

SPECTROMETRIE DE MASSE D'HETEROCYCLES DES ELEMENTS DU GROUPE IV

V*. INTERPRETATION AUTOMATIQUE DU SPECTRE DE MASSE DE COMPOSES HETEROCYCLIQUES COMPORTANT UN ELEMENT DU GROUPE IV

C. LAGEOT

*Laboratoire de Chimie Industrielle, Université de Droit, d'Economie et des Sciences,
13397 Marseille Cédex 4 (France)*

(Reçu le 14 mars 1975)

Summary

A program has been developed for the interpretation of mass spectra. The structures of compounds may be determined in a few seconds.

Résumé

L'utilisation d'un programme pour l'interprétation automatique du spectre de masse de composés permet d'accéder rapidement à la structure de ceux-ci. L'intérêt analytique d'une telle méthode, surtout pour les personnes non "initiales" à la spectrométrie de masse, est certaine. Les résultats obtenus en quelques secondes grâce à cette interprétation automatique ont été corrects pour tous les composés de cette série synthétisés jusqu'à présent.

Introduction

Dans de précédentes études nous avons analysé le comportement, sous bombardement électronique, de molécules ayant pour squelette commun un hétérocycle comportant un atome de silicium ou de germanium [1-3]. L'extension de ces travaux nous a amené à mettre au point un programme sur ordinateur permettant d'élucider la structure de chaque composé analysé, grâce à son spectre de masse.

Dans un but analytique, diverses approches pour l'interprétation de spectres de masse par ordinateur, ont été proposées. Nous n'entrerons pas dans le détail

* Pour la partie IV voir réf. 3.

des moyens de compilation et des méthodes de comparaison qui ont été utilisées [4-12]. Dans ce même but, des programmes spécifiques ont été élaborés pour résoudre le spectre de masse à basse résolution de cétones, d'amines, d'éthers, de thiols et de thioéthers [13-18].

D'autres études utilisaient la spectrométrie de masse à haute résolution pour déterminer la formule développée de peptides [19-23] et d'oestrogènes [24]. L'utilisation de la spectrométrie de masse à haute résolution, malgré l'inconvénient du prix de revient élevé d'un spectre (avec mesure à haute résolution des masses des ions les plus importants) a l'immense avantage de permettre la simplification appréciable du programme, permettant ainsi une détermination aisée de la structure du composé analysé.

Nous avons montré, dans de précédents travaux [1,2], les bonnes corrélations spectre de masse—structure d'une série d'hétérocycles des éléments du groupe IV. Des mécanismes communs de fragmentation ont pu être établis. Ce sont ces mécanismes communs qui ont été transcrits en langage machine. Nous pouvons donc, grâce à l'assistance de l'ordinateur, avoir une interprétation automatique des spectres de masse et accéder à la formule développée du composé analysé.

Pour cela nous procéderons en trois étapes: (1) détermination de l'ion moléculaire, (2) localisation et somme des compositions élémentaires des substituants de l'hétérocycle; (3) choix d'un code topologique pour générer les isomères à partir d'une composition élémentaire et adoption de discriminateurs tirés de l'IR, de la RMN ou de la synthèse pour n'avoir qu'une seule formule développée comme solution finale.

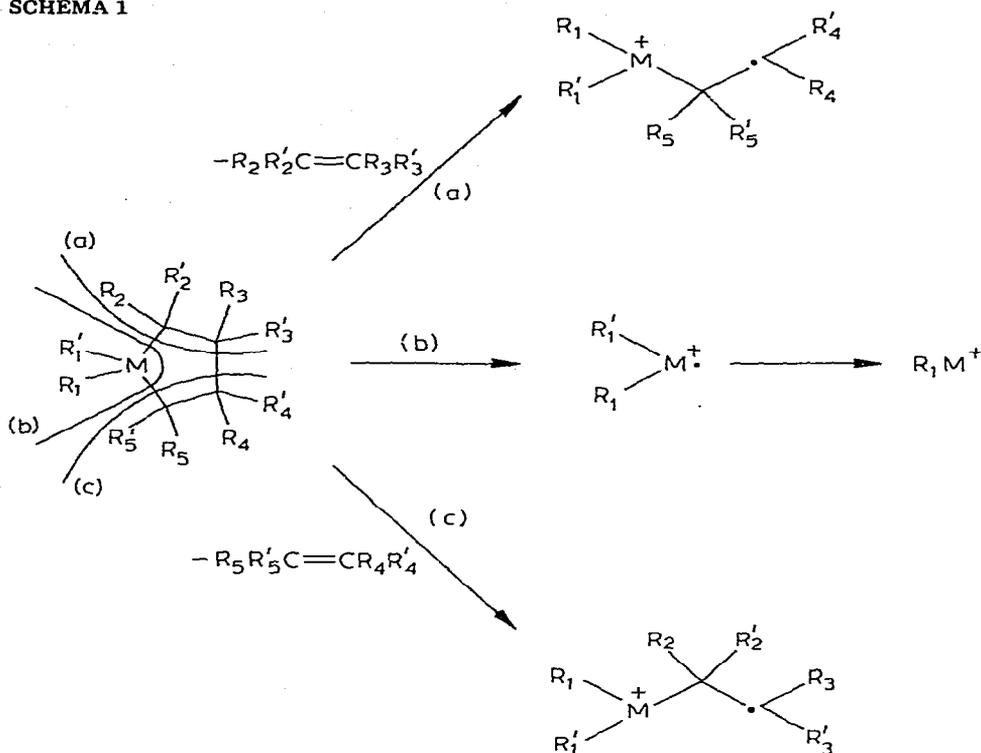
Méthode. Résultats

Nous avons vu que les principaux ions du spectre de masse d'un composé du Tableau 1 sont pour la plupart dûs à des mécanismes communs de fragmentation. Les dix ions les plus importants seront donc retenus et classés par ordre de masse croissante. Il est inutile d'en retenir davantage; ce nombre de dix a été suffisant pour donner une solution juste pour tous les composés synthétisés jusqu'à présent. Deux possibilités s'offrent alors: le dernier ion (de masse la plus élevée) est l'ion moléculaire ou il ne l'est pas.

Ion moléculaire

Si l'on dispose d'un appareillage moderne incluant une source à bombardement électronique et une source à ionisation chimique (ou à ionisation de champ) le dernier ion (de masse la plus élevée) indiqué par l'un des spectres peut être considéré avec une très grande probabilité comme étant l'ion moléculaire. Si l'on ne dispose que d'une source à bombardement électronique, le dernier ion peut ne pas être l'ion moléculaire. Pour que le programme ne donne pas comme résultat final: "pas de solution correcte", il faudra alors faire intervenir soit la présence éventuelle de pic métastable, soit l'écartement éventuel des deux pics les plus élevés en masse (il est bien évident que si ces deux pics sont séparés par 3 u.m.a., par exemple, il est hautement improbable que nous ayons affaire au départ de trois atomes d'hydrogène. Nous pourrions avoir affaire dans ce cas là, aux ions $M - H_2O^+$ et $M - CH_3^+$) pour que la solution finale soit juste. Cet ion

SCHEMA 1

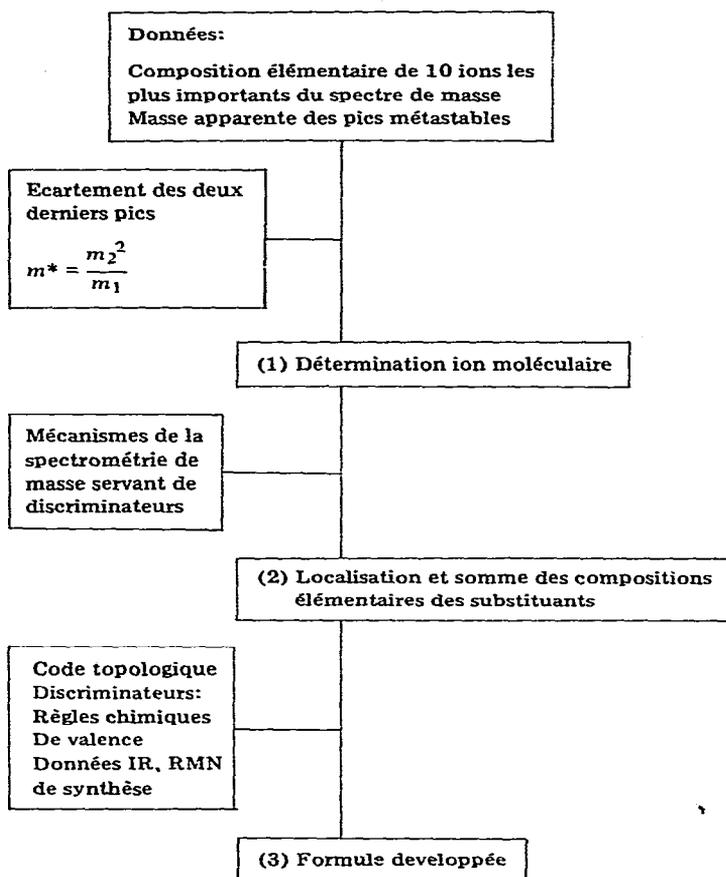


Le programme a été élaboré de telle sorte qu'il donne la formule brute de l'ion moléculaire puis les localisations et les sommes des compositions élémentaires des substituants ($R_2R'_2C=CR_3R'_3$, $R_4R'_4C=CR_5R'_5$ et $R_1MR'_1$) pour tous les hétérocycles pentagonaux comprenant un élément du groupe IV synthétisés jusqu'à présent (Tableau 1). Pour tester ce programme des molécules hypothétiques ont été envisagées, leur spectre a été simulé puis interprété (Tableau 2, molécules no. 13 et 14). Pour avoir finalement la formule développée de chacun des composés analysés, il faut établir un système qui permette d'obtenir tous les isomères à partir d'une formule brute. Il faut donc adopter un code topologique pour résoudre ce problème particulier. La liste des isomères obtenue sera réduite par une série de discriminateurs tenant compte des données IR, RMN ou de la synthèse du composé analysé.

Code topologique

A l'heure actuelle il n'existe aucune représentation universelle des molécules qui puisse permettre de résoudre tous les problèmes des chimistes. Différents modèles ont été proposés [25-31], chacun dans le but de résoudre des problèmes spécifiques. Notre problème consiste à énumérer tous les isomères à partir d'une formule brute. Après des essais infructueux [32-35] cette question vient d'être résolue [35]. Nous avons donc utilisé ce code topologique et nous allons donner quelques indications sur le principe de ce même code topologique pour nous familiariser avec les réponses qui seront données ultérieurement.

SCHEMA 2



La notation de ce code décrivant une molécule doit être univoque. Pour cela, l'énumération des atomes et des liaisons se fera dans l'ordre de leur connectivité à partir d'un point privilégié. On a montré en effet il y a fort longtemps, que chaque "graphe" a un centre unique appelé "centre de masse" ou "centroïde". C'est le point du graphe d'où partent les atomes de la manière la plus équilibrée possible. Sans entrer dans le détail de la notation (sachons seulement qu'une liaison est représentée par un point, une double liaison par deux points et que les atomes d'hydrogène ne sont pas représentés) qui a été décrite par ailleurs [35], nous donnerons à titre d'exemple la notation des premiers carbures saturés et des premiers isomères parmi les onze ayant pour formule brute $C_4H_8O_1$ (Tableau 3). Ce code permet, nous l'avons dit, de générer tous les isomères à partir d'une formule brute. Il convient donc parfaitement pour résoudre notre problème, dont nous allons donner ci après quelques résultats.

Résultats

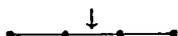
Nous allons prendre à titre d'exemple le composé no. 3, car il se prête à

TABLEAU 2

| | |
|----------------------|----------------------|
| Molécule numéro & 9 | Molécule numéro & 12 |
| GE 1 C10 H14 O 0 N 0 | GE 1 C12 H18 O 0 N 0 |
| GE 0 C 2 H 4 O 0 N 0 | GE 0 C 2 H 4 O 0 N 0 |
| GE 0 C 2 H 4 O 0 N 0 | GE 0 C 2 H 4 O 0 N 0 |
| GE 1 C 6 H 6 O 0 N 0 | GE 1 C 8 H10 O 0 N 0 |
| Molécule numéro & 10 | Molécule numéro & 13 |
| GE 1 C16 H18 O 0 N 0 | GE 1 C11 H16 O 1 N 0 |
| GE 0 C 2 H 4 O 0 N 0 | GE 0 C 2 H 4 O 0 N 0 |
| GE 0 C 2 H 4 O 0 N 0 | GE 0 C 2 H 4 O 1 N 0 |
| GE 1 C12 H10 O 0 N 0 | GE 1 C 7 H 8 O 0 N 0 |
| Molécule numéro & 11 | Molécule numéro & 14 |
| GE 1 C10 H18 O 0 N 0 | GE 1 C12 H18 O 1 N 0 |
| GE 0 C 2 H 4 O 0 N 0 | GE 0 C 2 H 4 O 1 N 0 |
| GE 0 C 2 H 4 O 0 N 0 | GE 0 C 3 H 6 O 0 N 0 |
| GE 1 C 6 H10 O 0 N 0 | GE 1 C 7 H 8 O 0 N 0 |

TABLEAU 3

EXEMPLE DE LA NOTATION DES QUELQUES CARBURES AYANT POUR FORMULE BRUTE $C_4H_8O_1$

| Molécule | Notation ^a | Formule développée | Notation ^a |
|------------|---|---------------------|-----------------------|
| Ethane |  . C C | $H_3CCH_2CH_2CH=O$ | C . . C . CC : 0 |
| Propane |  C . . C C | $H_2C=CHCH_2CH_2OH$ | C . . C : CC . 0 |
| n-Butane |  . C . C C . C | $H_2C=CHCH_2OCH_3$ | C . . C : CO . C |
| i-Butane |  C . . C C C | $H_3CCH_2CH=CHOH$ | C . : C . CC . 0 |
| n-Pentane |  C . . C . C C . C | $H_3CCH=CHCH_2OH$ | C : . C . CC . 0 |
| Néopentane |  C . . . C C C C | $H_2C=CHOCH_2CH_3$ | 0 . . C . CC : C |

^a Voir le texte pour explication de la notation.

toutes les déterminations considérées précédemment; pour ce composé les données fournies à l'ordinateur sont les suivantes:

| Ge | C | H | O | N | X |
|----|----|----|----|----|----|
| 01 | 07 | 15 | 01 | 00 | 00 |
| 01 | 08 | 16 | 00 | 00 | 00 |
| 01 | 07 | 13 | 00 | 00 | 00 |
| 01 | 05 | 12 | 01 | 00 | 00 |
| 01 | 05 | 12 | 00 | 00 | 00 |
| 01 | 03 | 09 | 01 | 00 | 00 |
| 01 | 02 | 07 | 01 | 00 | 00 |
| 01 | 02 | 06 | 00 | 00 | 00 |
| 01 | 01 | 03 | 00 | 00 | 00 |
| 00 | 03 | 07 | 00 | 00 | 00 |

Ce sont les compositions élémentaires des dix ions les plus importants du spectre de masse. La masse apparente du dernier pic métastable est fournie à l'ordinateur (soit $m^* = 169.5$). L'ordinateur calcule, grâce à cette masse du pic métastable, la composition de l'ion moléculaire. Ce résultat est confirmé par le test de l'écartement des deux derniers pics. Ce calcul donne la composition élémentaire suivante: $\text{Ge}_1\text{C}_3\text{H}_{18}\text{O}_1$. Grâce ensuite à la simulation des situations (transcription en langage machine des mécanismes intervenant en spectrométrie de masse) les sommes sous indiquées sont calculées (éqn. 1).

$$R_2 + R'_2 + 2C + R_3 + R'_3, R_4 + R'_4 + 2C + R_5 + R'_5 \text{ et } M + R_1 + R'_1 \quad (1)$$

Les résultats apparaissent dans le Tableau 4. Intervient alors le sous programme concernant le code topologique. Pour le premier résultat C_3H_6 , il n'y a aucune ambiguïté. Il n'existe qu'un seul fragment noté C.: CC soit $\text{H}_2\text{C}=\text{CHCH}_3$. Quant aux sommes $R_4 + R'_4 + 2C + R_5 + R'_5$, elles sont égales à $\text{C}_3\text{H}_6\text{O}_1$. Il existe six isomères correspondant à cette brute que le code topologique utilisé, représente par:

.C:CC.O :C.CC.O

.C:CO.C C.:CC:O

.C.CC:O C.:CCO

Cette liste est réduite par le fait que le fragment neutre expulsé issu du mécanisme a (ou c) doit posséder une double liaison carbone-carbone. Cette liste est encore réduite si nous faisons appel à des règles dites chimiques. En effet les structures contenant des groupements fonctionnels qui ne sont pas connus ou qui sont instables seront éliminés:

C=C-OH N=N-N H-C-N=O

C-C-NH N=N-H H-O-C=N-

C-C-OH H-O-C-O-H N-O-N

C=C-NH H-O-C-N-H N-O-O

Enfin cette liste est encore réduite si nous faisons appel, quand cela est nécessaire,

TABLEAU 4

LES RESULTATS DES CALCULATIONS DE L'EQUATION 1

| Ion moléculaire et fragments issus des mécanismes a, c et b | Notation des fragments | Fragments correspondants | Molécules |
|--|---------------------------|-----------------------------|-----------|
| Molécule numéro & 1 | | | |
| GE 1 C 6 H 14 O 1 N 0 X 0 | | | |
| GE 0 C 2 H 4 O 0 N 0 X 0 | &CC | | |
| GE 0 C 2 H 4 O 1 N 0 X 0 | C.&OC | | |
| GE 1 C 2 H 6 O 0 N 0 X 0 | .CC | | |
| Molécule numéro & 2 | | | |
| GE 1 C 7 H 16 O 1 N 0 X 0 | | | |
| GE 0 C 2 H 4 O 0 N 0 X 0 | &CC | | |
| GE 0 C 3 H 6 O 1 N 0 X 0 | C.&.CCO | | |
| GE 1 C 2 H 6 O 0 N 0 X 0 | .CC | | |
| Molécule numéro & 3 | | | |
| GE 1 C 8 H 18 O 1 N 0 X 0 | | | |
| GE 0 C 3 H 6 O 0 N 0 X 0 | C.&CC | | |
| GE 0 C 3 H 6 O 1 N 0 X 0 | C.&.CCO | | |
| GE 1 C 2 H 6 O 0 N 0 X 0 | .CC | | |

à des données tirées soit de la synthèse, soit des spectres IR ou RMN. Pour le composé no. 3 le résultat de la somme $R_4 + R'_4 + 2C + R_5 + R'_5$ est noté C.:.CCO soit $H_2C=C(OH)CH_3$. Enfin la somme $R_1 + R'_1$ est notée .CC soit H_3C-CH_3 . A l'aide de ces résultats la formule développée peut être reconstituée.

Conclusion

Ainsi grâce aux dix ions les plus importants du spectre de masse et certaines fois grâce à quelques données tirées des spectres IR, RMN ou de la synthèse, la structure du composé analysé a toujours été déterminé pour cette série. La réponse de l'ordinateur est plus complète et beaucoup plus rapide (quelques secondes pour un IBM 1130) que celle que pourrait donner un "spectroscopiste de masse" entraîné. Du point de vue analytique, l'intérêt d'une telle méthode est évidente. A l'heure actuelle, il est possible à l'aide d'un programme adapté à un type de composé, de connaître la structure de ses dérivés. Pour généraliser ce type de démarche, il est nécessaire de mieux connaître les mécanismes réactionnels de la spectrométrie de masse, donc de mieux connaître le phénomène d'ionisation qui est à la base de celle-ci [36-38].

Bibliographie

- 1 C. Lageot, J.C. Maire, G. Manuel et P. Mazerolles, *J. Organometal. Chem.*, 54 (1973) 131.
- 2 C. Lageot, J.C. Maire, G. Manuel et P. Mazerolles, *J. Organometal. Chem.*, 60 (1973) 55.
- 3 C. Lageot, J.C. Maire, P. Rivièrè, M. Massol et J. Barrau, *J. Organometal. Chem.*, 66 (1974) 49.
- 4 P.C. Purs, *Anal. Chem.*, 43 (1971) 1812.
- 5 L.R. Crawford et J.D. Morrison, *Anal. Chem.*, 43 (1971) 179.
- 6 D.H. Smith, *Anal. Chem.*, 44 (1972) 536.
- 7 H.S. Hertz, R.A. Hites et K. Biemann, *Anal. Chem.*, 43 (1971) 681.
- 8 K. Biemann, P. Bommer et D.M. Desiderio, *Tetrahedron Lett.*, (1964) 1725.
- 9 C. Merritt, P. Issenberg, M.L. Bazinet, B.N. Green, T.O. Merron et J.G. Murray, *Anal. Chem.*, 37 (1965) 1037.
- 10 R. Venkataraghanan, F.W. Mc Lafferty et J.W. Amy, *Anal. Chem.*, 39 (1967) 178.
- 11 A.L. Burlingham, D.H. Smith et R.W. Olsen, *Anal. Chem.*, 40 (1968) 13.
- 12 R.H. Hites et K. Biemann, *Anal. Chem.*, 39 (1967) 965; idem, 40 (1968) 1217.
- 13 J. Lederberg, G.L. Sutherland, B.G. Buchanan, E.A. Feigenbaum, A.V. Robertson, A.M. Duffield et C. Djerassi, *J. Amer. Chem. Soc.*, 91 (1969) 2973.
- 14 A.M. Duffield, A.V. Robertson, C. Djerassi, B.G. Buchanan, G.L. Sutherland, E.A. Feigenbaum et J. Lederberg, *J. Amer. Chem. Soc.*, 91 (1969) 2977.
- 15 G. Schroll, A.M. Duffield, C. Djerassi, A.B. Delfino, B.G. Buchanan, G.L. Sutherland et J. Lederberg, *J. Amer. Chem. Soc.*, 92 (1969) 7740.
- 16 A. Buchs, A.M. Duffield, G. Schroll, C. Djerassi, A.B. Delfino, B.G. Buchanan, G.L. Sutherland, E.A. Feigenbaum et J. Lederberg, *J. Amer. Chem. Soc.*, 92 (1970) 6831.
- 17 A. Buchs, A.B. Delfino, A.M. Duffield, C. Djerassi, B.G. Buchanan, E.A. Feigenbaum et J. Lederberg, *J. Amer. Chem. Soc.*, 53 (1970) 1394.
- 18 Y.M. Sheikh, A. Buchs, A.B. Delfino, G. Schroll, A.M. Duffield, C. Djerassi, B.G. Buchanan, G.L. Sutherland, E.A. Feigenbaum et J. Lederberg, *Org. Mass. Spectrom.*, 94 (1970) 493.
- 19 A. Mandelbaum, P.V. Fennessey et K. Biemann, *Proc. Amer. Conf. Mass. Spectrom. Allied Topics*, 15 (1967) 111.
- 20 W.J. Richter, B.R. Simoneit, D.U. Smith et A.L. Burlingham, *Anal. Chem.*, 41 (1969) 1392.
- 21 R. Venkatarachavan, F.W. Mc Lafferty et G.E. Van Lear, *Org. Mass. Spectrom.*, 2 (1969) 1.
- 22 M. Senn, R. Venkatarachavan et F.W. Mc Lafferty, *J. Amer. Chem. Soc.*, 88 (1966) 5593.
- 23 K. Biemann, C. Cone, B.R. Nebster et G.P. Arsenault, *J. Amer. Chem. Soc.*, 88 (1966) 5598.
- 24 D.H. Smith, B.G. Buchanan, R.S. Engelmöre, A.M. Duffield, A. Yed, E.A. Feigenbaum, J. Lederberg et C. Djerassi, *J. Amer. Chem. Soc.*, 94 (1972) 5062.
- 25 E.J. Corey, *Quant. Rev. Chem. Soc.*, 25 (1971) no. 4.
- 26 J.B. Hendrikson, *J. Amer. Chem. Soc.*, 93 (1971) 6847 et 6854.
- 27 J.E. Dubois et H. Viellard, *Bull. Soc. Chim. Fr.*, (1968) 900, 905 et 913.
- 28 D. Lefkowitz, *J. Chem. Soc.*, 7 (1967) 186.
- 29 W.J. Wisswesser, *A Line Formula Chemical Notation*, Crowell, New York, N.Y., 1954.
- 30 L. Spialter, *J. Amer. Chem. Soc.*, 4 (1964) 261.
- 31 H.L. Morgan, *J. Chem. Soc.* 6 (1966) 81.
- 32 A. Caley, *Ber.*, 8 (1875) 1056.
- 33 H.R. Henze et C.M. Blair, *J. Amer. Chem. Soc.*, 53 (1931) 3077.
- 34 A.T. Balaban, *Rev. Chim. Acad. Rey. Pop. Roum.*, 12 (1967) 875.
- 35 J. Lederberg, G.L. Sutherland, B.G. Buchanan, E.A. Feigenbaum, A.V. Robertson, A.M. Duffield et C. Djerassi, *J. Amer. Chem. Soc.*, 91 (1969) 2973.
- 36 C. Lageot, *Org. Mass. Spectrom.*, 5 (1971) 839.
- 37 C. Lageot, *Org. Mass. Spectrom.*, 5 (1971) 845.
- 38 C. Lageot, *Org. Mass. Spectrom.*, 5 (1971) 891.