

ORDINATEUR ET SYNTHÈSE ORGANIQUE
UTILISATION DES MÉCANISMES RÉACTIONNELS

R. BARONE *, M. CHANON, J. METZGER

I.P.S.O.I. - Université d'Aix-Marseille III - Rue Henri Poincaré - Marseille - France.

(Received in France 11 June 1974; received in UK for publication 24 June 1974)

Un précédent travail (1) nous a permis de résoudre quelques uns des problèmes posés par l'utilisation de l'ordinateur en synthèse organique. Rappelons les brièvement :

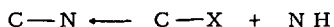
1) Stratégie de synthèse : La stratégie développée est celle proposée par E. J. COREY (2).

Partant du produit à synthétiser (la cible), on remonte étape par étape, dans un sens inverse à celui de la synthèse telle qu'elle est menée au laboratoire, jusqu'aux produits de départ.

2) Représentation des molécules : Les molécules sont décrites par une table de connectivité analogue à celles citées dans la littérature (3a). Un sous-programme permet de les dessiner à partir de la table de connectivité et des coordonnées des atomes.

3) Représentation des réactions : Une réaction se décrit par une relation entre deux membres.

Par exemple la formation d'amine peut se résumer au schéma suivant :



On doit tout d'abord rechercher un motif dans la cible (ici le motif C—N), puis effectuer des transformations sur la table de connectivité afin d'obtenir les précurseurs (ici C—N est remplacé par C—X + NH). Les motifs sont représentés par des tables de connectivité semblables à celles qui décrivent les molécules. A chaque motif on associe une ou plusieurs transformations. Un sous-programme permet de retrouver les motifs en effectuant une comparaison atome par atome (3b), un autre permet d'apporter les modifications nécessaires.

Bien que les résultats obtenus soient encourageants, deux limitations apparaissent. La première réside dans le fait que les motifs sont codés sur la base des fonctions (formation de la fonction alcool, acide, amine, etc...), ce qui entraîne l'écriture d'un nombre important de motifs pour couvrir l'ensemble de la chimie organique, de plus, un oubli est toujours possible. La seconde limitation est une conséquence de la première : en ne codant que les réactions qui sont connues, l'ordinateur ne pourra jamais proposer de nouvelles réactions, ce qui est un des objectifs de cette étude.

Pour parvenir à un tel résultat : élaboration automatique de nouveaux procédés de

synthèse, il semble qu'il faille utiliser des méthodes plus souples, plus générales, qui laissent une certaine "initiative" à l'ordinateur. C'est à quoi tendent certains travaux, comme ceux d'HENDRICKSON (4), ou ceux d'UGI et GILLESPIE (5). Aborder le problème de synthèse au moyen des fonctions chimiques correspond au découpage de la chimie organique tel qu'on l'envisageait il y a quelques années. Depuis, la connaissance des mécanismes a permis un regroupement cohérent des réactions, aussi nous a-t-il paru raisonnable de suivre cette évolution et de coder les réactions non plus sur la base des fonctions, mais sur celle des mécanismes.

Par exemple, les réactions de substitution nucléophile seront représentées :



où Y = centre nucléophile et L = groupement libérable (X, OH, ...). Les motifs font intervenir des centres nucléophiles et électrophiles qui ne sont pas directement accessibles dans la molécule telle qu'elle est codée. Aussi, le premier travail du programme consiste à rechercher dans la cible, les centres nucléophiles, électrophiles, les cycles (aromatiques et non aromatiques), et à les sauvegarder sous forme binaire. Ces informations seront utilisées lors de la recherche des motifs. Le programme chargé de la perception de ces caractéristiques utilise des techniques proches de celles citées par COREY (6).

Actuellement, les centres nucléophiles codés sont : les hétéroatomes, $\underline{C} \equiv N$, $\underline{C} - Z$ (Z = groupe électron - attracteur), comme centre électrophile nous avons retenu uniquement l'atome d'hydrogène. Enfin, les mécanismes codés sont : substitution nucléophile sur carbone saturé et insaturé, élimination, addition nucléophile et électrophile sur liaison multiple. Nous y avons adjoint certaines réactions d'oxydation et de réduction et les équilibres du type céto-énolique. Les réactions codées ici ne sont pour l'instant qu'une première approche très limitée, dont l'utilité est surtout de rédiger le programme qui sera ensuite généralisé aux autres mécanismes.

Le programme, tel qu'il est conçu, propose des solutions erronées car une réaction est représentée par un motif isolé sans tenir compte de l'influence exercée par le reste de la molécule. Une étude est en cours afin de coupler la recherche des motifs avec une discrimination chimique capable de tenir compte de l'ensemble de la molécule. De même, il ne permet pas l'élimination automatique des précurseurs identiques. Aussi, avons-nous effectué un découpage pour l'exemple traité, afin de ne présenter que des solutions dignes d'intérêt. Le tableau I montre les résultats obtenus pour la synthèse du triméthyl 2, 4, 5 thiazole. Les principales voies de synthèse ont été retrouvées (7 - 12).

Remarques : 1) Si nous n'avons plus à coder tous les motifs possibles, il faut pouvoir retrouver tous les centres nucléophiles et électrophiles, cependant ce problème est moins vaste.

2) Cette approche est extrêmement générale, et permettra de couvrir à l'aide d'un petit nombre de schémas, un très grand nombre de réactions. Cependant certaines

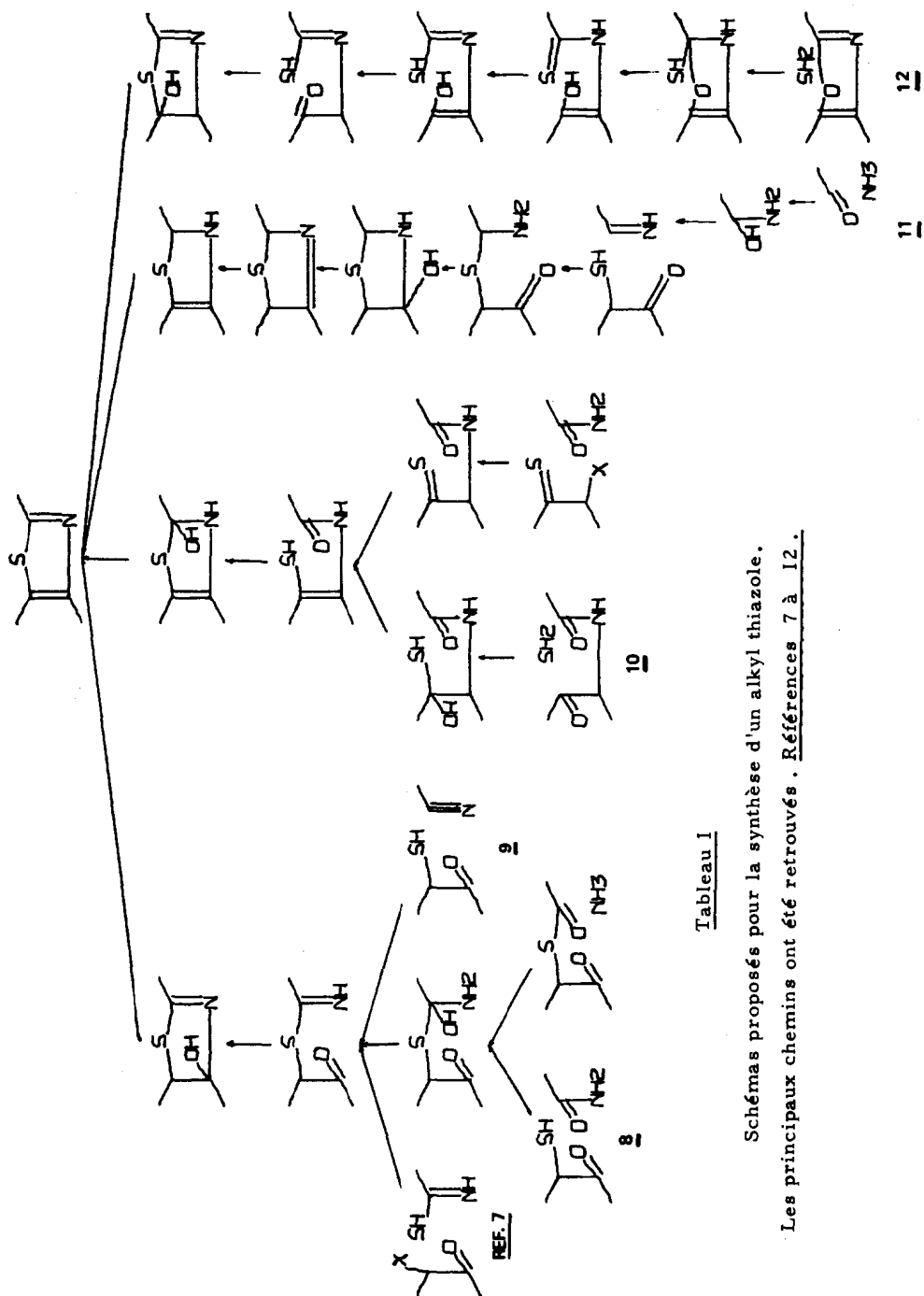


Tableau I

Schémas proposés pour la synthèse d'un alkyl thiazole.

Les principaux chemins ont été retrouvés. Références 7 à 12.

réactions ne seront pas représentées car leur mécanisme n'est pas très bien connu, ou ne répond pas à un schéma général clairement défini, comme par exemple certaines réactions d'oxydation et de réduction, certaines transpositions ; aussi est-il nécessaire de conserver, pour ces cas particuliers, une représentation sur le principe des fonctions.

En conclusion, nous pouvons dire que cette méthode est très prometteuse puisqu'avec un nombre réduit de motifs, le programme propose déjà des chemins de synthèse corrects. De plus, en faisant abstraction de la nature des atomes, il est raisonnable d'espérer que l'ordinateur sera capable de proposer des schémas originaux qui pourront être la source de nouvelles méthodes de synthèse.

REFERENCES

- (1) R. BARONE, M. CHANON, J. METZGER, *Rev. Inst. Français Pétrole*, 1973, 5, 771.
- (2) E.J. COREY, W.T. WIPKE, *Science*, 1969, 166, 178.
- (3) M.F. LYNCH, J.D. HARRISON, W.G. TOWN, J.A. ASH, *Computer Handling of Chemical Structure Information*, Macdonald London and Elsevier, Inc. New-York (a) page 14, (b) page 73.
- (4) J.B. HENDRICKSON, *J. am. chem. Soc.*, 1973, 93, 6847 et 6854.
- (5) I. UGI, P. GILLESPIE, *Angew. Chem. Intern. Ed. Engl.* 1971, 10, 914 et 915.
- (6) E.J. COREY, W.T. WIPKE, R.D. CRAMER III, W.J. HOWE, *J. am. chem. Soc.*, 1972, 94, 431.
- (7) J.M. SPRAGUE, A.H. LAND, *Elderfield - Heterocyclic Compounds*, tome V, Ed. Wiley, 1957, page 496.
- (8) M. OHTA, K.M. SATO, *J. Pharm. Soc. Japan* 1947, 67, 227.
- (9) K. MIYATAKE, S. KOJIMA, G. OHTA, G. OUCHI, *J. Pharm. Soc. Japan*, 1955, 75, 1054.
- (10) J.M. SPRAGUE, A.H. LAND, *Elderfield - Heterocyclic Compounds*, tome V, Ed. Wiley, 1957, page 503.
- (11) F. ASINGER, M. THIEL, G. PESHEL, *Ann. Chem.*, 1958, 619, 145.
- (12) G.H. KEYES, L.G.S. BROOKER, U.S. Patent, 2.652.396 (1953), *Chem. Abstracts* : 1954 - 48 - 486 b.