match no. 4 pp. 93 - 101 1978

DETERMINATION DES STRUCTURES TERPENIQUES SUR ORDINATEUR : PROGRAMME DETER-TERP

Robert Jacquier, Jean-Michel Kornprobst(*)

Laboratoire des Produits Naturels, Faculté des Sciences, Université de Dakar-Fann, République du Sénégal.

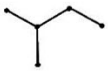
(received: December 1977)

Résumé. Le programme DETER-TERP identifie les fragments générés sur un graphe connexe par un certain nombre de coupures avec l'arbre représentatif de l'unité isoprénique. Ce programme permet de traiter des graphes connexes ayant jusqu'à 30 points et 5 cycles (triterpènes pentacycliques).

Introduction.

La reconnaissance automatique sur ordinateur du caractère " terpènique " d'un graphe représentant le squelette carboné d'une molécule donnée fait actuellement l'objet de diverses études (1) ; nous présentons ici les résultats de l'une d'entre elles.

Le programme DETER-TERP s'applique actuellement à des graphes connexes (2) ayant un nombre de points multiple de 5. Les " unités isopréniques ", représentées par l'arbre:



peuvent être reliées les unes aux autres de façon quelconque.

L'idée originale du programme consiste à couper certaines liaisons afin de faire apparaitre les différentes unités isopréniques composant la molécule initiale. Un certain nombre de critères accélérant le processus de décomposition seront également exposés.

Représentation matricielle utilisée (3).

Le graphe connexe contenant p points est représenté numériquement par une matrice triangulaire dont les côtés sont numérotés de 1 à p, respectivement de haut en bas et de gauche à droite. L'existence d'une liaison entre les atomes i et j est matérialisée par la présence du chiffre 1 au croisement de la ligne i et de la colonne j. Cette représentation matricielle présente les avantages suivants :

1 - La numérotation du graphe initial est arbitraire car il existe une bijection entre la matrice et le graphe étiqueté.

2 - La diagonale de la matrice contient les degrés de chacun des points du graphe ; nous l'appellerons " diagonale des degrés ", en abrégé " ddd ".

3 - La matrice permet d'obtenir immédiatement :

a) Le nombre de cycles. Le graphe étant connexe, le nombre de cycles est donné par :

$$N_{cycles} = \sum q - \sum p + 1$$

Le terme $\sum q$ représente le nombre des lignes du graphe ; il est donné par le nombre des chiffres 1 contenus dans la matrice.

b) Un test de cohérence des données fournies au programme. Pour tout graphe, la somme des degrés des points est égale au double du nombre des lignes (2) ; par conséquent, on doit vérifier :

$$\sum d_i = 2 \sum q$$

c) Le nombre des coupures à réaliser pour obtenir des unités de cinq points (pas nécessairement isopréniques à ce stade du calcul). Si le graphe est un arbre (2), le nombre de coupures à effectuer est égal à : $1/5 \sum p - 1$ et, pour chaque cycle, il faut envisager une coupure supplémentaire. Le nombre de coupures à réaliser pour un graphe connexe quelconque

est donc égal à :

$$N_{coupures} = 1/5 \sum p - 1 + N_{cycles}$$

soit

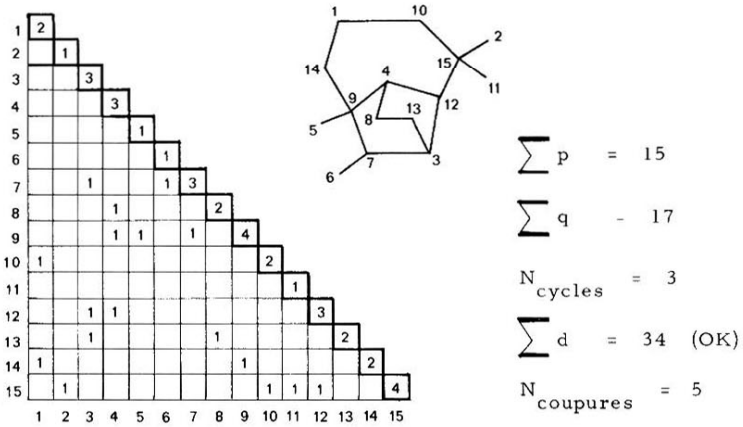
$$N_{coupures} = 1/5 \sum p - 1 + \sum q - \sum p + 1$$

$$= \sum q - 0,8 \sum p$$

La représentation matricielle utilisée et les renseignements qui en découlent sont illustrés pour le graphe représentatif du longifolène.

- Figure 1 -

Représentation matricielle du longifolène



Algorithmes simplificateurs.

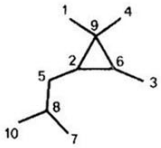
Deux algorithmes permettent de réduire le nombre des liaisons à examiner pour la recherche des coupures.

1 - Principe de la dégradation des degrés des sommets. Ce principe s'énonce comme suit : Tout sommet de degré 1 lié à un sommet de degré supérieur dégrade le degré de celui-ci d'une

unité. Ainsi, on ne cherchera pas à couper de liaisons entre un sommet de degré 1 et un sommet de degré supérieur. Ce principe de dégradation n'est transmissible qu'une seule fois. Il est illustré ci-dessous pour l'acide chrysanthémique.

Acide Chrysanthémique

	1	2	3	4	5	6	7	8	9	10
1	1									
2		3								
3			1							
4				1						
5		1			2					
6		1	1			3				
7							1			
8					1		1	3		
9	1	1		1		1			4	
10								1		1



$\sum p = 10$

$\sum q = 10$

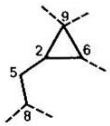
$N_{cycles} = 1$

$\sum d = 20$ (OK)

$N_{coupures} = 2$

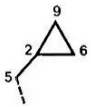
Liste des liaisons initiales : (1-9) ; (2-9) ; (2-5) (2-6) ; (3-6) ; (4-9) ; (5-8) ; (6-9) ; (7-8) ; (8-10).

La dégradation des sommets de degré 1 conduit au graphe suivant :



Pour ce graphe, il reste à considérer les liaisons : (8-5) ; (5-2) ; (2-9) ; (6-9) ; (2-6).

La transmission de la dégradation conduit au nouveau graphe :



Il reste finalement à considérer les liaisons : (2-5) ; (2-6) ; (2-9) ; (6-9).

Avant application de l'algorithme le nombre d'essais nécessaires pour envisager toutes les possibilités de coupure est donné par $C_2^{10} = 45$; après application de l'algorithme, le nombre d'essais est ramené à $C_2^4 = 6$.

2- Principe de la non-génération de sommets de degré 0. Ce principe, appliqué après le précédent sur le graphe " dégradé " permet de réduire encore le nombre des possibilités de coupures.

Ainsi, on ne coupera pas de liaisons reliant un sommet de degré 1 à un autre sommet, quel que soit son degré ; on ne coupera pas deux fois autour d'un sommet de degré 2, ... etc.

L'application de ce deuxième principe au graphe " dégradé " de l'acide chrysanthémique, élimine la coupure de la liaison 2-5. Au total, après application des deux algorithmes simplificateurs, il reste $C_2^3 = 3$ possibilités de coupures au lieu des 45 initiales (la solution est donnée par les coupures 2-6 et 2-9.) Pour l'exemple du longifolène, l'application des deux algorithmes réduit le nombre des possibilités de coupures de 6188 à 1287.

Principe du programme DETER-TERP.

L'arbre représentant l'unité isoprénique est le seul arbre à 5 points ayant la partition (1, 1, 1, 2, 3). Le principe du programme consiste à comparer la partition des fragments obtenus après coupure, avec la partition (1, 1, 1, 2, 3). Cette comparaison se fait en deux temps :

1- Les coupures réalisées, le programme recalcule la " ddd " et vérifie si elle est formée de n fois la partition (1, 1, 1, 2, 3). Cette condition, qui est nécessaire pour que le graphe initial soit formé de n unités isopréniques, n'est cependant pas suffisante.

2- Si la condition précédente est satisfaite, chacun des fragments est alors comparé avec l'unité isoprénique, cette comparaison se faisant pour le nombre des points et la partition de leurs degrés.

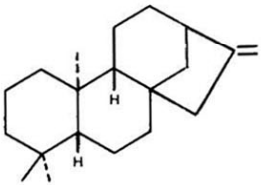
Lorsque tous les fragments sont reconnus identiques à l'unité isoprénique, les coupures trouvées forment une solution.

ANNEXE

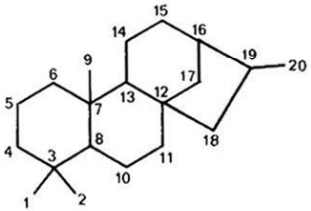
Le programme DETER-TERP fonctionne actuellement sur un ordinateur IBM 1130.

Les listings ci-dessous ont été obtenus avec les graphes de l'ent-Kaurène et de l'acide Rétigéranique. Dans les listings, le paragraphe DEGRE DES SOMMETS donne le numéro du sommet, suivi entre parenthèses, du degré initial et du degré après dégradation.

ent-Kaurène

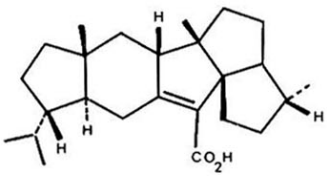


Molécule

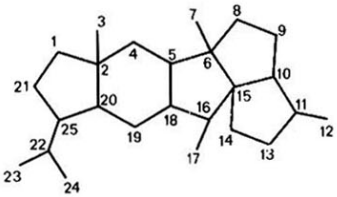


Graphe représentatif

acide Rétigéranique



Molécule



Graphe représentatif

```
PROGRAMME DETER-TERP

ENT-KAURENE

LIAISONS ENTREES

   1- 3    2- 3    3- 4    3- 8    4- 5    5- 6    6- 7
   7- 8    7- 9    8-10    7-13   10-11   11-12   12-13
  12-18   13-14   14-15   15-16   16-17   17-12   16-19
  19-18   19-20

NOMBRE DE SOMMETS  = 20

NOMBRE DE LIAISONS = 23

NOMBRE DE CYCLES   =  4

SOMME DES DEGRES   = 46       (OK)

NOMBRE DE COUPURES A REALISER =  7

LIAISONS SUSCEPTIBLES D'ETRE ROMPUES
   3- 4    4- 5    5- 6    6- 7    3- 8    7- 8    8-10
  10-11   11-12    7-13   12-13   13-14   14-15   15-16
  12-17   16-17   12-18   16-19   18-19

DEGRE DES SOMMETS
 1(1,1)     2(1,1)     3(4,2)     4(2,2)     5(2,2)
 6(2,2)     7(4,3)     8(3,3)     9(1,1)    10(2,2)
11(2,2)    12(4,4)    13(3,3)    14(2,2)    15(2,2)
16(3,3)    17(2,2)    18(2,2)    19(3,2)    20(1,1)

UNE SOLUTION EST OBTENUE PAR COUPURE DE 7 LIAISONS ...
   5- 6    3- 8   10-11    7-13   14-15   12-17   18-19

UNE SOLUTION EST OBTENUE PAR COUPURE DE 7 LIAISONS ...
   5- 6    3- 8   10-11    7-13   14-15   16-17   12-18


FIN DU CAS
```

PROGRAMME DETER-TERP

ACIDE RETIGERANIQUE (SESTERTERPENE PENTACYCLIQUE)

LIAISONS ENTREES

1- 2	2- 3	2- 4	4- 5	5- 6	6- 7	6- 8
8- 9	9-10	10-11	6-15	15-10	11-12	11-13
13-14	14-15	15-16	16-17	16-18	5-18	18-19
19-20	20-25	25-22	22-24	22-23	25-21	1-21
2-20						

NOMBRE DE SOMMETS = 25

NOMBRE DE LIAISONS = 29

NOMBRE DE CYCLES = 5

SOMME DES DEGRES = 58 (OK)

NOMBRE DE COUPURES A REALISER = 9

LIAISONS SUSCEPTIBLES D'ETRE ROMPUES

1- 2	2- 4	4- 5	5- 6	6- 8	8- 9	9-10
10-11	11-13	13-14	6-15	10-15	14-15	15-16
5-18	16-18	18-19	2-20	19-20	1-21	20-25
21-25						

DEGRE DES SOMMETS

1(2,2)	2(4,3)	3(1,1)	4(2,2)	5(3,3)
6(4,3)	7(1,1)	8(2,2)	9(2,2)	10(3,3)
11(3,2)	12(1,1)	13(2,2)	14(2,2)	15(4,4)
16(3,2)	17(1,1)	18(3,3)	19(2,2)	20(3,3)
21(2,2)	22(3,1)	23(1,1)	24(1,1)	25(3,2)

UNE SOLUTION EST OBTENUE PAR COUPURE DE 9 LIAISONS ...

2- 4	8- 9	13-14	6-15	10-15	5-18	18-19
1-21	20-25					

UNE SOLUTION EST OBTENUE PAR COUPURE DE 9 LIAISONS ...

4- 5	5- 6	8- 9	13-14	10-15	15-16	2-20
19-20	21-25					

UNE SOLUTION EST OBTENUE PAR COUPURE DE 9 LIAISONS ...

4- 5	5- 6	9-10	10-15	14-15	15-16	2-20
19-20	21-25					

UNE SOLUTION EST OBTENUE PAR COUPURE DE 9 LIAISONS ...

4- 5	9-10	6-15	10-15	14-15	5-18	2-20
19-20	21-25					

FIN DU CAS

Bibliographie et Notes

1- A. T. BALABAN, F. HARARY, J. M. KORNPROBST, S. MARCUS, unpublished

2- Pour tout renseignement concernant la théorie des graphes voir F. HARARY " Graph Theory ", Addison-Wesley Publishing Company, Reading, Massachussetts 1972.

3- Une revue sur la représentation matricielle des molécules a été publiée récemment par D. H. ROUVRAY dans A. T. BALABAN " Chemical Applications of Graph Theory " Academic Press, 1976, chapitre 7.