

Nouveau service de diffusion sélective de l'AFDAC, en 1977

En 1977, l'AFDAC exploitera les bandes Chemical Abstracts Subject Index Alert (CASIA), fournies par Chemical Abstracts Service (CAS). L'interrogation portera sur les composés définis et sur les familles de composés, au moyen de la nomenclature, des numéros de registre, des formules moléculaires.

L'abonnement à une question coûtera environ 1 000 F par an.

Dans le cadre d'un groupe de travail spécialisé de l'Union des Industries Chimiques, il a été décidé fin 1975 de mener une expérience en vue de comparer, en diffusion sélective (DSI), les performances des bandes Chemical Abstracts Subject Index Alert (CASIA) et des bandes Chemical Abstracts Condensates (CACon) fournies par Chemical Abstracts Service (CAS) : rappel, pertinence, délais de signalement, coûts de traitement.

Cette étude visait à une connaissance approfondie du fichier CASIA, déjà accessible expérimentalement on-line chez Lockheed, et au développement d'un nouveau service de DSI en fonction de l'apport spécifique de ce fichier.

I. Couverture des bandes

Les bandes CASIA et les bandes CACon signalent tous les documents résumés dans les fascicules imprimés. Leur organisation est séquentielle, dans l'ordre des numéros d'extraits CA. Les bandes CACon sont hebdomadaires, tandis que les bandes CASIA sont fournies tous les quinze jours; cependant, pour certains documents, le signalement dans CASIA est plus tardif que le signalement dans CACon : une bande CASIA ne comprend donc pas tous les numéros d'extraits des deux bandes CACon correspondantes, par contre elle comprend un certain nombre de numéros d'extraits antérieurs.

II. Contenu des bandes

Les bandes CACon comportent, pour chaque extrait : le numéro d'extrait CA, le titre, les auteurs, la référence bibliographique, l'organisme, la section CA et la sous-section, la langue, le type de document, les groupes de mots-clés figurant dans les fascicules hebdomadaires, et, pour les brevets, le pays de dépôt, le numéro, les classes, les dates de priorité. Le titre et les groupes de mots-clés sont rédigés en langage naturel.

Les bandes CASIA comportent, pour chaque extrait : le numéro d'extrait, la section CA et la sous-section ainsi que les signalements des documents qui figureront dans les index semestriels, « Chemical Substance Index » et « General Subject Index ». Ces signalements sont rédigés en langage contrôlé sans qu'il soit fait véritablement appel à un thésaurus; les composés chimiques sont décrits d'après les règles de nomenclature fixées par CAS; les concepts généraux ne peuvent être choisis librement par l'indexeur, mais doivent être sélectionnés parmi une liste de rubriques prédéfinies. L'utilisateur peut acquérir la connaissance

de ce langage contrôlé au moyen de l'Index Guide, des index semestriels cités ci-dessus, et des « CASIA Search Aids » fournis par CAS.

Les signalements des documents dans les bandes CASIA peuvent être décrits schématiquement de la façon suivante :

- Les composés chimiques sont indexés par le « CA preferred index name » établi suivant la nomenclature CAS : « heading parent » éventuellement suivi d'un « substituent »; la nomenclature est complétée par le numéro de registre et la formule moléculaire du composé.

L'ensemble est suivi de « text modifications » établies en langage naturel.

- Les concepts généraux sont décrits par les « concept headings » du langage contrôlé, suivis des « text modifications » en langage naturel.

III. Résultats de l'étude

L'étude a consisté à comparer une quarantaine de profils sur les deux fichiers, pendant une période de trois mois. Elle sera publiée en 1977 dans « Journal of Chemical Information and Computer Sciences ». Ses conclusions peuvent être résumées ainsi :

1. Délais de signalement

88 % des références apparaissent en même temps dans CASIA et dans CACon; 7 % apparaissent dans CASIA avec un retard compris entre 1 et 2 mois; 2 % avec un retard compris entre 2 et 3 mois; 3 % avec un retard supérieur. Peu à peu CAS rattrape ce retard : en 1978 les signalements dans CASIA et dans CACon devraient être entièrement synchronisés.

2. Pertinence et rappel

Les taux de pertinence et de rappel sont excellents sur CASIA ($\geq 84\%$) et très supérieurs à ceux obtenus sur CACon, pour les questions concernant des composés définis ou des familles de composés, associés ou non à des concepts.

L'apport spécifique du fichier CASIA par rapport au fichier CACon est dû à l'indexation systématique, pour les index semestriels, des composés définis décrits dans les documents primaires, au moyen de la nomenclature CAS. L'utilisateur, s'il possède une connaissance suffisante des règles de nomenclature, peut rechercher les composés définis, les familles de composés et les structures contenant une sous-structure donnée.

3. Coûts de traitement

Le coût de conversion d'une bande CASIA, dans un format compatible avec le programme de lecture de texte PRETEXT II, est supérieur de 60 % au coût de conversion de deux bandes CACon.

Le coût d'exploitation des profils sur une bande CASIA est supérieur de 22 % au coût d'exploitation sur deux bandes CACon. Ces différences proviennent principalement du plus grand nombre de références signalées sur CASIA pendant la durée de l'expé-

N° 465. — Synthèses de thiéno[2,3-d]pyrimidines substituées en 2 et 4 II,

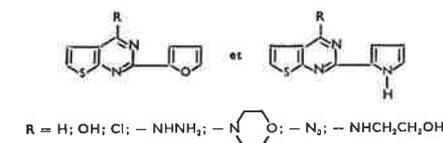
par J. BOURGUIGNON, M. MOREAU, G. QUÉGUINER et P. PASTOUR.

(Laboratoire de chimie organique
de l'Institut National Supérieure de Chimie Industrielle de Rouen,
76130 Mont Saint-Aignan.)
(Manuscrit reçu le 6.5.75.)

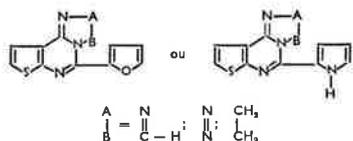
Nous avons synthétisé deux nouvelles séries de thiéno[2,3-d]pyrimidines possédant en 2 les groupes furyle-2 ou pyrrolyle-2, en 4 les groupes hydroxyle ou chloro ou hydrazino ou morpholino ou azido ou hydroxy-2 éthylamino.
Certains de ces substituants nous ont permis d'aboutir à des systèmes hétérocycliques comportant trois noyaux condensés : s-triazolo, tétrazolo et dihydroimidazo thiéno[3,2-e]pyrimidines.

De nombreux composés renfermant la structure thiéno[2,3-d]pyrimidine sont connus. Pour notre part, nous avons précédemment synthétisé quatre séries de thiéno[2,3-d]pyrimidines possédant en 2 les substituants pyrrolyle-2, thiényle-2, phényle ou méthyle (1).

Nous décrivons les préparations de deux nouvelles séries de composés du même type substitués en 2 par les noyaux furannique ou pyrrolique et en 4 par les groupes hydroxyle, chloro, hydrazino, morpholino, azido et hydroxy-2 éthylamino.



Certains de ces composés sont cyclisés en s-triazolo, tétrazolo et dihydroimidazo thiéno[3,2-e]pyrimidines.

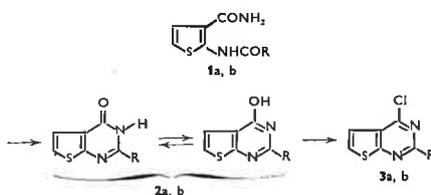


I. — Synthèses de composés disubstitués en 2 et 4.

Remarque. — Dans tout ce qui suit les lettres a et b correspondent respectivement aux produits substitués en 2 par les groupes furyle-2 et pyrrolyle-2.

L'acylation, en présence d'une base, de l'amino-2 thénamide-3 (2) nous a permis d'obtenir les composés 1 dont la cyclisation par la soude à 10 % mène aux oxo-4 dihydro-3,4 thiéno[2,3-d]pyrimidines 2. Le traitement de ces composés par l'oxychlorure de phosphore conduit aux dérivés chlorés 3.

Lorsque R est le groupe furyle-2, les synthèses ressemblent à celles des composés où R est le groupe thié-

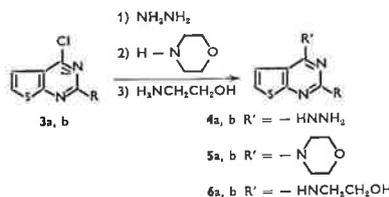


nyle-2. Par contre, dans le cas des dérivés pyrroliques quelques préparations sont délicates car les produits de réaction sont instables ou difficiles à isoler.

C'est ainsi que dans le cas de la synthèse du (pyrrolyl-2) amino-2 thénamide-3 1b nous avons été obligés d'introduire de manière simultanée le chlorure de l'acide pyrrole carboxylique-2 et la base dans l'amino-2 thénamide-3. Si l'on opère de manière usuelle, en chauffant le mélange des réactifs et de la base, on obtient des goudrons inexploitable.

De la même manière, la chloro-4 (pyrrolyl-2)-2 thiéno[2,3-d]pyrimidine 3b ne peut être obtenue que si la réaction de chloration de l'oxo-4 dihydro-3,4 thiéno[2,3-d]pyrimidine 2b est réalisée très rapidement.

L'attaque des composés chlorés 3 par divers nucléophiles azotés permet de préparer les produits 4, 5 et 6.



Les hydrazines 4 donnent accès aux thiéno[2,3-d]pyrimidines 7. Une oxydation classique par l'oxyde mercurique jaune a permis d'aboutir au dérivé furannique 7a alors que nous avons été obligés de traiter 4b par l'oxygène en présence d'éthylate de sodium pour obtenir le dérivé pyrrolique 7b.

rience, et de la taille de l'enregistrement de chaque signalement : en moyenne 960 caractères pour CASIA, 540 caractères pour CACon.

IV. Service proposé pour 1977 par l'AFDAC

Un service régulier de diffusion sélective portant seulement sur les bandes CASIA, et impliquant l'abandon du traitement des bandes CACon, n'est pas envisageable : d'abord les références bibliographiques, ne figurant pas sur les bandes CASIA, ne peuvent être éditées sur les sorties ; ensuite, les questions portant sur des concepts seulement semblent donner des résultats nettement moins satisfaisants que sur les bandes CACon. Par ailleurs, le traitement simultané de la totalité des bandes CASIA et CACon serait évidemment très coûteux, puisque le coût de traitement d'une bande CASIA est nettement supérieur au coût de traitement de deux bandes CACon.

C'est pourquoi l'AFDAC propose de développer en 1977 un service expérimental de diffusion sélective sur une partie des bandes CASIA, tout en poursuivant le service opérationnel sur les bandes CACon. Elle limitera le contenu des bandes CASIA en choisissant :

- Les sections interrogeables : une liste (en annexe) est proposée ; cette liste est encore susceptible de modifications, selon les réactions des utilisateurs.
- Les informations interrogeables : ne seront conservées sur les bandes que les données relatives aux structures : « heading parents », « substituents », « text modifications », numéros de registre, formules moléculaires.

L'interrogation pourra donc porter sur les numéros de sections CA, la nomenclature, les numéros de registre, les formules moléculaires et les « text modifications ».

Le programme de lecture de texte PRETEXT II sera utilisé, permettant les troncatures droite et gauche, les logiques booléennes, les logiques syntaxiques AVEC et IGNORE particulièrement intéressantes pour l'interrogation de la nomenclature.

Les sorties comporteront probablement les numéros d'extraits CA, les sections, et les entrées des index semestriels relatives aux composés : nomenclature, numéros de registre, formules moléculaires, « text modifications ».

Extrait CACon

**CAC *084(021)*150586U*S:028* BSCFAS*
 ρSYNTHESES OF 2- AND 4-SUBSTITUTED THIENO(2,3-D)PYRIMIDINES. II.
 BOURGUIGNON, J.; MOREAU, M.; QUÉGUINER, G.; PASTOUR, P. (LAB. CHIM. ORG. INST. NATL. SUPER. CHIM. IND. ROUEN) (MONT ST.-AIGNAN FR.).
 BULL. SOC. CHIM. FR. (PUB. 000075) (11-12, PT. 2) 248 3-7 (FR)
 ρTHIENOPYRIMIDINE ρTHIOPHENECARBOXAMIDE ACYLAMINO CYCLIZATION ρTHIENOTRIAZOLOPYRIMIDINE ρIMIDAZOTHIENOPYRIMIDINE ρTETRAZOLOTHIENOPYRIMIDINE

PROFIL : CA1 114

CARTE : 002

Extrait CASIA

**CA *084(021)*150586U*S:028* CS7611*
 ρ/HP=2-FURANCARBOXYLIC ACID/RN=000088142/MF=C5H4O3/TM=CHLORINATION OF.
 ***/CH=RING CLOSURE AND FORMATION/TM=OF ACYLAMINOTH IOPHENECARBOXAMIDES ***/HP=TETRAZOLO(1,5-C)THIENO(3,2-E)PYRIMIDINE/SB=5-(1H-PYRROL-2-YL)-/RN=058980099/MF=C10H6N6S ***/HP=TETRAZOLO(1,5-C)THIENO(3,2-E)PYRIMIDINE/SB=5-(2-FURANYL)-/RN=058980088/MF=C10H5N5OS ***/HP=1 MIDAZO(1,2-C)THIENO(3,2-E)PYRIMIDINE/SB=5-(1H-PYRROL-2-YL)-/RN=058980077/MF=C12H10N4S ***/HP=IMIDAZO(1,2-C)THIENO(3,2-E)PYRIMIDINE/SB=5-(2-FURANYL)-/RN=058980066/MF=C12H9N3OS ***/HP=THIENO(3,2-E)-1,2,4-TRIAZOLO(4,3-C)PYRIMIDINE/SB=5-(1H-PYRROL-2-YL)-/RN=058980055/MF=C11H7N5S ***/HP=THIENO(3,2-E)-1,2,4-TRIAZOLO(4,3-C)PYRIMIDINE/SB=5-(2-FURANYL)-/RN=058980044/MF=C11H6N4OS ***/HP=METHANE/SB=TRIMETHOXY-/RN=000149735/MF=C4H10O3/TM=CYCLIZATION OF, WITH HYDRAZINOTHIENOPYRIMIDINES ***/HP=THIENO(2,3-D)PYRIMIDINE/SB=2-(1H-PYRROL-2-YL)-/RN=058980033/MF=C10H7N3S ***/HP=THIENO(2,3-D)PYRIMIDINE/SB=2- CARTE : 012

PROFIL : CSA 114

Le prix annuel d'un profil CASIA sera calculé d'après la formule suivante :

$300 + 26n + 0,30r$ (en francs)
 n étant le nombre de termes du profil,
 r le nombre de références retrouvées.

Le prix moyen annuel d'un profil sera de 1 000 F, incluant le coût de formulation et le coût de traitement informatique.

L'équipe technique de l'AFDAC se tient à la disposition des lecteurs de *L'actualité chimique* pour donner tous renseignements complémentaires et pour coder leurs profils CASIA. Envoyez rapidement commentaires et sujets à l'AFDAC. Le service est opérationnel depuis janvier 1977.

Annexe

Liste des sections interrogeables sur CASIA en 1977 à l'AFDAC

- 1 Pharmacodynamics
- 2 Hormone Pharmacology
- 3 Biochemical Interactions
- 4 Toxicology
- 5 Agrochemicals
- 17 Foods
- 22 Physical Organic Chemistry
- 23 Aliphatic Compounds
- 24 Alicyclic Compounds
- 25 Noncondensed Aromatic Compounds
- 26 Condensed Aromatic Compounds
- 27 Heterocyclic Compounds (one Hetero Atom)

- 28 Heterocyclic Compounds (more than one Hetero Atom)
- 29 Organometallic and Organometalloidal Compounds
- 30 Terpenoids
- 31 Alkaloids
- 32 Steroids
- 33 Carbohydrates
- 34 Synthesis of Amino Acids, Peptides, and Proteins
- 59 Air Pollution and Industrial Hygiene
- 60 Sewage and Wastes
- 61 Water
- 62 Essential Oils and Cosmetics
- 63 Pharmaceuticals
- 64 Pharmaceutical Analysis
- 78 Inorganic Chemicals and Reactions
- 79 Inorganic Analytical Chemistry
- 80 Organic Analytical Chemistry