

SHOU-SIN SUNG

**Un exemple d'application de la méthode statistique  
en biologie moléculaire**

*Journal de la société statistique de Paris*, tome 116 (1975), p. 44-50

[http://www.numdam.org/item?id=JSFS\\_1975\\_\\_116\\_\\_44\\_0](http://www.numdam.org/item?id=JSFS_1975__116__44_0)

© Société de statistique de Paris, 1975, 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/>

## II

### ARTICLES

#### UN EXEMPLE D'APPLICATION DE LA MÉTHODE STATISTIQUE EN BIOLOGIE MOLÉCULAIRE

##### Qu'est devenue la théorie des régions K et L en matière de carcinogénèse chimique?

*The author applies statistical analysis methods to see whether K-L region theory applied to chemical substances makes possible to distinguish carcinogen from the other substances.*

*Der Verfasser verwendet die Methode der statistischen Analyse um zu sehen, ob die Theorie der Regionen K und L auf chemische Substanzen angewendet es ermöglicht die krebserregenden Substanzen von den anderen Substanzen zu unterscheiden*

Un des buts essentiels de l'étude de la carcinogénèse par les substances chimiques est de faire des approches en vue d'élucider à long terme les mécanismes intimes d'action de ce processus pathologique au niveau moléculaire. Les connaissances de base acquises dans ce domaine jusqu'à ce jour sont de nature à nous inciter d'envisager une telle entreprise. En effet, environ 4 000 composés chimiques ont été successivement essayés du point de vue de leur pouvoir carcinogène par les cancérologues et biochimistes du monde entier. Un inventaire a été dressé par le service de santé publique américain (voir « Survey of compounds which have been tested for their carcinogenic activity », 1951, 1329 composés) et a été régulièrement mis à jour (Supplément I, 1957, 779 composés; Supplément II, 1969, 1 048 composés et Supplément III, 1972, environ 1 000 composés). On y découvre quelque 20 % de substances carcinogènes. Les plus puissantes d'entre elles sont capables de provoquer chez les animaux expérimentés quasi 100 % de tumeurs à la dose d'une fraction de milligramme! Certaines de ces substances font partie du goudron de tabac et des poussières urbaines! Curieusement, ces composés pathogènes sont réparties dans presque toutes les familles de la chimie organique. Aucune tentative d'attribuer l'action carcinogène à des

groupements fonctionnels chimiques particuliers n'a été couronnée de plein succès. Prenons par exemple un groupe de 15 hydrocarbures aromatiques constitués par l'accolement de 5 cycles benzéniques sans aucun groupement fonctionnel particulier. Ils ont le même nombre de carbones et certains d'entre eux sont exactement isomères. La seule différence distinctive est la disposition relative des cycles les uns par rapport aux autres. Parmi ces 15 composés extrêmement voisins, 10 ne sont pas carcinogènes et 5 le sont dont le fameux 3,4-benzopyrène très puissant. Devant cet état de choses, on est en droit de se demander si les mécanismes intimes de cancérisation ne devraient pas se chercher au niveau submoléculaire (répartition des charges sur la molécule par exemple). Malheureusement, les phénomènes physiopathologiques évoluant à une telle échelle sont difficilement accessibles aux expériences classiques.

Lorsque l'on essaie d'établir statistiquement une corrélation entre le pouvoir carcinogène d'une molécule et certaines de ses propriétés moléculaires ou submoléculaires plus élémentaires, on s'aperçoit que tandis que les données concernant le premier sont abondantes, celle concernant ces dernières sont nettement insuffisantes. Cette étude est heureusement devenue possible grâce à l'avènement de la mécanique ondulatoire et des ordinateurs puissants qui permettent d'estimer par des calculs théoriques, dans des cas particuliers, certaines propriétés moléculaires ou électroniques dont on soupçonne quelque chose à voir avec le pouvoir carcinogène. Mais nous sommes naturellement conscients des inconvénients de recourir à ces méthodes d'approche indirectes qui ne pourraient se justifier qu'en l'absence des données expérimentales correspondantes.

Un sujet qui a eu quelque succès retentissant il y a une quinzaine d'années dans ce domaine est la théorie des régions K et L. Cette théorie a été bâtie, à partir des éléments calculés mathématiquement par une méthode relevant de la mécanique ondulatoire. Il n'y a pas lieu d'expliquer en détail dans cet article en quoi consiste la méthode de calcul utilisée. Pour la discussion qui suit, il suffit de savoir que cette méthode de calcul ne donne que des mesures de certaines propriétés moléculaires exactement comme peuvent le faire les expériences. Rien de plus. Dans notre cas précis, on va discuter certains indices de réactivité chimique calculés d'une molécule en rapport avec son pouvoir carcinogène qui, lui, est déterminé expérimentalement. Certes, ces indices peuvent aussi être déterminés par les expériences : les indices de région K d'une molécule peuvent être mesurés par sa réactivité avec le tétra-oxyde d'osmium et ceux de région L par celle avec l'anhydride maléique. Ces données expérimentales ont bien servi de contrôle pour nos indices calculés, mais leur nombre est nettement insuffisant pour alimenter une analyse statistique valable. Ainsi se justifie le recours aux entités calculées.

Avant d'entrer dans les détails, nous pouvons annoncer que cette théorie construite il y a une vingtaine d'années en examinant un certain nombre de données, de l'ordre de 37, n'a pu être vérifiée lorsqu'on veut l'appliquer ultérieurement à d'autres lots de molécules. La question de son abandon pourrait ainsi être logiquement posée. Nous verrons au cours de cet exposé, que les raisons de cette impasse sont extrêmement simples. En effet, l'application des méthodes statistiques a permis facilement une remise au point en faisant ressortir ce qu'il y a de positif dans cette théorie.

Rappelons maintenant ce qu'est cette théorie des régions K et L. Tout d'abord, quelles sont ces substances qui font l'objet de notre étude? Dans la figure 1, je n'ai pas fait figurer les premières molécules à l'aide desquelles a été bâtie la théorie que vous trouverez dans la référence [1]. Pour donner une idée sur la nature des substances traitées, je n'ai fait figurer que 23 hydrocarbures aromatiques supplémentaires qui forment un groupe homogène avec les premières. Ces molécules sont toutes constituées par l'accolement de cycles benzéniques

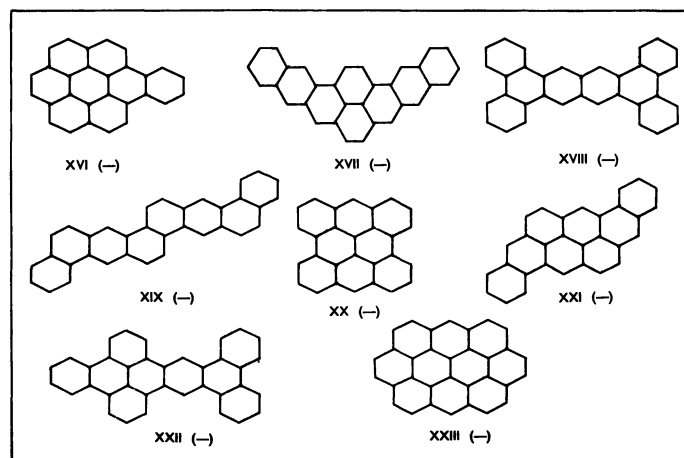
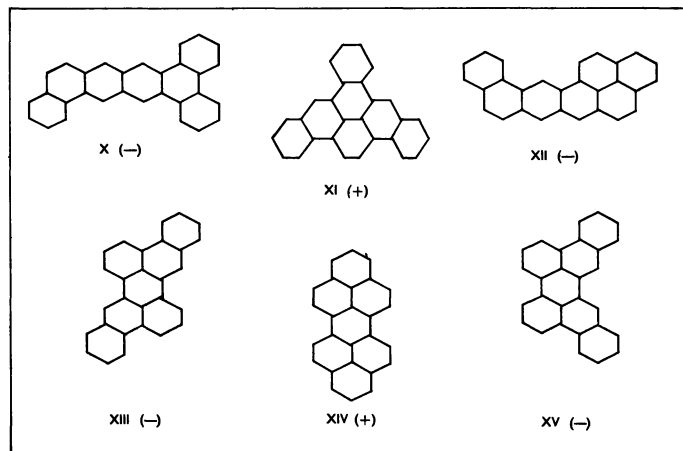
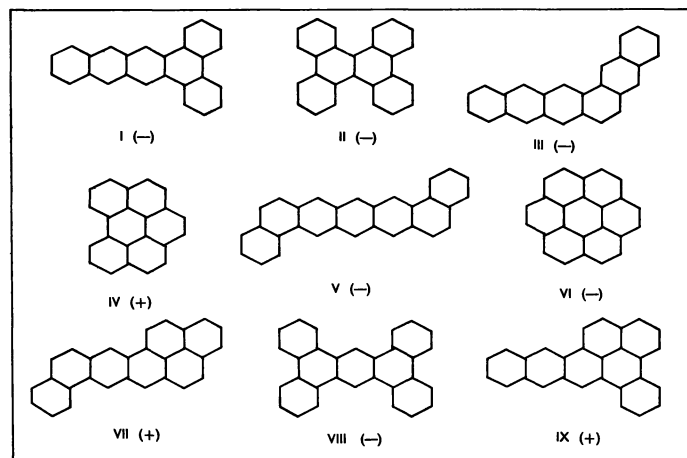
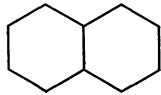
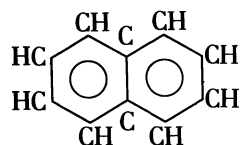


Fig. 1

sans aucun groupement fonctionnel particulier. Je dois préciser un peu sur la convention d'écriture utilisée, car elle est très simplifiée. Prenons l'exemple d'un des plus simples,

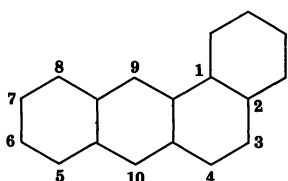
le naphthalène qui ne figure pas dans la liste présente.  On y a omis d'écrire tous

les C et H, et la structure est supposée entièrement conjuguée. Au complet, on devrait écrire

 . Nous étudions ce groupe de molécules en cherchant si certaines propriétés

de réactivité chimique ne conditionnent pas chez la molécule qui les possède l'apparition de son pouvoir carcinogène. Car, je le répète, cette famille d'hydrocarbures contient une grosse proportion de carcinogènes.

Venons-en maintenant aux détails. Tout d'abord, qu'appelle-t-on régions K et L? Primitivement, on appelait chez un hydrocarbure aromatique la région K la liaison mésophénanthrénique et la région L les deux sommets mésanthracéniques. Prenons l'exemple

classique, le 1,2-benzanthracène :  Chez cette molécule, la liaison 3,4

constitue la région K et les sommets 9 et 10 la région L. Il est évident qu'avec cette définition, beaucoup de molécules manquent soit l'une soit l'autre de ces régions, soit même toutes les deux. Et pourtant, ces molécules peuvent très bien être carcinogènes. Pour pouvoir être englobés dans une même théorie tous les hydrocarbures, plus tard, les principaux promoteurs de cette théorie, c'est-à-dire professeur et M<sup>me</sup> Pullman [1, 2] d'une part et l'équipe du P<sup>r</sup> Daudel [5] d'autre part ont proposé d'élargir la notion des régions K et L. La région K sera la liaison la plus réactive d'une molécule et la région L sera constituée par la paire de deux sommets les plus réactifs. Ainsi les anciennes régions K et L ne seront plus que des cas particuliers, car quand une molécule contient une région mésophénanthrénique, c'est toujours elle qui est la liaison la plus réactive; et quand la molécule possède deux sommets mésanthracéniques, ce sont toujours eux qui sont les sommets les plus réactifs. Avec cette définition élargie, tous les hydrocarbures polybenzéniques que nous étudions sont pourvus à la fois de ces deux régions.

La théorie des régions K et L selon professeur et M<sup>me</sup> Pullman [1, 2] dit que pour être cancérigène, un hydrocarbure polybenzénique doit posséder une région K active ayant un indice de réactivité égal ou inférieur à un certain seuil en même temps une région L au contraire peu active ayant un indice égal ou supérieur à un autre seuil. A l'aide de ces deux seules règles, les auteurs de la théorie des régions K et L. ont pu expliquer, selon eux, la carcinogénicité ou la non-carcinogénicité de 37-39 hydrocarbures avec seulement quatre exceptions. On pourrait se permettre de se douter si ce n'est pas trop beau pour être vrai. Car, par habitude, l'on sait qu'un phénomène aussi complexe que le processus de cancérisation ne pourrait ne comporter que deux facteurs seulement. Effectivement ces dernières années, plusieurs faits expérimentaux isolés ont été signalés [5] qui sont de nature à jeter confusion sur la théorie des régions K et L. Je vous en cite trois pour exemples.

1) Un fait expérimental essentiel à l'appui de la thèse des régions K et L a été le fait qu'environ 25 % de 1, 2, 5, 6-dibenzanthracène badigeonné sur la peau de souris ont été retrouvés par Bhargava et Heidelberger [6] sous forme d'un métabolite dont la nature fait penser que la fixation de l'hydrocarbure a été faite, conformément à la théorie, sur la région K de la molécule. Or, en 1969, cette même expérience a été répétée par Diringer et Heidelberger [7] et l'on aperçoit que l'on avait commis lors de la première expérience une erreur d'identification.

2) Cavalieri et Calvin [8] en irradiant le benzo (a) pyrène avec l'hydrochlorure de l-méthylcytosine ont obtenu un 6-(l-méthylcytos-5-yl)-benzo (a) pyrène. La fixation a donc été faite non sur la région K, comme le prévoit la théorie, mais sur un atome de carbone.

3) Le 1, 2, 3, 4-dibenzanthracène a été souvent cité comme exemple pour illustrer la théorie des régions K et L en raison du fait qu'il est dépourvu de région mésophénanthrénique (région K au sens strict), mais possède une région L mésanthracénique typique. Il doit donc être, selon la théorie, non carcinogène et il a été connu jusqu'ici comme tel. Or, en 1970, Lijinski et coll. [9] ont trouvé 9 souris sur 30 porteuses de tumeurs après avoir été badigeonnées avec ce produit. Cette expérience a été confirmée presque simultanément par l'équipe de Van Duuren [10].

Ces faits isolés, pour gênants qu'ils soient, n'apportent cependant pas d'arguments décisifs pour rayer d'un trait de plume la théorie des régions K et L. La confusion n'est due au fait que promoteurs et critiques sont tous deux peu familiers au concept de la statistique. Pour éclaircir la situation et par conséquent pour mettre la théorie des régions K et L à la place qu'elle mérite, il faudrait dès le départ abandonner tout vain espoir que le processus de cancérisation puisse ne comporter qu'un petit nombre de facteurs qui sont en fait, selon toute vraisemblance, inconnus en nombre et en nature. Notre problème immédiat est d'essayer de reconnaître le rôle d'un facteur particulier parmi beaucoup d'autres, problème inabordable sans méthodes statistiques qui ont été conçues pour des problèmes semblables mais à une échelle beaucoup plus grande.

Nous avons donc calculé de la même façon que nos prédécesseurs les indices de réactivité pour 23 hydrocarbures supplémentaires qui n'ont pas été analysés auparavant et nous avons fait une analyse statistique globale sur un total de 62 molécules. On trouvera la liste complète de ces composés dans la référence [11]. La figure 2 présente la situation des 62

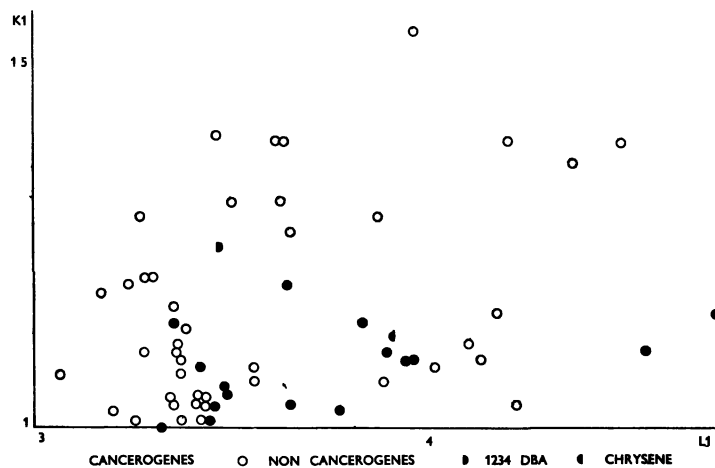


Fig. 2

hydrocarbures en fonction de leurs deux paramètres de réactivité : indices de régions K et L. On y constate :

- 1) que les composés carcinogènes se groupent effectivement volontiers dans le quartier où les indices de région K sont petits et les indices de région L sont grands. Mais il serait malaisé de vouloir délimiter des frontières rigides nettes;
- 2) que le rôle des indices de région K semble plus net que celui des indices de région L;
- 3) qu'enfin la position de quelques composés particuliers n'influence pas beaucoup sur la physionomie de la situation en général comme illustre le cas 1, 2, 3, 4-dibenzanthracène et le chrysène dont la carcinogénicité n'est pas admise unanimement.

Le tableau I exprime toutes ces remarques en langage statistique. On y voit en effet :

TABLEAU I

*Étude statistique des indices de réactivité des régions K et L. (1).*

Effectif	Non carcinogènes 43	Carcinogènes 18	Signification de la différence
Indices de région K {	Moyennes . .	1,15	p < 5 % p' < 1 %
	Variances . .	0,0198	
Indices de région L {	Moyennes . .	3,62	non significative (p ~ 8 %) non significative
	Variances . .	0,1445	

1) que les indices de région K sont en moyenne significativement plus petits pour le groupe de carcinogènes que pour celui des non carcinogènes;

2) que la différence de variances entre ces deux groupes est particulièrement significative, ce qui souligne que les carcinogènes se groupent remarquablement dans une zone particulière sur le spectre de distribution des indices de région K;

3) que les indices de région L ne sont pas significativement différents entre le groupe des carcinogènes et celui des non-carcinogènes ni en moyenne ni surtout en variance.

Cette façon d'analyser notre problème réconcilie tous les faits connus à ce jour concernant la théorie des régions K et L, en particulier ceux signalés plus haut qui lui sont apparemment incompatibles. Ainsi, moyennant quelques restrictions sur sa prétention, la théorie des régions K et L est assise sur une base plus solide. Statistiquement parlant, un hydrocarbure aromatique a beaucoup plus de chance d'être carcinogène s'il possède une région K chimiquement très active. Maintenant, est-ce que cela signifie que la région K active est un facteur direct dans le déclenchement du processus de la cancérisation? Il pourrait être rappelé qu'une analyse statistique positive n'implique jamais une interprétation de causalité. Pour pouvoir se prononcer dans ce sens, il faudrait d'autres faits corroboratifs. Jusqu'à une date récente, on croyait, d'après les faits expérimentaux dont on disposait alors, que les hydrocarbures aromatiques agissaient « per se » sans modifications métaboliques préalables.

1. Dans cette analyse, on n'a pas tenu compte du 1,2,3,4-dibenzanthracène et le chrysène a été considéré comme carcinogène.

Mais petit à petit [3], on est revenu sur cette conviction. On pense maintenant que ces composés ayant en général un potentiel d'ionisation assez faible, agissent peut-être sous forme cathionique ou une autre forme de radicaux. Ils pourraient agir également sous une forme électriquement excitée au lieu dans son état fondamental. Un fait essentiel à ne pas oublier, c'est que ce sont des composés chimiquement très stables et par conséquent peu enclins à s'engager dans une réaction chimique. De plus, ils agissent biologiquement *in situ* et à la dose de l'ordre de milligramme à la différence des azoïques par exemple qui agissent à distance et avec des doses 100 plus fortes. Tout cela fait penser que même s'ils subissent des changements avant d'agir biologiquement, ceux-ci ne seraient que des plus simples. En tout cas, puisque notre analyse statistique n'est basée au départ sur aucune hypothèse interprétative, les conclusions qui en sont dégagées sont essentiellement des constatations objectives. Tout autre mécanisme moléculaire à découvrir dans l'avenir dans le domaine de la carcinogénèse chimique devrait être compatible avec cette déduction mathématique.

SHOU-SIN SUNG

*Maître de recherche au C. N. R. S.*

- 
- [1] PULLMAN A. et B. — Cancérisation par les substances chimiques et structure moléculaire. Masson, Paris, 1955.
- [2] PULLMAN B. — *in* : Theoretical physics and biology, M. Marois, North Holland Publishing Co. Amsterdam, 1969.
- [3] The symposia on quantum chemistry and biochemistry, Academic press. New York, London, 1, 1969.
- [4] DAUDEL P. et R. — Chemical carcinogenesis and molecular biology, Interscience Publishers. New York, London, Sydney, 1966.
- [5] SUNG S.-S. — C. R. Acad. Sc., Série D., 275, 1307, 1972.
- [6] BHARGAVA P.-M. et HEIDELBERGER C. — *J. Am. Chem. Soc.*, 78, 3671, 1956.
- [7] DIRINGER H. et HEIDELBERGER C. — *Cancer Res.*, 29, 2127, 1969.
- [8] CAVALIERE E. et CALVIN M. — *Photochem. Photobiol.*, 14, 614, 1971.
- [9] LIJINSKI W., GARCIA H. et SAFFIOTTI U. — *J. Nat. Cancer Inst.*, 44, 641, 1970.
- [10] VAN DUUREN B.-L., SIVAK A., GOLDSCHMIDT B.-M., KATZ C. et MEOCHIONE S. — *J. Nat. Cancer Inst.*, 44, 1167, 1970.
- [11] SUNG S.-S. — C. R. Acad. Sc., série D., 274, 1597, 1972.