

CHROM. 7533

IDENTIFICATION DE COMPOSÉS ALIPHATIQUES SATURÉS PAR CHROMATOGRAPHIE EN PHASE GAZEUSE

II. COMPOSÉS À DEUX DIRECTIONS DE DÉVELOPPEMENT

MAURICE CHASTRETTÉ

Laboratoire de Chimie Organique Physique, Université Lyon 1, 43 Boulevard du 11 Novembre 1918, 69100-Villeurbanne (France)

GÉRARD LENFANT

Laboratoire de Chimie Organique, E.N.S.C.L. Cité Scientifique, BP 40, 59650-Villeneuve d'Ascq (France)

et

ANNIE REMY

Centre Interuniversitaire de Traitement de l'Information, Université des Sciences et Techniques, Cité Scientifique, BP 36, 59650-Villeneuve d'Ascq (France)

(Reçu le 19 avril 1974)

SUMMARY

Gas chromatographic identification of saturated aliphatic compounds. II. Compounds developed in two directions

Retention-structure relations are used to identify the carbon chains R or R' of saturated aliphatic compounds obeying the general formula RX or R-X-R' (X being the chemical function). A general programme, written in Algol 60, enables us to propose one or more structures for a compound whose family and retention time are known.

INTRODUCTION

Nous avons proposé récemment une méthode d'identification des composés chimiques par chromatographie gaz-liquide, à l'aide d'un ordinateur¹. Cette méthode utilise une relation générale de topologie-information reliant les structures des composés et leurs grandeurs de rétention²⁻⁶. L'identification est réalisée à l'aide d'un programme écrit en Algol 60, dans le cas de composés du type RX où R est une chaîne carbonée saturée et X une fonction chimique.

Dans cet article, nous étendons ce traitement au cas plus complexe des composés du type R-X-R' où les deux chaînes carbonées R et R' liées au groupe fonctionnel X peuvent présenter des interactions.

PRINCIPE DE LA MÉTHODE

L'identification d'un composé chimique à partir de sa rétention se fait en deux étapes.

Établissement d'un modèle mathématique

Ce modèle mathématique relie la rétention du composé dans des conditions chromatographiques bien déterminées à sa structure exprimée par son topomodèle dans le langage du système Darc⁷⁻¹⁰. De tels modèles ayant été décrits pour plusieurs familles chimiques dans nos travaux antérieurs²⁻⁶ nous ne rappellerons pas la méthode qui permet de les établir. Ce traitement fait intervenir des paramètres structuraux I_i caractéristiques des positions de l'environnement carboné de la fonction chimique et des interactions entre ces positions. Les paramètres I_i dépendent de la température selon la relation :

$$I_i = a_i/T + b_i \quad (1)$$

Utilisation du modèle pour l'identification

La relation générale rétention-structure² ainsi obtenue permet de calculer la rétention d'un composé dont on connaît la structure, à toute température comprise dans le domaine étudié. Le programme établi permet d'engendrer des structures, de calculer les grandeurs de rétention correspondantes, de les comparer à la grandeur de rétention expérimentale et de sélectionner les structures pour lesquelles les grandeurs calculées et observées sont suffisamment voisines.

Pratiquement on introduit dans l'ordinateur les coefficients a_i et b_i des paramètres I_i de la relation générale, le nom de la famille chimique à laquelle est supposé appartenir le composé à identifier, la température, et enfin la valeur de la grandeur de rétention observée. Le traitement fournit une liste de composés dont la structure est compatible avec la valeur expérimentale introduite.

ÉTABLISSEMENT DU PROGRAMME

Le programme est établi à partir des principes suivants :

(1) Les directions de développement sont divisées en branches explorées selon la méthode décrite précédemment.

(2) Les interactions entre les directions de développement apparaissant lors de l'exploration des deux directions de développement sont repérées.

Proposons-nous, par exemple, d'identifier un ester R-COO-R' dont le logarithme du temps de rétention relatif à l'acétate de méthyle est YD . On recherche d'abord si un ester méthylique R-COO-CH₃ ou un acétate CH₃-COO-R' constitue une solution acceptable. Cette recherche s'effectue selon le traitement précédemment présenté¹ pour les composés à une direction de développement. On envisage successivement pour chaque direction de développement DD (selon R) et DD' (selon R') trois branches de développement II parcourues chacune dans le sens dicté par la règle des priorités entre les positions occupées dans le topomodèle.

Ensuite, on se propose de trouver un ester R-COO-R' qui soit solution. La méthode de traitement de ce cas général consiste à effectuer en partant de l'élément

structural —C—COO—C— la génération d'une des deux directions de développement (on choisit DD') selon les trois types de branche $I1$ précédemment définis¹ tout en développant simultanément l'autre direction de développement (DD), en tenant compte des interactions correspondantes. Mais une telle systématique impose de pointer les différentes positions de substitutions successives de DD' afin de repérer les interactions avec les positions de DD . On définit alors (Fig. 1) une matrice carrée 12/12 d'interaction MI [I, J] à laquelle est adjoint un tableau booléen qui permet de

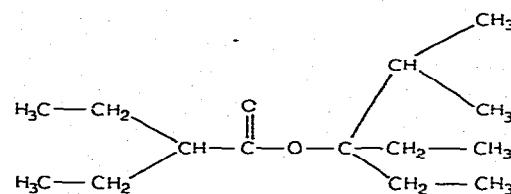


Tableau booléen $TP[L]$	DD Branche $IP1=2$ DD' Branche $IP2=3$	$A_1 \rightarrow A_2 \rightarrow B_{11} \rightarrow B_{21}$ $B_{12} \rightarrow B_{22}$ $B_{13} \rightarrow B_{23}$
vrai	A_1 ↓ A_2 ↓ A_3 ↓ B_{11}	1 1 0 1 0 0 1 0 0
vrai		1 1 0 1 0 0 1 0 0
vrai		1 1 0 1 0 0 1 0 0
vrai		1 1 0 1 0 0 1 0 0
vrai		1 1 0 1 0 0 1 0 0
faux		0 0 0 0 0 0 0 0 0
vrai	B_{21} ↓ B_{12} ↓ B_{13}	1 1 0 1 0 0 1 0 0
faux		0 0 0 0 0 0 0 0 0
faux		0 0 0 0 0 0 0 0 0
vrai	B_{31} ↓ B_{32} ↓ B_{33}	1 1 0 1 0 0 1 0 0
faux		0 0 0 0 0 0 0 0 0
faux		0 0 0 0 0 0 0 0 0

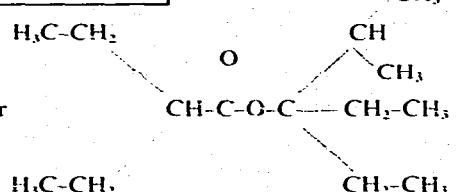


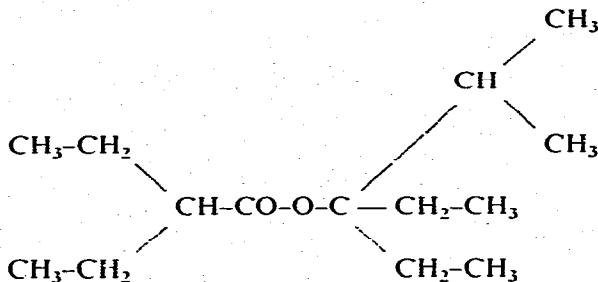
Fig. 1. Constitution de la matrice-solution relative à l'ester

repérer les interactions en y inscrivant "vrai" ou "faux", si la position correspondante de *DD* est occupée ou non. Pour chaque étape de la progression, on calcule au moyen de la relation rétention-structure générale le logarithme du temps de rétention relatif théorique *Y* et on teste la relation 2:

$$|YD - Y| \leq 0.006 \quad (2)$$

où 0,006 représente l'erreur expérimentale admise pour le logarithme du temps de rétention relatif.

La Fig. 1 représente la matrice-solution correspondant à l'ester suivant:



Le passage de cette matrice solution au nom du composé dans le langage habituel des chimistes organiciens se fait par l'intermédiaire du Descripteur par Environnement Limité (DEL)⁸ obtenu facilement à partir de la matrice. Une table de correspondance entre les DEL et les groupes alkyles (pour le premier environnement limité seulement) permet d'exprimer la solution sous une forme plus immédiatement compréhensible. Ainsi la matrice-solution de la Fig. 1 peut être représentée en utilisant les DEL par:



et, d'après le Tableau I, l'ordinateur fournira la solution suivante:

<i>p</i>	Groupe Alkyle	Fonction	Groupe alkyle	<i>p'</i>
<i>DD</i>			<i>DD'</i>	
0	CH (Et) ₂	-COO-	C (iPr) (Et) ₂	0

TABLEAU I

CORRESPONDANCE ENTRE LE DEL DES GROUPES ALKYLES R LIMITÉS AU PREMIER *E_B* ET LEUR REPRESENTATION HABITUELLE

No.	DEL	Groupes alkyles	No.	DEL	Groupes alkyles
1	(0000)	Me	19	(3320)	C (iPr) ₂ (Et)
2	(1000)	Et	20	(3330)	C (iPr) ₃
3	(2000)	<i>i</i> Pr	21	(1111)	CH ₂ (<i>t</i> Bu)
4	(3000)	<i>t</i> Bu	22	(2111)	CH (Me) (<i>t</i> Bu)
5	(1100)	Pr	23	(2211)	CH (Et) (<i>t</i> Bu)
6	(2109)	<i>s</i> Bu	24	(2221)	CH (<i>i</i> Pr) (<i>t</i> Bu)
7	(3100)	<i>t</i> Am	25	(3111)	C (Me) ₂ (<i>t</i> Bu)
8	(2200)	CH (Et) ₂	26	(3211)	C (Me) (Et) (<i>t</i> Bu)
9	(3200)	C (Me) Et ₂	27	(3221)	C (Me) (<i>i</i> Pr) (<i>t</i> Bu)
10	(3300)	C (Et) ₃	28	(2222)	CH (<i>t</i> Bu) ₂
11	(1110)	<i>i</i> Bu	29	(3222)	C (Me) (<i>t</i> Bu) ₂
12	(2110)	CH (Me) (<i>i</i> Pr)	30	(3311)	C (Et) ₂ (<i>t</i> Bu)
13	(2210)	CH (Et) (<i>i</i> Pr)	31	(3321)	C (Et) (<i>i</i> Pr) (<i>t</i> Bu)
14	(3110)	C (Me) ₂ (<i>i</i> Pr)	32	(3331)	C (iPr) ₂ (<i>t</i> Bu)
15	(3210)	C (Me) (Et) (<i>i</i> Pr)	33	(3322)	C (Et) (<i>t</i> Bu) ₂
16	(3310)	C (Et) ₂ (<i>i</i> Pr)	34	(3332)	C (iPr) (<i>t</i> Bu) ₂
17	(2220)	CH (<i>i</i> Pr) ₂	35	(3333)	C (<i>t</i> Bu) ₃
18	(3220)	C (Me) (<i>i</i> Pr) ₂			

Le nombre de positions p extérieures à l' E^1_B occupées est indiqué aussi le cas échéant. Lorsque l'exploration est terminée, les noms de tous les composés dont la structure est telle que la relation 2 soit vérifiée ont été imprimés.

ORGANIGRAMME DU PROGRAMME

Le programme est écrit en Algol 60 pour un ordinateur: Bull-Honeywell M 40. La Fig. 2 représente l'organigramme du programme principal. Après l'introduction des données propres aux directions de développement DD (notée B) et DD' (notée BP) intervient la procédure "Direction de développement B " équivalente au programme précédemment établi¹ et dans laquelle sont incluses les procédures "Branche" et

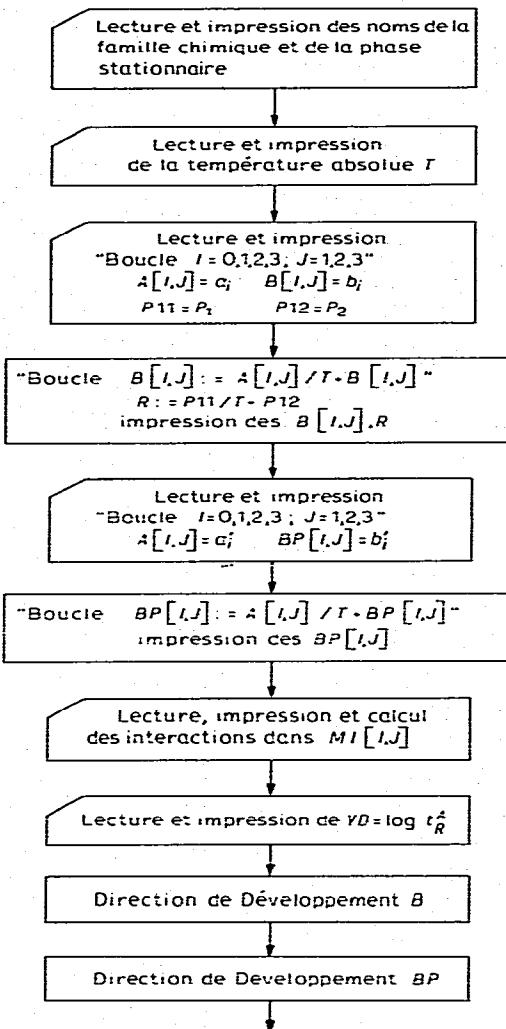


Fig. 2. Organigramme du programme principal d'identification des composés à deux directions de développement.

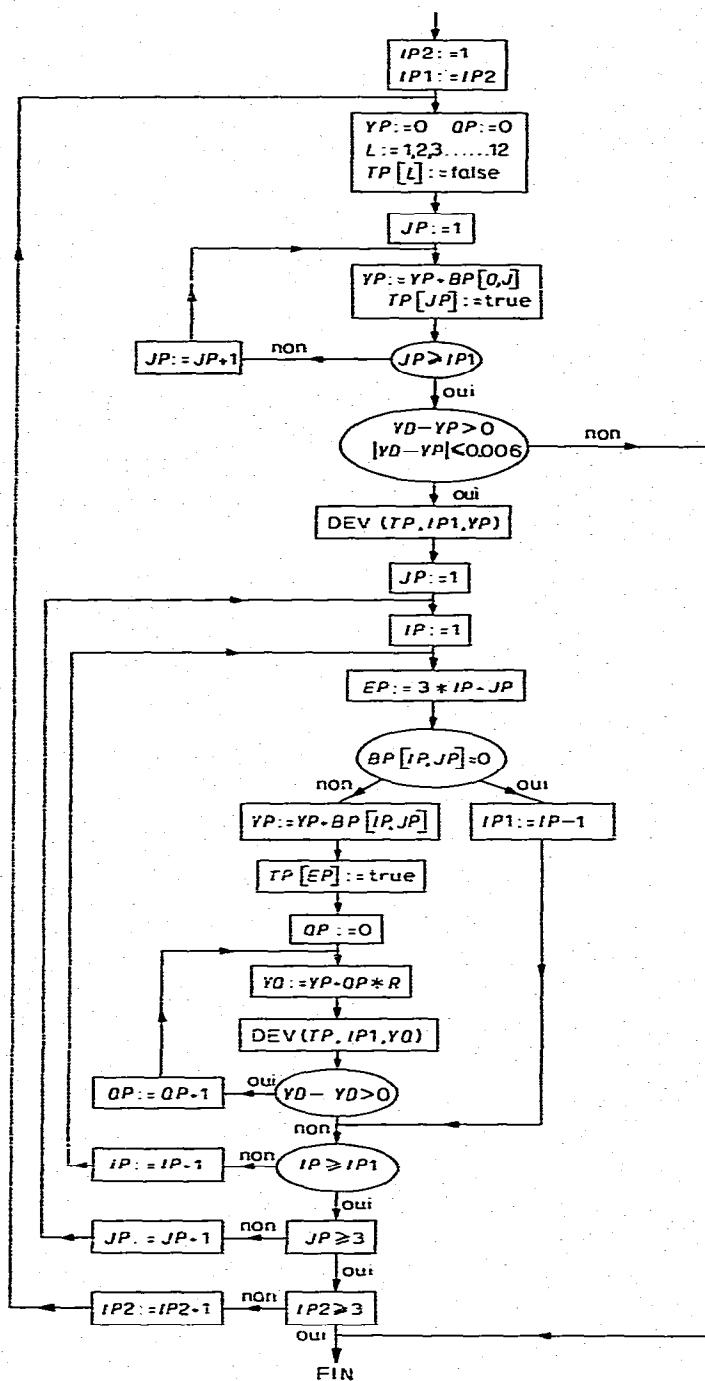


Fig. 3. Organigramme de la procédure DEV (TP, IP1, Y).

“P” également définies antérieurement. Ensuite on envisage de la même façon la procédure “Direction de développement BP”.

On choisit alors de développer DD' selon la branche notée $IP2$ de type 1 ($IP2 = 1$) tout en développant DD selon la procédure “DEV” (Fig. 3), les positions étant repérées dans le tableau booléen unicolonne $TP[L]$. Dans cette procédure “DEV” les termes B_{ij} sont introduits à l'aide de la procédure “Braca” qui est très semblable à la procédure “Branche”.

TEST DU PROGRAMME

Le programme d'identification ainsi élaboré est général, en ce sens que toutes les positions situées dans chaque segment d'environnement limité E^1_B selon chaque direction de développement ont été envisagées. En outre dans la matrice d'interaction, la possibilité d'existence de toutes les interactions entre directions de développement est mentionnée dans le but de ne pas diminuer la généralité du programme. En d'autres termes tous les paramètres de substitution et d'interaction associés aux positions correspondantes pour une famille chimique donnée sont supposés connus, c'est à dire déterminés au préalable au moyen du modèle mathématique. Dans la pratique, ceci n'est pas toujours réalisé. En particulier il y a lieu, dans la recherche des composés à deux directions de développement, de supprimer des données les paramètres de substitution correspondant aux positions pour lesquelles les paramètres d'interaction ne peuvent être définis. En tenant compte de ces remarques, le programme général a été testé pour l'ensemble des données qui sont à notre disposition, à savoir les logarithmes des grandeurs de rétention relatives de cétones et d'esters aliphatiques saturés dont nous avons précédemment étudié les relations rétention-structure²⁻⁶.

YD = 0 3600					
DIRECTION DE DÉVELOPPEMENT B					
DIRECTION DE DÉVELOPPEMENT BP					
2 DIRECTIONS DE DÉVELOPPEMENT					
O	Et	—CO—	Et	O	Y=0 3589
YD = 0 5140					
DIRECTION DE DÉVELOPPEMENT B					
1	Pr	—CO—	Me		Y=0 5193
DIRECTION DE DÉVELOPPEMENT BP					
Me	—CO—		Pr	1	Y=0 5193
2 DIRECTIONS DE DÉVELOPPEMENT					
O	Pr	—CO—	Et	O	Y=0 5144
O	Et	—CO—	Pr	O	Y=0 5144

Fig. 4. Exemples d'identification de cétones aliphatiques saturées. Température: 210°.

Les cétones aliphatiques saturées R-CO-R' présentent deux directions de développement identiques, c'est à dire qu'il n'y a pas à les distinguer dans l'expression des résultats. Par ailleurs la méthode utilisée dans la recherche des structures conduit à l'obtention de couples de solutions symétriques. C'est ce que l'on remarque sur la Fig. 4 qui est une copie des résultats donnés directement par l'ordinateur pour deux valeurs YD des logarithmes des temps de rétention relatifs obtenus à 210° sur SE-30^{2,4,6}.

Les esters aliphatiques saturés ont deux directions de développement différentes: il y a donc lieu d'imprimer les groupes alkyles au bon endroit de part et d'autre du groupe fonctionnel. Par exemple, pour $YD = 0.4870$ (Fig. 5) l'impression suivante serait fausse:



Le programme a été testé pour les 45 esters de notre échantillon à cinq températures, et la Fig. 5 donne un exemple de résultats obtenus sur ordinateur. Pour les

YD = 0.3960
DIRECTION DE DÉVELOPPEMENT B
DIRECTION DE DÉVELOPPEMENT BP
Me —CO— O sBu O
Y = 0.3960
2 DIRECTIONS DE DÉVELOPPEMENT
O Et —CO— O iPr O
Y = 0.3962
YD = 0.4870
DIRECTION DE DÉVELOPPEMENT B
DIRECTION DE DÉVELOPPEMENT BP
Me —CO— O Pr t
Y = 0.4876
Me —CO— O CH(Me)Pr O
Y = 0.4902
2 DIRECTIONS DE DÉVELOPPEMENT
O Et —CO— O Pr O
Y = 0.4878
YD = 0.4710
DIRECTION DE DÉVELOPPEMENT B
DIRECTION DE DÉVELOPPEMENT BP
Me —CO— O tAm O
Y = 0.4736
2 DIRECTIONS DE DÉVELOPPEMENT
O tBu —CO— O Et O
Y = 0.4737
O Et —CO— O tBu O
Y = 0.4738

Fig. 5. Exemples d'identification d'esters aliphatiques saturées. Température: 210°.

TABLEAU II

IDENTIFICATION DE COMPOSÉS ALIPHATIQUES SATURÉS. INFLUENCE DE LA TEMPERATURE SUR LE NOMBRE DE SOLUTIONS

Échantillon	Solutions proposées			
	210°	180°	160°	140°
Me-CO-Bu	Me-CO-Bu	Me-CO-Bu	Me-CO-Bu Me-CO-Bu (<i>t</i> Bu)	Me-CO-Bu Me-CO-Bu (<i>t</i> Bu) Et-CO-Pr
<i>t</i> Bu-CO- <i>i</i> Pr	<i>t</i> Bu-CO- <i>i</i> Pr Me-CO-CH (Et) ₂	<i>t</i> Bu-CO- <i>i</i> Pr	<i>t</i> Bu-CO- <i>i</i> Pr	<i>t</i> Bu-CO- <i>i</i> Pr
Pr-CO-Bu	Pr-CO-Bu Pr-CO- <i>t</i> Am	Pr-CO-Bu	Pr-CO-Bu <i>s</i> Bu-CO- <i>s</i> Bu	Pr-CO-Bu <i>s</i> Bu-CO- <i>s</i> Bu
Et-CO-O- <i>t</i> Bu	Et-COO- <i>t</i> Bu Me-COO- <i>t</i> Am <i>t</i> Bu-CO-O-Et	Et-CO-O- <i>t</i> Bu Me-CO-O- <i>t</i> Am <i>t</i> Bu-CO-O-Et Pr-CO-O-Et	Et-CO-O- <i>t</i> Bu Me-CO-O- <i>t</i> Am <i>t</i> Bu-CO-O-Et Pr-CO-O-Et	— Me-CO-O- <i>t</i> Am <i>t</i> Bu-CO-O-Et

cétones et les esters plusieurs solutions peuvent être obtenues pour une même valeur Y_D , ce que confirme d'ailleurs l'expérience. Mais généralement les solutions ne se retrouvent pas toutes aux autres températures, ce qui facilite l'identification. Ceci est illustré dans le Tableau II pour des échantillons préalablement identifiés.

Il nous a été possible d'obtenir, pour les 400 valeurs de grandeurs de rétention dont nous disposions, des tests d'identification excellents qui nous assurent de la validité du programme.

CONCLUSIONS

La méthode utilisée se révèle très rapide et peu coûteuse. En effet le temps de calcul en mémoire centrale nécessaire à l'identification de 45 esters (62 sec) est à peine supérieur au temps de compilation du programme.

D'autre part, elle est susceptible de généralisations intéressantes. En particulier, un traitement analogue peut s'appliquer aux composés à plusieurs directions de développement par l'introduction de plusieurs matrices d'interaction 12/12. En outre, cette méthode peut se généraliser à l'identification de composés dont la chaîne carbonée présente des ramifications dans des segments d'environnement extérieurs à l' E^1_B . Enfin, elle permet d'envisager l'identification de certains composés ne figurant pas dans la population ayant servi à l'établissement du modèle mathématique.

RÉSUMÉ

Des relations de topologie-information entre la structure de composés chimiques et leur temps de rétention sont utilisées pour identifier les chaînes carbonées R et R' de composés aliphatiques saturés de formule générale RX ou R-X-R'. Un programme général écrit en Algol 60 permet de proposer une ou plusieurs structures pour un composé dont on connaît la famille chimique et le temps de rétention.

BIBLIOGRAPHIE

- 1 M. Chastrette, G. Lenfant, A. Remy et M. Cohen-Makabeh, *J. Chromatogr.*, 84 (1973) 275.
- 2 M. Chastrette, G. Lenfant et J. E. Dubois, *C.R. Acad. Sci., Sér. C*, 265 (1967) 602.
- 3 M. Chastrette et G. Lenfant, *C.R. Acad. Sci., Sér. C*, 271 (1970) 79.
- 4 G. Lenfant, M. Chastrette et J. E. Dubois, *J. Chromatogr. Sci.*, 9 (1971) 220.
- 5 M. Chastrette et G. Lenfant, *J. Chromatogr.*, 68 (1972) 19.
- 6 M. Chastrette et G. Lenfant, *J. Chromatogr.*, 77 (1973) 255.
- 7 J. E. Dubois, D. Laurent et H. Viillard, *C.R. Acad. Sci., Sér. C*, 264 (1967) 1019.
- 8 J. E. Dubois et H. Viillard, *Bull. Soc. Chim. Fr.*, (1968) 900, 905 et 913.
- 9 J. E. Dubois et D. Laurent, *C.R. Acad. Sci., Sér. A*, 268 (1969) 405.
- 10 J. E. Dubois et D. Laurent, *Bull. Soc. Chim. Fr.*, (1969) 2449.