show[1] such a wide variation that they raise the question as to whether there is any marked difference in the nature of the smoke from the cigarettes commonly used in different countries. There is no suggestion that cigar smoke, which is alkaline, is a causative agent, whereas cigarette smoke is usually strongly acidic. Pipe tobacco smoke is also stated to be acidic. The thirteen countries referred to above do not include any from behind the "curtain," but it was stated by Dr. Savittski at the Cancer Conference in Brazil in 1954 that there is no relation between tobacco and lung cancer in Russia and a like claim has been made for Poland.

After trying to obtain Russian cigarettes of the tubed "Dyeli" variety for over 2 years, a few hundred were obtained in Rumania recently, thanks to help from the Rumanian Academy, and what follows deals mainly with a comparison of the smoke from these and from seven brands from other countries, including an English variety on which the great majority of my work has been done. Although only 350 "Dyeli" have been smoked in three experiments (Example 4, Table 1), consistent results were obtained. All experiments have been done under standard conditions, but varying atmospheric conditions during smoking have some effect on the results. Further, in view of the smaller number of cigarettes smoked in Examples 2, 3, 5, 6, 7 and 8, the figures after the decimal point (except for the pH) may be rather less reliable.

In order to make the results as comparable as possible, they are based on 100 g of tobacco smoked rather than on the number of cigarettes, which vary considerably in weight.

The striking fact that emerges, however, is the pH of 8·9 for the "Dyeli," whereas all the other seven values are markedly acidic. This type with 3 cm of tobacco at the front and a 5 cm empty tube at the mouth end is much the most difficult to light and its 3 cm of tobacco takes as long to burn as the 7 to 8 cm of tobacco in the other varieties. Contrary to a common notion, the tubes condense very little of the smoke products (only 6 mg approximately per tube), but they probably cool the smoke appreciably before it enters the mouth. Incidentally the filter tips of a popular

---

[1] R. Doll, *Advances in Cancer Research* (Vol. III).

TABLE 1. PRODUCTS IN SMOKE FROM 100 g OF TOBACCO

Example No.	Tobacco tested	Crude hentriacontane (g)	Volatile in MeOH (g)	M (g)	Soluble in water		Soluble in NaOH solution (g)	Soluble in AcOH and HCl (g)	Esters (g)	M_N (g)
					Amount (g)	pH				
1.	English	0·23	0·35	6·8	2·7	4·2	2·1*	0·14	0·39	1·34
2.	U.S. (Chesterfield)	0·23	0·76	9·65	3·5	4·8	2·8	0·45	0·3	2·47
3.	U.S. (Camel)	0·36	1·08	8·95	3·1	5·2	2·4	0·5	0·54	2·27
4.	Russian (Dyeli)	0·26	0·8	8·75	3·2	8·9	2·8†	0·43	0·58	1·7
5.	Rumanian Marasesti	0·16	0·54	6·86	2·1	5·1	2·3	0·48	0·23	1·63
6.	French (Gitane)	0·2	0·88	8·75	2·98	5·7	2·32	0·77	0·77	1·9
7.	German (Peer)	0·16	1	9·2	3·7	4·9	2·77	0·15	0·77	1·74
8.	Black Russian (Sobranie)	0·24	(Lost)	9·6	3·5	4·9	3	0·33	0·74	1·95
9.	English (MeOH extracted)	0·05	Trace?	7	3·97	5·5	2·24	0·17	0·27	0·35

* 0·22 g phenolic. † 0·26 g phenolic.

English brand are a convenient source of at least one crystalline compound that can only be isolated from M_N (Table 1) with difficulty.

The pH of 8·9 (Example 4) is due to the presence of nicotine and the non-production of any quantity of water-soluble acids, although as can be seen from Table 1 the Dyeli smoke has almost the highest proportion of sodium hydroxide soluble "free acids" (including phenols) of any of the eight varieties. The presence of the nicotine is proved by distillation under reduced pressure, when, on the addition of picric acid to the distillate, only pure nicotine dipicrate separates almost at once.

The cigarettes in Example 8 are stated to be made in the United Kingdom from Russian tobacco.

Although most of my work has been done on the final neutral M_N fraction from Example 1 and some crystalline products isolated and analysed, it is to the acidic and phenolic constituents in Example 1 that my further remarks will be mainly directed.

It may be said, however, that, although M_N is predominately hydrocarbon in character, it contains very small amounts of aromatic hydrocarbons, but appreciable amounts of nitrogenous material (even after treatment with potassium) and oxygen-containing compounds, some of which have been obtained crystalline and analysed but not yet identified.

It is now generally agreed that the 3:4-benzopyrene content of M_N cannot account for the effect claimed for cigarette smoke. Thus none has been isolated therefrom and its presence to the extent of about 1 p.p.m. is based on spectrographic evidence, which is usually qualified by benzopyrene *or* an alkyl derivative.

This being so, the question arises as to what other factor or factors might separately or together be responsible for the lung cancers claimed to be due to the cigarette smoke. It seems that two such factors are present that could produce the result and they will be referred to at the end of this paper.

EXPERIMENTAL

Method of smoking and the isolation of the products

All carbon–hydrogen analyses and molecular-weight determinations are by Drs. Weiler and Strauss at Oxford.

The cigarettes have been "smoked" twelve at a time in an apparatus designed to simulate average natural smoking. After a variety of solvents for the absorption of the products had been tried in an effective five-fold absorbing system, methanol was found to be both effective and convenient. It has been used for all the work to be described.

The resulting methanolic solutions in all the eight examples deposited a greasy solid. (Only in the case of Example 4 did this lead to trouble during the experiments, through frequent choking at a constriction in the apparatus.) This grease distils (vacuum) and several crystallisations from benzene–ethanol raises the m.p. of the waxy plates to 69° (Found: C, 85·1; H, 14·5; mol. wt., 332. Calc. for $C_{31}H_{64}$: C, 85·3; H, 14·7 per cent; mol. wt., 436). Although the found molecular weight differs appreciably from the calculated value, there is little doubt that the purified material is hentriacontane. Wright and Wynder[2] claim that 90 per cent of the crude mixture is this hydrocarbon, but no mass-spectrometric study of it appears to have been made (see also Kosak *et al.*[3]). Further, hentriacontane has been shown to be present in natural tobacco leaf and Example 9 shows that little of it is formed by burning extracted tobacco.

Evaporation of the methanolic solution to dryness on the water bath (finally under reduced pressure) gives a syrup *M* (Table 1) and a yellow rather choking distillate due to volatile products. Some of these have been isolated, purified and analysed and appear (apart from a trace of phenols) to be mainly terpenoid in character. So far, only dipentene has been identified from the mixture of at least eleven compounds (Found: C, 87·6; H, 12·1. Calc. for $C_{10}H_{16}$: C, 88·2; H, 11·8 per cent). Its tetrabromide melts at 124–125° alone or mixed with authentic dipentene tetrabromide.

Separation of mixture M

Mixture *M* has been shaken with a mixture of water and benzene (twice), when a trace of an acidic tar is left, which is easily soluble in ethanol. The pH values (Table 1, column 7) refer to the aqueous solutions so obtained. The acids in this solution will include fatty ones simpler than caproic, which is only very slightly soluble in water, and other easily soluble types such as oxalic acid. If this solution, which contains 40 per cent of *M*, is distilled under reduced pressure, the distillate has a pH of 4·7 and readily reduces silver nitrate, thus indicating the presence of

---

[2] G. Wright and E. L. Wynder, *Proc. Amer. Ass. Cancer Res.* **2**, 159 (1956).

[3] Kosak *et al.*, *J. Nat. Cancer Inst.* **17**, 3375 (1956).

formic acid. A number of acids have been claimed to be present in tobacco smoke by earlier workers, and Kosak[4] states that formic, acetic, butyric, valeric and caproic acids are in the smoke from the destructive distillation of tobacco, whilst others such as succinic and citric might be present. The conditions in a burning cigarette would probably yield a very different mixture. In harvested tobacco leaf malic and citric acids amount to from 9 to 14 per cent of the weight of the dried leaves and oxalic to 2·5 per cent.

The mixture in the above-mentioned benzene solution (approximately 60 per cent of M) was shaken twice in ethereal solution with 4% sodium hydroxide. (If the dark alkaline solution is allowed to stand, a small amount of a sodium salt separates, from which a crystalline acid can be isolated.)

TABLE 2. COMPOSITION OF FRACTIONS

A—methyl esters, b.p. 83–150°/2 mm from the A_2 fraction of "free" acids
B—methyl esters undistilled from the A_2 fraction of "free" acids
C—methyl esters undistilled from the A_3 fraction
D—methyl esters undistilled from the acids soluble in light petroleum after hydrolysis (see p. 172).

	A (per cent)	B (per cent)	C (per cent)	D (per cent)
Eluted before C_{14}—contains 2 per cent of lauric*	32		10	13
C_{14} Myristic	4			1
C_{14} to C_{16}, X_1, X_2, X_3†	3·5			1·5
C_{16} Palmitic	32		4	26·5
C_{16} Palmitoleic	1·5			1·0
C_{16} to C_{18}, X_4, X_5‡	2	Same acids		1
C_{18} Stearic	4			1·5
C_{18} Oleic	7			9·0
C_{18} Linoleic and linolenic	13		2·5	33
C_{20} Arachidic	0·5			0·5
C_{20} X_6—unsaturated	0·5			
C_{26} Cerotic	0			
Non-volatile	0	12	83·5	15
	100	100	100	100

* The 30 per cent contains about 15 per cent of saturated and 15 per cent of unsaturated acids, of which 20 per cent are straight chain and 10 per cent branched or cyclic. There are at least fifteen and probably more compounds present.

† X_1, X_2 and X_3 are neither branched nor cyclic; one is unsaturated.

‡ X_4 and X_5 are saturated. Neither is highly branched nor cyclic. One appears to be C_{17} and the other an isomer of C_{18}.

Acidification of the dark alkaline solution of the *free* acids and phenols in the smoke and ether extraction left 0·36 g of dark solid acids (A_1), insoluble in benzene also, but easily soluble in ethanol. These acids are complex non-fatty acids and probably more highly oxygenated compounds. The "acids" recovered from the ether (1·38 g) were separated from the phenols (0·22 g) and found to be only partly

---

[4] Kosak, *Experimentia* **10**, 69 (1954).

soluble in boiling light petroleum (A_2) (even C_{22} behenic acid is readily soluble in this solvent), and as shown below this fraction is a very complex mixture. Little information has so far been obtained on the acids insoluble in light petroleum but soluble in ether (A_3) and none on those insoluble in ether (A_1). From the acids soluble in light petroleum palmitic and one or two other acids have been separated in a pure form by conversion to methyl esters, followed by chromatography, distillation and hydrolysis. A vastly deeper insight into the composition of this fraction has been obtained, however, through the kindness of the Research Director of Unilever, Limited, Port Sunlight, where Mr. J. Clifford, by a combination of chemical methods followed by vapour-phase chromatographic analysis of the methyl esters, has obtained the results shown in Table 2.

Phenols

Dry tobacco leaves are said to contain on an average 2·7 per cent of their weight of phenols and tannins, including chlorogenic acid, and, although Commins and Lindsay[5] identified phenol, the three cersols and dihydroxybenzenes, and 1- and 2-naphthols in cigarette smoke by means of the ultra-violet peaks of their methyl esters, no one seems to have realised the extent or complexity of the phenolic mixture in the smoke.

Thus the 0·22 g of phenols (Example 1, Table 1, column 5, from 89 cigarettes) contains 0·21 g of ether-soluble material, of which 0·11 g is easily volatile in steam, leaving 0·1 g of *C* in the distilling flask. The 0·11 g on distillation gave 0·09 g, b.p. 50–55°/2 mm of *A* and the remainder up to 90°/2 mm of *B*.

The Research Dept. of I.C.I., Ltd., Dyestuffs Division, has very kindly examined these fractions by gas–liquid chromatography and the mass-spectrometer and they report that *A* consists of phenol (50–60 per cent), *o*-cresol (10 per cent), *m*- and *p*-cresols (20 per cent) and guaiacol (10 per cent) with small amounts of three other substituted phenols. In addition to small amounts of the above, *B* contains several more complex compounds, including three of the hydroaromatic type such as tetrahydro-2-naphthol. Fraction *C* consists of more complex phenols and neither gas–liquid chromatography or infra-red spectroscopy shed light on its composition. The mass-spectrometer examination of the fractions obtained at 250° and 350° (the materials are stable under these conditions) gave a large number of peaks. At the lower masses the groups of peaks are separated by mass 14 (i.e. CH_2), whilst in the 350° fraction the outstanding peak was at mass 272 with other prominent ones at 298 and 342. The 272 peak could be due to quite a number of polycyclic phenols, including a dihydroxybenzopyrene.

The benzene-soluble material not extracted by sodium hydroxide

This was extracted successively three times with 10% acetic acid and then three times with 10% hydrochloric acid to remove remaining basic substances. (Most of the nicotine is present in the aqueous acidic solution (column 4).) Then the esters present in the mixture were hydrolysed by being heated under reflux with methanolic potash and the neutral mixture M_N was obtained (see Table 1).

The alcohols liberated in the hydrolysis are also a considerable mixture, including

---

[5] B. T. Commins and A. J. Lindsay, *Brit. J. Cancer* **10**, 504 (1956).

crystalline ones, but only the simplest—b.p. 90°/2 mm—has been analysed so far (Found: C, 83·0, 83·07, 83·1; H, 10·5, 10·8, 10·8; mol. wt., 210, 230. $C_{16}H_{24}O$ requires: C, 82·8; H, 10·3 per cent; mol. wt., 232. $C_{18}H_{28}O$ requires: C, 83·1; H, 10·8 per cent; mol. wt., 260). Its reduction has not been examined and so no deduction can be made as to its ring nature.

In Example 9 (Table 1) the English cigarettes have been extracted with methanol before smoking, whereby 43·7 per cent of the weight of the tobacco is removed, of which 39 per cent is water soluble. The "tobacco" remaining on "smoking" gives practically no methanol-volatile materials and only little hentriacontane. The high proportion of the weight of a cigarette soluble in water will be due mainly to carbohydrate-type materials, and this 39 per cent when smoked will be responsible for many of the products found in *M*. It is stated that the higher the carbohydrate content the better the cigarette! Cigar tobacco contains very little carbohydrate materials. Unfortunately the literature contains little exact knowledge on the carbohydrates or terpenoids in tobacco and none, so far as I am aware, on their subjection to 770°, approximately the temperature in the glow of a cigarette.

CONCLUSIONS

It has been shown that the acidic and phenolic constituents of cigarette smoke (Example 1) form a very complex mixture (and M_N is no less complex) and, whilst the 3:4-benzopyrene content of smoke is much too little to account for its postulated action, insufficient attention has been given to other potentially harmful compounds present in considerable amounts.

Attention is now called briefly to the possibility that some of the acidic and phenolic constituents might account for many of the observed results. Broadly speaking, heavy cigarette smokers in cities show, after some years, a greater incidence of lung cancer than similar smokers in rural areas and a very much greater incidence than non-smokers. That some of the latter fall victims is to be expected when fully one-third of the smoke of a cigarette drifts from its burning end and affects those within reach, particularly on public transport.

Other variations such as shown by the United States–United Kingdom statistics on lung cancer and the Russian claim to immunity will probably only be explained when the nature of the constituents of the various smokes is fully known and their actions on living cells understood.

It is known that some aromatic hydrocarbons are soluble in acids and it has now been found that formic acid exerts a selective solvent action on some hydrocarbons in mixtures. The lungs of city dwellers particularly are soot covered and we showed[6] that city smoke contained at least one potent carcinogen that is not benzopyrene. It has been claimed that carcinogens absorbed on soot are inactive, but since 50% acetic acid and cold propionic and butyric acids will dissolve appreciable amounts of benzopyrene and acetic acid will also extract the above-mentioned carcinogenic agent from soot, it could, therefore, elute such a carcinogen from lung soot and thus render it active.

As for the phenol present (at least 1 mg per cigarette with the cresols and possibly considerably more, since these compounds being appreciably soluble in water would

---

[6] G. R. Clemo, E. J. Miller and F. C. Pybus, *Brit. J. Cancer* **9**, 137 (1955).

probably be partly removed in the water extraction under Column 4), its irritant and corrosive action is universally recognised. It could largely explain the bronchitic action of cigarette smoke and also render lung sites susceptible to attack by other smoke constituents even if its prolonged irritant action did not itself produce cancer.

Acknowledgements—In addition to those mentioned above, thanks are due to the Royal Society for enabling me to start this work, to Mr. J. H. Smith for making the smoking apparatus, and to the special M.R.C. fund for financing the cost of analyses, etc.