JOURNAL DE LA SOCIÉTÉ STATISTIQUE DE PARIS

SHOU-SIN SUNG
PAUL DAMIANI

Analyse multi-dimensionnelle du pouvoir carcinogène des hydrocarbures aromatiques polycycliques

Journal de la société statistique de Paris, tome 117 (1976), p. 14-24
<http://www.numdam.org/item?id=JSFS_1976__117__14_0>

© Société de statistique de Paris, 1976, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (http://publications-sfds.math.cnrs.fr/index.php/J-SFdS) implique l'accord avec les conditions générales d'utilisation (http://www.numdam.org/conditions). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM
Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
http://www.numdam.org/

III

ARTICLES

ANALYSE MULTI-DIMENSIONNELLE DU POUVOIR CARCINOGÈNE DES HYDROCARBURES AROMATIQUES POLYCYCLIQUES (1)

Cette analyse fait suite à celle faite récemment par Herndon qui a établi une formule de régression multiple exprimant le pouvoir carcinogène des hydrocarbures aromatiques polycycliques non substitués en fonction des indices de réactivité des régions K *et* L *et un troisième facteur représentant la dimension moléculaire. Dans la présente analyse, nous allons montrer que le rôle apparent de ce troisième facteur pourrait être remplacé et expliqué par deux autres facteurs : énergie d'excitation électronique et potentiel d'ionisation moléculaires.*

This analysis follows the one performed recently by Herndon who established a multiple regression formula expressing the carcinogenicity of unsubstituted polycyclic aromatic hydrocarbons in function of K *—and* L *— region reactivity indices and a third factor representing the molecular size. In the present analysis, we are showing that the apparent role of molecular dimension can be explained by two other factors : molecular electronic excitation and ionization potential energies.*

Diese Analyse folgt derjenigen Herndons, welcher eine vielfache Regressions-Formel ausgestellt hat, die die krebsförderuden Eigenschaften der polyzyklischen aromatischen Kohlenwasserstoffe anbetreffs ihrer Reaktivitat in der K *— und* L *—Zone angibt, und in der ein dritter Faktor die Molekulgrosse darstellt. In der vorliegenden Analyse zeigen wir, dass die scheinbare Rolle der molekulären Grösse anhanden zweier weiteren Faktoren, welche die elektronische Erregung des Molekels und die Potential-Energie der Ionisation darstellen, erklärt werden kann.*

Dans un travail précédent, l'un de nous [1] a procédé à une analyse statistique simple sur les données qui ont été à l'origine de la formulation de la théorie des régions K et L [2-5] en matière de carcinogenèse par les hydrocarbures aromatiques polycycliques. Les résultats de cette analyse suggèrent une remise au point de la théorie dans un contexte nouveau lui permettant de se reposer sur une base nouvelle plus confortable. Compte tenu de la réserve habituelle à faire au sujet du risque d'erreur possible, on peut dire que la réactivité par la région K dans une molécule paraît être un facteur important pour l'apparition de son pouvoir carcinogène. Le rôle de la région L était moins manifeste du fait probablement de la faiblesse de l'effectif analysé. Nous verrons, en effet, qu'à la suite de la présente analyse, son rôle

1. L'essentiel de cet article a fait l'objet d'une communication au Ve Congrès international de biophysique qui s'est tenu du 4 au 9 août 1975 à Copenhague (Danemark).

Journal de la Société de statistique de Paris, tome 117, n° 1, 1976.

sera également confirmé. L'essentiel de la théorie des régions K et L est donc, à quelque nuance près, conservé. Les difficultés n'apparaissent que lorsque l'on prétend que les actions des régions K et L sont les seules à la fois nécessaires et suffisantes pour la détermination du pouvoir carcinogène. On pouvait se douter que c'était trop beau pour être vrai. Car, par intuition, on pressent qu'un phénomène aussi complexe que le processus de cancérisation ne peut ne comporter que deux facteurs seulement. Herndon [6, 7] dans une récente revue sur la question a rappelé que de nombreux auteurs ont formulé, dès l'époque de la présentation de cette théorie, des critiques sur ce point. Leur conviction est aussi que le pouvoir carcinogène ne peut être corrélé avec un petit nombre de paramètres tels que les indices électroniques de réactivité. Herndon attribue cependant les causes des « défaillances » de ces indices à l'imperfection de la méthode de calcul (simple méthode de Hückel) par laquelle ces indices ont été obtenu, tout en affirmant que les paramètres calculés par la méthode de « self-consistent Hückel MO » (SC-β [8], [9]) et par la méthode de « Pariser-Parr-Pople SCF-MO » [10], deux méthodes infiniment plus élaborées, ont donné essentiellement les mêmes conclusions. Cet auteur a procédé en 1974 à deux analyses statistiques [6, 7] sur les indices de réactivité calculés par lui-même avec deux méthodes différentes, toutes deux extrêmement simples. Ces deux analyses, l'une portant sur 51 cas, l'autre sur 58 cas, ont amené des résultats analogues. Aussi dans ce qui suit, pouvons-nous ne nous référer qu'à la dernière [7].

Dans le tableau I, figurent 62 hydrocarbures aromatiques polycycliques non-substitués dont le pouvoir carcinogène a été expérimentalement plus ou moins étudié avec leur divers paramètres électroniques. L'analyse de Herndon [7] porte sur les 57 premiers + le naphtho (2,3-e) pyrène. A l'aide des données des 10 premiers composés dont le pouvoir carcinogène a été particulièrement bien exploré et l'indice d'Iball [11] de ce pouvoir proposé par certains expérimentateurs, Herndon a établi la formule de régression multiple suivante :

Indice de carcinogenicité $= -47,8 + 274,5\, P_K - 179,1\, P_L - 4,3\, \Delta$ (I) où P_K et P_L sont des indices de réactivité des régions K et L que Herndon a assimilé aux indices de liaison de Pauling [12] et Δ représente la dimension moléculaire dont la valeur est : $\Delta = | N - 24 |$, N étant le nombre d'atomes de carbone que contient l'hydrocarbure. Le coefficient de corrélation multiple R (I) de l'équation I est : R (I) $= 0,935$. A l'aide de cette équation, Herndon a calculé le signe (+ pour carcinogène et — pour non carcinogène) de carcinogénicité des 48 autres hydrocarbures qu'il s'est proposé d'analyser et a trouvé que pour 33 d'entre eux, le signe correspondait à celui déterminé expérimentalement alors que pour les 15 non carcinogènes restants, le signe a été trouvé positif.

Au lieu d'attribuer les « défaillances » des indices de réactivité à l'imprécision de leur calcul d'obtention, comme l'a fait Herndon, nous nous sommes demandés, quant à nous, si l'adjonction de nouveaux facteurs dans la formule de régression ne pouvait aussi améliorer son coefficient de corrélation. Nous avions remarqué [13], en effet, que la corrélation (1) entre le pouvoir carcinogène et certaines énergies d'excitation moléculaires qui a été signalée pour la première fois par Mason [14, 15], est au moins comparable en degré à celle entre ce même pouvoir carcinogène et les indices de région K, et ce malgré quelque critique qu'elle a rencontrée [16, 17]. Les histogrammes des figures 1 et 2 illustrent bien cette comparaison. Aussi avons-nous cru intéressant d'insérer ce « facteur » dans la formule de régression du type de l'équation (I) et de voir ce qui en résulte. Les énergies d'excitation électroniques en question seront désignées par ΔE_1 et ΔE_2 qui représentent respectivement la différence

1. Le terme « corrélation » est utilisé ici abusivement comme le font généralement les biophysiciens dans un sens plus large que celui attribué par les statisticiens. Il désigne des liens de tout genre, monotone ou non, entre variables aléatoires qualitatives ou quantitatives.

TABLEAU I

K : indices simples de région K ; L : indices simples de région L ; N : nombre d'atomes de carbone que contient l'hydrocarbure ; ΔE_1 et ΔE_2 : deux énergies d'excitation électronique calculées par la simple méthode de Hückel.

Hydrocarbures	K	L	N	ΔE_1	ΔE_2	Pouvoir carcinogène
1. Benz (a) anthracène	1,03	3,42	18	0,90	1,17	5
2. Tribenzo (a, e, i) pyrène	1,19	3,67	28	0,79	1,08	17
3. Benzo (g) chrysène	1,15	4,82	22	1,06	1,24	18
4. Dibenz (a, h) anthracène	1,05	3,51	22	0,94	1,16	26
5. Naphtho (2, 3-b) pyrène	1,00	3,44	24	0,61	0,95	27
6. Dibenzo (a, e) pyrène	1,14	3,87	24	0,84	1,16	50
7. Dibenzo (a, h) pyrène	1,08	3,44	24	0,61	1,10	68
8. Benzo (a) pyrene	1,03	3,68	20	0,74	1,17	72
9. Dibenzo (a, i) pyrène	1,01	3,46	24	0,68	1,02	74
10. Dibenzo (a,l) pyrene	1,02	3,81	24	0,80	1,07	82
11. Phénanthrène	1,07	3,07	14	1,21	1,37	—
12. Benzo (c) phénanthrène	1,10	4,61	18	1,13	1,23	+
13. Chrysène	1,12	3,95	18	1,04	1,31	+
14. Benzo (c) chrysène	1,09	3,99	22	1,10	1,15	+
15. Dibenz (a, j) anthracène	1,04	3,51	22	0,98	1,11	+
16. Phénanthra (2, 3-b) pyrène	1,03	3,46	28	0,64	1,00	—
17. Naphtalène	1,26	3,68	10	1,24	1,62	—
18. Anthracène	1,20	3,31	14	0,83	1,41	—
19. Naphthacene	1,19	3,25	18	0,59	1,07	—
20. Triphénylène	1,38	4,55	18	1,37	1,37	—
21. Pentacène	1,18	3,18	22	0,44	0,84	—
22. Benzo (a) naphtacène	1,01	3,27	22	0,65	1,01	—
23. Dibenz (a, c) anthracène	1,24	3,49	22	1,00	1,21	—
24. Dibenzo (e, l) pyrène	1,38	4,25	24	1,11	1,23	—
25. Anthracéno (2, 1-a) anthracène	1,11	3,38	26	0,70	1,14	—
26. Tetrabenz (a, c, h, j) anthracène	1,38	3,65	30	1,15	1,18	—
27. Dibenzo (a, l) pentacène	1,02	3,44	30	0,54	0,90	—
28. Dinaphtho (1, 2-b, k) chrysène	1,04	3,45	34	0,77	0,96	—
29. Benzo (a) pentacene	1,02	3,21	26	0,49	0,86	—
30. Tétrabenzo (a, c, j, l) naphtacène	1,39	3,48	34	0,85	1,08	—
31. Dibenzo (a, c,) naphtacene	1,20	3,30	26	0,71	1,08	—
32. Tétrabenzo (a, c, f, h) naphtalène	1,35	4,42	26	1,02	1,22	—
33. Dibenzo (a, j) perylène	1,30	3,65	28	0,71	1,05	—
34. Dibenzo (a, m) pérylène	1,30	3,52	28	0,71	1,07	—
35. Tétrabenzo (a, c, hi, qr) pentacène	1,38	3,65	36	1,06	1,11	—
36. Benzo (e) pyrène	1,08	4,06	20	0,99	1,22	—
37. Picène	1,11	4,15	22	1,00	1,18	—
38. Benzo (c) pentaphène	1,03	3,44	26	0,86	0,98	—
39. Pentaphène	1,01	3,45	22	0,87	0,96	—
40. Naphto (2, 3-a) naphtacène	1,00	3,29	26	0,67	0,82	—
41. Dibenzo (c, g) phénanthrène	1,09	4,18	22	1,07	1,19	—
42. Tribenz (a, c, h) anthracène	1,06	3,58	26	1,05	1,16	—
43. Tribenzo (a, c, j) naphtacène	1,03	3,40	30	0,78	1,05	—
44. Benzocoronene	1,16	3,37	28	0,94	1,05	—
45. Benzo (b) chrysène	1,09	3,39	22	0,81	1,11	—
46. Dibenzo (b, g) phénanthrène	1,07	3,39	22	0,84	1,08	—
47. Dibenzo (a, l) naphtacène	1,03	3,37	26	0,72	1,16	—
48. Dibenzo (a, j) naphtacène	1,03	3,37	26	0,71	1,04	—
49. Pyranthène	1,03	3,43	30	0,52	0,93	—
50. Dinaphtho (2, 3, b, h) pyrène	1,01	3,39	32	0,55	0,75	—
51. Benzène	1,53	4,00	6	2,00	2,00	—
52. Pyrène	1,06	3,92	16	0,89	1,32	—
53. Pérylène	1,28	3,91	20	0,69	1,35	—
54. Anthanthrène	1,03	3,46	22	0,58	1,04	—
55. Anthraceno (1, 2-a) anthracène	1,10	3,38	26	0,72	1,15	—
56. Phénanthra (3, 2-b) pyrène	1,03	4,27	28	0,65	0,92	—
57. Coronène	1,15	4,22	24	1,08	1,08	—
58. Naphthodianthrène	1,28	3,28	28	0,35	0,88	—
59. Peropyrène	1,10	3,93	26	0,57	1,11	+
60. Benzo (e) naphto (2, 3-1) pyrène	1,14	3,37	28	0,68	1,00	—
61. Benzo (fgh) pérylène	1,09	4,00	22	0,88	1,12	+
62. Ovalène	1,08	3,58	32	0,67	0,94	—

Nota: 1. Toutes les variables sauf *N* sont exprimés en unites β, un parametre de calcul dans la simpe méthode de Hückel.
2 Les indices complexes des régions *K* et *L* ne figurent pas dans ce tableau; on peut les trouver dans la référence (18).

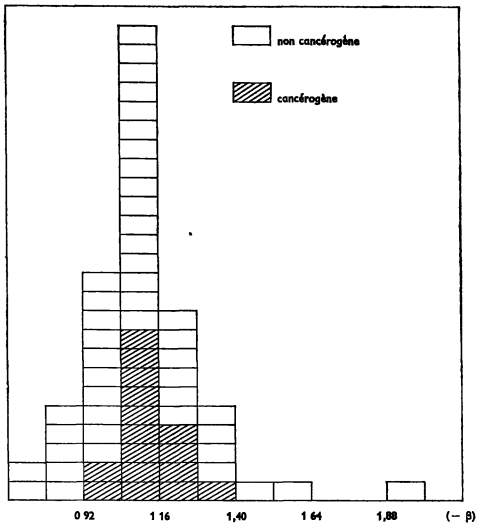

FIG. 1

Histogramme de distribution de ΔE_2

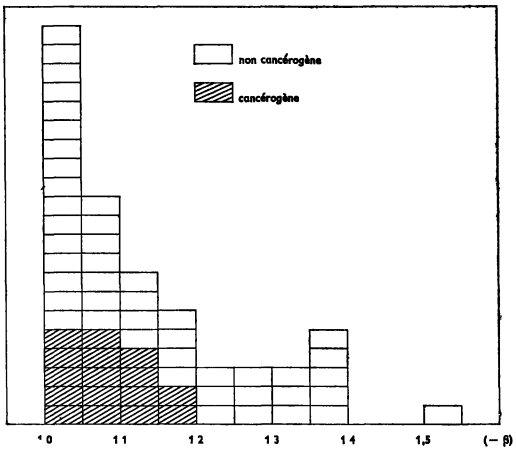

FIG. 2

Histogramme de distribution des indices simples de région K

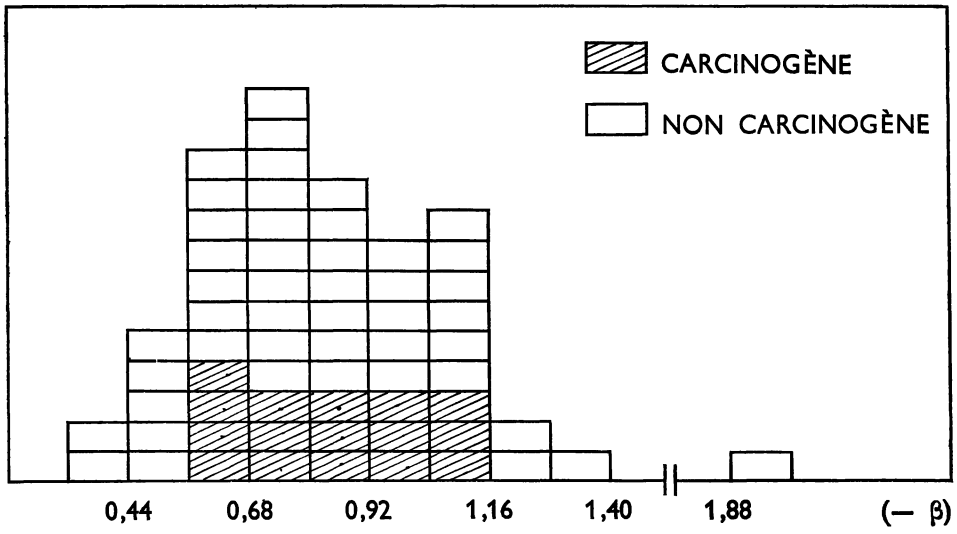

FIG. 3

Histogramme de distribution de ΔE_1

d'énergie entre l'état fondamental de la molécule et son premier et second état excité, calculées par la simple méthode de Hückel. A noter tout de suite que ΔE_1, telles que calculées par la méthode de Hückel en question et dans le cas particulier des hydrocarbures que nous étudions, représentent aussi, à un facteur constant près, les énergies de potentiel d'ionisation. Comme le montrent les figures 1 et 3, les relations entre le pouvoir carcinogène et ΔE_1 et ΔE_2 ne sont pas monotones, nous avons dû procéder, pour notre analyse quantitative, à un changement de variables en posant :

$$E_1 = | \Delta E_1 - \overline{\Delta E_1^c} | \qquad \text{et} \qquad E_2 = | \Delta E_2 - \overline{\Delta E_2^c} |$$

où ΔE_1^c et ΔE_2^c représentent respectivement la moyenne des ΔE_1 et ΔE_2 du groupe des carcinogènes. Physiquement, cela signifie qu'il y a une valeur propice d'énergie d'excitation et de potentiel d'ionisation pour le pouvoir carcinogène et que celui-ci s'atténue à mesure que l'on s'éloigne de ce point.

Si l'on procède exactement comme l'a fait Herndon, en utilisant ses données concernant P_K, P_L et Δ, mais simplement en adjoignant deux autres « facteurs » concernant ΔE_1 et ΔE_2, on obtient, en se limitant aux données des 10 premiers cas du tableau I, la formule de régression suivante :

Indice de carcinogénicité $= -48{,}7 + 267{,}3\, P_K - 158{,}4\, P_L - 4{,}9\, \Delta - 42{,}2\, E_2 - 39{,}1\, E_1$ (II)

L'équation (II) a un coefficient de corrélation multiple R (II) $= 0{,}948$. La progression est peu importante, mais non négligeable.

Dans l'espoir de pouvoir lever cette ambiguïté, nous avons voulu procéder d'une autre façon en essayant de tenir compte de toutes les données que nous avons à notre disposition, c'est-à-dire que les 62 données doivent apporter toutes leur contribution propre. Pour cela, nous avons dû renoncer à nous servir des indices chiffrés de carcinognénicité, car pour 6 ou 7 carcinogènes, l'indice d'Iball n'est pas connu dans la littérature. Nous avons donc assigné la valeur 1 aux carcinogènes et 0 aux non carcinogènes. Pour les indices de réactivité des régions K et L, nous continuons à utiliser les indices simples des régions K et L, c'est-à-dire respectivement les énergies d'ortholocalisation et de paralocalisation. Car, si Herndon avait trouvé que les indices complexes des régions K et L étaient moins aptes que les siens, il n'avait pas essayé les indices simples correspondants que nous trouvons meilleurs que les complexes [18].

Le tableau II indique les coefficients de corrélation simples deux à deux entre le pouvoir carcinogène d'un hydrocarbure et sept de ses indices électroniques caractéristiques :

TABLEAU II

1 : pouvoir carcinogène (0 ou 1); 2 : indice simple de région K; 3 : indice complexe de région K; 4 : indice simple de région L; 5 : indice complexe de région L; 6 : $\Delta = | N - 24 |$ défini plus haut; 7 : $E_2 = | \Delta E_2 - \Delta E_2^c |$ défini plus haut; et 8 : $E_1 = | \Delta E_1 - \Delta E_1^c |$ défini plus haut.

Numeros de variables	1	2	3	4	5	6	7	8
1	1,000	—0,258	—0,206	+0,258	+0,173	—0,288	—0,285	—0,167
2		1,000	+0 924	+0,300	+0,360	+0,421	+0,263	+0,430
3			1,000	+0,428	+0,520	+0,480	+0,282	+0,417
4				1,000	+0,941	—0,038	—0,007	+0,147
5					1,000	+0,061	+0,082	+0,219
6						1,000	+0,643	+0,442
7							1,000	+0,775
8								1,000

Nota : On rappelle que les seuils de signification pour 60 degrés de liberté sont les suivants : risque 10 % : 0,211; risque 5 % : 0,250; risque 2 % : 0,295 et risque 1 % : 0,325.

L'examen du tableau II permet de faire les quelques remarques suivantes : 1° seuls les indices simples des régions K et L, la variable de dimension moléculaire Δ et la variable d'énergie d'excitation E_2 sont corrélés avec le pouvoir carcinogène à un degré de signification de l'ordre de 0,05; 2° les indices simples des régions K et L sont corrélés avec leurs homologues complexes à un très haut degré de signification, puisque leur coefficient de corrélation valent respectivement 0,924 et 0,941; 3° la variable de dimension moléculaire Δ a le plus fort coefficient de corrélation avec le pouvoir carcinogène, mais elle est en même temps très fortement corrélée avec les indices simples de région K, et avec E_2 et E_1.

A la suite des remarques ci-dessus, il était à prévoir que si l'on établit une formule de régression multiple du pouvoir carcinogène en fonction des sept autres variables que nous nous proposons d'analyser, la contribution des indices complexes des régions K et L sera négligeable. Pour établir ladite formule, nous avons fait appel à un programme sur ordinateur [1] qui fait entrer automatiquement l'une après l'autre les variables suivant l'importance de leur coefficient partiel de corrélation et en donnant à chaque étape les coefficients de régression de chaque variable et le coefficient de corrélation multiple de l'ensemble. Le tableau III résume les résultats des diverses étapes de cette opération.

TABLEAU III

K : indice simple de région K; *L :* indice simple de région L; Cte *:* Constante; *R :* coefficient de corrélation multiple. Δ, E_2 et E_1 ont été définis plus haut.

Étape	Coefficients de régression des variables dans la formule de régression multiple						Coefficient de corrélation multiple R
	Cte	K	L	Δ	E_2	E_1	
1re	—0,635	—	0,238	—0,034	—	—	0,380
	0,181	—1,292	0,421	—	—	—	0,436
2e	0,087	—1,041	0,390	—0,018	—	—	0,456
3e	0,098	—1,053	0,395	—0,004	—0,589	—	0,477
4e	0,271	—1,208	0,384	—0,0004	—0,979	—0,408	0,487
5e	0,274	—1,212	0,385	—	—0,986	—0,410	0,487

La formule de régression multiple finale est donc :

$$\text{Indice de carcinogénicité} = 0{,}274 - 1{,}212\ K + 0{,}385\ L - 0{,}986\ E_2 - 0{,}410\ E_1 \quad \text{(III)}$$

avec un coefficient de corrélation multiple R (III) = 0,487. On remarque tout d'abord que les indices complexes des régions K et L n'y figurent pas, leur contribution au coefficient de corrélation multiple étant inférieur à 0,01. Un fait notable et digne d'un très grand intérêt est le suivant : bien qu'au départ, la variable Δ ait parmi les six autres le plus fort coefficient de corrélation avec le pouvoir carcinogène, sa contribution propre au coefficient de corrélation de l'ensemble ne cesse de diminuer à chaque étape pour disparaître complètement dans la formule finale. Ceci voudrait dire que le rôle de Δ figurant dans la formule (I) de Herndon n'est qu'apparent et pourrait être remplacé totalement par deux autres variables telles que E_2 et E_1. La contribution des énergies d'excitation et du potentiel d'ionisation est ainsi plus

1. Nous remercions très sincèrement M. Dimbour, conseiller mathématique au Centre interégional de Calcul électronique (C. I. R. C. E.) où nous avons effectué tous nos calculs, de son assistance technique extrêmement dévouée et efficace.

clairement mise en évidence par la comparaison entre le coefficient de corrélation de la formule (III), R (K, L, E_2, E_1) = 0,487 et celui de l'équation où ne sont entrées que les variables K et L, R (K, L) = 0,436 faisant apparaître une progression de 0,051.

Discussion

Nous venons donc d'obtenir, par une méthode d'analyse multidimensionnelle portant sur 62 molécules, une formule de régression où le pouvoir carcinogène d'un hydrocarbure aromatique polycyclique est exprimé en fonction de 4 variables distinctes : les indices de réactivité chimique des régions K et L, l'énergie d'excitation électronique ΔE_2 et le potentiel d'ionisation ΔE_1 (1), chacun de ces 4 « facteurs » ayant sa contribution propre. Regardons maintenant de plus près la formule III. On note tout d'abord que les signes de ces 4 facteurs sont bien en accord avec les faits, puisque la théorie des régions K et L prévoit que la réactivité par la région K (2) est favorable à l'apparition du pouvoir carcinogène alors que celle par la région L lui est défavorable, d'où le sens opposé des signes du facteur K et du facteur L. De même que les signes de E_2 et E_1 doivent être négatifs, puisque (rappelons que $E_2 = |\Delta E_2 - \Delta E_2^c|$ et $E_1 = |\Delta E_1 - \Delta E_2^c|$) les carcinogènes ont leur valeur de ΔE_2 et ΔE_1 centrées sur une valeur déterminée propice, toute autre valeur s'éloignant de ce point ayant pour effet de diminuer le pouvoir carcinogène. Quant à l'importance respective de ces 4 variables, compte tenu de l'unité de chacune d'elles, elle conforme aussi à notre attente. Ainsi, comme l'a déjà signalé Herndon [6], la contribution de la réactivité par la région K est plus importante que celle par la région L. De même, le coefficient du facteur E_2 est plus fort que celui du facteur E_1. Par ailleurs, l'importance du facteur K paraît être comparable à celle du facteur E_2 et l'importance du facteur L à celle du facteur E_1, avec une légère prépondérance des indices de réactivité. Dans l'ensemble, l'esprit de la théorie des régions K et L a donc été confirmé. Mais les réactivités par les régions K et L ne pourront plus être considérées comme étant les seules en cause. Cette conception rigide a d'abord été battue en brèche par Herndon [6] en introduisant dans sa formule un troisième facteur représentant la dimension moléculaire. La présente analyse apporte quelques précisions complémentaires en montrant que le rôle de la dimension moléculaire n'est en réalité qu'apparent et pourrait être remplacé par celui de deux autres facteurs.

Quant à la signification physique de ce phénomène que nous venons de mettre en avant, son évocation entre évidemment dans le domaine de la spéculation. Car, comme le savent bien les statisticiens, on ne peut jamais être assez prudent en passant du stade de la constatation statistique à celui de l'interprétation causale. Nous avons bien vu plus haut que l'effet apparent de la dimension moléculaire devait s'effacer devant l'introduction des facteurs E_2 et E_1, il n'est donc pas exclu que dans l'avenir ceux-ci ne subissent le même sort devant l'introduction d'autres facteurs à découvrir. En admettant que ΔE_2 et ΔE_1 ont bien réellement un rôle propre, nous pourrons risquer une interprétation qui ne deviendra convaincante qu'après corroboration par d'autres faits théoriques et/ou expérimentaux indépendants. On sait que l'action de la réactivité par les régions K et L dans le mécanisme du processus

1. On n'oublie pas que ΔE_1 représente dans notre cas particulier aussi bien un potentiel d'ionisation qu'une énergie d'excitation électronique. La présente analyse ne permet pas de faire la distinction. C'est pour des raisons de simplicité que nous disons tout le long de cet article qu'il s'agit d'un potentiel d'ionisation.

2. Pour les indices des régions K et L que nous avons introduits dans la formule III, les petites valeurs traduisent des réactivités plus fortes; pour les indices utilisés par Herndon dans la formule I, c'est l'inverse.

de cancérisation par les hydrocarbures aromatiques n'est pas seulement suggérée par l'analyse statistique, mais elle est également appuyée par de nombreux fait expérimentaux [2]. Il est vrai que ces dernières années, on a signalé quelques faits expérimentaux isolés qui sont de nature à jeter la confusion à ce consensus quasi général. Par exemple, Cavalieri et Calvin [19] en irradiant à 3 500 Å le benzo(a)pyrène en présence de l'hydrochlorure de 1-méthylcytosine, ont obtenu un 6-(1-méthylcystos-5-yl)benzo(a)pyrène. La fixation a donc été faite non sur la région K, mais sur un atome de carbone. Pour répondre à cela, Umans et Pullman [20] ont fait une analyse sur un nombre très restreint, il est vrai, de données qui sont disponibles à cet effet, et ont trouvé que si la réaction *in vitro* des hydrocarbures avec les ADN est compatible avec un mécanisme de fixation du réactif par un atome de carbone, celle *in vivo* est bien mieux expliquée par un mécanisme de fixation par la région K. Il est vrai aussi que récemment, la carcinogénicité des époxydes dérivés des hydrocarbures aromatiques soit de la région K soit d'une liaison un peu moins réactive, fait l'objet d'investigation intense et certaines expériences ne semblent pas confirmer le rôle de la région K. Mais il faut reconnaître que toutes les expériences n'amènent pas encore des conclusions concordantes. Pour résumer en peu de mots, on peut dire que *in vitro*, les époxydes se sont révélés par rapport aux hydrocarbures dont ils dérivent plus mutagènes envers certains micro-organismes [30] et plus transformants envers les fibroblastes en culture [31, 32]. Par contre, *in vivo*, ces époxydes ont été trouvés en général moins carcinogènes que leur molécules-parentes [34]. De plus, la formation *in vitro* des époxydes se fait souvent en dehors de la région K [33]. L'explication de ce paradoxe semble résider, comme l'a souligné l'équipe de Chouroulinkov [34], dans le fait que ces époxydes sont des corps extrêmement instables et que les expériences *in vitro* doivent être gênées moins que celles *in vivo* par ce caractère. De toute façon, un argument statistique est un argument de constatation et par conséquent indépendant de toute autre considération. Et dans la recherche d'une interprétation cohérente de tous ces faits expérimentaux apparemment contradictoires, ce caractère des arguments statistiques est d'une importance non négligeable.

Jusqu'à une date récente, on croyait, d'après les données expérimentales dont on disposait alors, que les hydrocarbures agissaient *per se* sans modifications métaboliques préalables. Puis, petit à petit [3], on est revenu de cette conviction. Cependant, deux faits essentiels sont à ne pas perdre de vue dans notre raisonnement : 1° les hydrocarbures aromatiques sont des composés réputés chimiquement extrêmement stables, et par conséquent peu enclins à s'engager dans des réactions chimiques sans intervention d'une force extérieure; 2° ils agissent biologiquement *in situ* et à la dose de l'ordre du milligramme à la différence des azoïques par exemple qui agissent à distance et avec des doses environ 100 fois plus fortes. Pour tenir compte de tous ces faits à la fois, on pourrait penser aux deux possibilités suivantes : 1° l'organisme facilite la fixation des substances cancérogènes sur des récepteurs cellulaires par un ou des systèmes enzymatiques; cette hypothèse fait effectivement l'objet d'investigation de nombreux auteurs (1); 2° les hydrocarbures peuvent ne pas agir dans leur état fondamental, mais dans un des états électroniquement excités ou sous une forme ionique [22], les deux possibilités n'étant du reste pas forcément exclusives. Cette présente analyse apporte quelques indications en faveur du second mécanisme sans pour autant infirmer le premier. En effet, les deux entités calculées, ΔE_1 et ΔE_2, expriment les énergies de deux bandes d'absorption des hydrocarbures aromatiques. Selon la classification

1. On sait aujourd'hui que la formation des époxides est effectivement catalysée par des monooxygénases NADPH — dépendantes qui se trouvent dans la fraction microsomique cellulaire [21].

CENTRE DE RECHERCHES ET D'ÉTUDES ... ET STATISTIQUES

de Clar [23], les bandes de ces composés dans le visible et le proche U. V. sont dénommées p, α et β. ΔE_1 correspond à l'énergie de la bande p, mais ΔE_2, telle calculée par la simple méthode de Hückel, correspond, selon la conception théorique de Dewar et Longuet-Higgins [24], aux deux bandes α et β à la fois (1), la bande β étant en général une centaine de fois plus intense que la bande α. Quant à l'intensité de la bande p, elle se situe entre celle des bandes α et β. Ceci étant précisé, rappelons encore que les indices simples des régions K et L représentent une partie essentielle de la barrière de potentiel dans la réaction de fixation respective par les régions K et L des hydrocarbures sur les récepteurs cellulaires. Les valeurs numériques de ces barrières de potentiel, exprimées dans le tableau I en unités β (un paramètre de calcul dans la méthode de Hückel) équivalent dans les 70 kcal/mol, valeurs nettement plus élevées que celles rencontrées dans les réactions chimiques habituelles. Dans ces conditions, on conçoit aisément que *in vitro* on n'a pu obtenir une fixation [25] de benzo(a)-pyrène sur des constituants d'ADN qu'à l'aide d'une irradiation à 3 600 Å. *In vivo*, on peut concevoir que la fixation puisse avoir lieu par les deux mécanismes mentionnés plus haut : à savoir enzymatique ou/et photochimique. Nous ne nous étendons pas dans le présent travail sur le premier mécanisme qui fera l'objet d'une prochaine analyse. Contentons-nous dans celle-ci d'envisager la possibilité que la fixation des hydrocarbures par la région K sur un récepteur cellulaire soit de nature photochimique : c'est-à-dire que la réaction de fixation serait facilitée par un apport d'énergie lumineuse d'une certaine valeur équivalent à la barrière de potentiel. Mais ceci amènerait à introduire en biologie une notion tout à fait nouvelle : existe-t-il à l'intérieur d'un organisme supérieur des sources lumineuses? A la suite d'une découverte faite récemment par White, Wiecko et Roswell [26] que certaines réactions chimiluminescentes peuvent se comporter comme des sources d'énergie d'excitation électronique transférable, Sung et Buu-Hoï [28] ont procédé à une recherche bibliographique pour savoir s'il y a des réactions biochimiques équivalentes à celles signalées par les auteurs américains. On a trouvé que de telles réactions ne manquent pas dans le processus métabolique normal. Par exemple, dans le processus de dégradation du tryptophane, il y a un intermédiaire qui semble bien se trouver dans un état excité. Il en est de même dans le domaine des lipides et des terpènes, l'oxydation biologique passe souvent par la formation transitoire d'oxy-2-oxétane, substance que White et collaborateurs eux-mêmes ont utilisée par thermolyse pour engendrer deux espèces en état d'excitation électronique. Cette nouvelle notion importante sur la possibilité de transfert d'énergie d'excitation électronique dans le processus physiologique normal fera probablement dans les prochaines années l'objet d'attention d'autres chercheurs.

Au moment où nous mettions en chantier ce présent article, nous apprenons que Popp [29], en analysant 37 hydrocarbures de notre groupe, a trouvé une très bonne relation entre le pouvoir carcinogène de ces composés et certaine disposition de leurs bandes d'absorption p et α. Il a remarqué, en effet, que les carcinogènes ont presque toujours leur bande p distante de leur bande α d'un intervalle énergétique déterminé (de l'ordre de 0,15 e. v.). Cette remarque est tout à fait en harmonie avec notre présente analyse et nous allons regarder plus en détail dans un prochain travail comment se situe la découverte de Popp par rapport à nos deux facteurs ΔE_1 et ΔE_2.

Mais ce que nous voulons surtout souligner, c'est que pendant longtemps, on était à la recherche d'une corrélation simple entre le pouvoir carcinogène et une seule propriété

1. Pour un effectif de 40 molécules (27), les coefficients de corrélation sont les suivants : $\Delta E_2 (\alpha) = 0{,}90$; $\Delta E_2 (\beta) = 0{,}93$ et coefficient de corrélation multiple $\Delta E_2 (\alpha, \beta) = 0{,}95$.

moléculaire. Chaque auteur faisait valoir la corrélation qu'il a lui-même trouvée et rejetait toutes les autres comme si le processus de cancérisation par les substances chimiques ne comportait qu'un seul facteur. On entre désormais dans une nouvelle phase où on mettra à profit les techniques avancées de la science statistique, telle que l'analyse multi-dimensionnelle, pour situer tous les « facteurs envisagés ces dernières années les uns par rapport aux autres : certains seront reconnus comme étant des facteurs indépendants, d'autres ne feront que double emploi. Et lorsque l'énorme puzzle sera enfin totalement déchiffré, on aura une équation de régression multiple comportant probablement au moins une vingtaine de facteurs.

Shou-Sin Sung,
maitre de recheche au C. N. R. S. Institut de cancérologie et d'immunogénétique, département de cancérogenèse chimique, 94800 Villejuif

et Paul Damiani,
administrateur de l'I. N. S. E. E.

BIBLIOGRAPHIE

[1] Sung S. S. — *J. Soc. Statis.* Paris, n° 1, 1975, p. 44.

[2] Pullman A. et B. — *Cancérisation par les substances chimiques et structure moléculaire.* Masson, Paris, 1955.

[3] Pullman B. et A. — *The Jerusalem Symposia on Quantum Chemistry and Biochemistry*, I, Physico-Chemical Mechanisms of Carcinogenesis, Academic Press, New York, London, 1969.

[4] Pullman B. — *Theoretical Physics and Biology.* M. Marois, ed., North-Holland Publishing C°, Amsterdam, 1969.

[5] Daudel P. et R. — *Chemical Carcinogenesis and Molecular Biology.* Interscience Publishers, New York, London, Sidney, 1966.

[6] Herndon W. C. — *Transact.* N. Y. Acad. Sci., sér. II, vol. 36, 1974, p. 200.

[7] Herndon W. C. — *Intern. Quant. Chem.*, n° 1, 1974, p. 109.

[8] Boyd G. V. et Singer N. — *Tetrahedron*, 22, 1966, p. 3383.

[9] Meyer A. T. et Bergmann E. D. — *Jerusalem Symposia on Quantum Chemistry and Biochemistry*, I, Physico-Chemical Mechanisms of Carcinogenesis, Academic Press, New York, London, 1969.

[10] Hoffman F. — *Theor. Chim. Acta*, 15, 1969, p. 393.

[11] Iball J. — *Amer. J. Cancer*, 35, 1939, p. 188.

[12] Pauling L., Brockway L. O. et Beach J. T. — *J. Amer. Chem. Soc.*, 57, 1935, p. 2705.

[13] Sung S. S. — *Ann. Soc. Sci.*, Brux., 89, II, 1975, p. 319.

[14] Mason R. — *Nature*, 181, 1958, p. 820.

[15] Mason R. — *Brit. J. Cancer*, 12, 1958, p. 469.

[16] Pullman A. et B. — *Nature*, 196, 1962, p. 228.

[17] Pullman A. — *Biopolymers*, in *Symposia*, n° 1, 1964, p. 47.

[18] Sung S. S. — C. R. Acad. Sci., 274 D, 1972, p. 1597.

[19] Cavalieri E. et Calvin M. — *Photochem. Photobiol.*, 14, 1971, p. 614.

[20] UMANS R. S. et PULLMAN B. — *Intern. J. Quant. Chem.*, 5, 1971, p. 575.

[21] Par exemple : P. SIMS, P. L. GROVER et al., Nature, 252, 1974, p. 326.

[22] DIPPLE A., LAWLEY P. D. et BROOKS P. — *Europ. J. Cancer*, 4, 1968, p. 493.

[23] CLAR E. — *Polycyclic Hydrocarbons*. Academic Press, Londres, New York et Springer-Verlag, Berlin, Gottingen et Heidelberg, 1964.

[24] DEWAR M. J. S. et LONGUET-HIGGINS H. C. — *Proc. Phys. Soc.*, A 67, 1954, p. 795.

[25] Ts'o P. O. P. et LU P. — *Proc. Nat. Acad. Sci.*, 51, 1964, p. 272.

[26] WHITE E. H., WIECKO J. et ROSWELL D. — *J. Amer. Chem. Soc.*, 91, 1969, p. 5194.

[27] SUNG S. S. et LAZAR P. — *J. Chim. Phys.*, 66, 1969, p. 1372.

[28] SUNG S. S. et BUU-HOI N. P — C. R. Acad. Sci., 270 D, 1970, p. 2052.

[29] POPP F. A. — Communication personnelle, sous presse.

[30] AMES B. N , SIMS P. et GROVER P. L. — *Science*, 176, 1972, p. 47.

[31] GROVER P. L., SIMS P., HUBERMAN E , MARQUARDT H., KUROKI T. et HEIDELBERGER C. — *Proc. Nat. Acad. Sci.*, U. S. A., 68, 1971, p. 1098.

[32] MARQUARDT H., KUROKI T., HUBERMAN E., SELKINK J. K., HEIDELBERGER C., GROVER P. L. et SIMS P. — *Cancer Res.*, 32, 1972, p. 716.

[33] DAUDEL P., DUQUESNE M., VIGNY P., GROVER P. L. et SIMS P. — Febs Letters, 57, 1975, p. 250.

[34] CHOUROULINKOV I., GENTIL A. et LASNE C. — C. R. Acad. Sci., 281 D, 1975, p. 207.