

THESIS / THÈSE

MASTER EN SCIENCES INFORMATIQUES

Conception d'une base de données graphique en chimie organique

Bovesse, Michel

Award date:
1985

Awarding institution:
Universite de Namur

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

FACULTES UNIVERSITAIRES NOTRE-DAME DE LA PAIX

NAMUR

INSTITUT D'INFORMATIQUE

**CONCEPTION D'UNE BASE DE DONNEES GRAPHIQUE
EN CHIMIE ORGANIQUE**

Michel BOVESSE

*Mémoire présenté en vue
de l'obtention du grade*

de

Licencié et Maître en Informatique

Année académique 1984-1985

*Ce travail a été encadré et dirigé par
Madame Monique Noirhomme ;*

*L'aspect chimique a été orienté par
Monsieur Alain Krief ;*

*La critique de l'analyse conceptuelle a
été assurée par
Monsieur Jean-Luc Hainaut ;*

*Enfin, la dactylographie est due à
Madame Christiane Coppée.*

Je les remercie tous pour leur aide efficace.

P L A N

<u>CHAPITRE I : INTRODUCTION</u>	1
<u>CHAPITRE II : APERCU DES SYSTEMES EXISTANTS</u>	3
1. Systèmes-experts	3
1.1. Aspects généraux des Systèmes-experts en synthèse organique	3
1.2. Système-expert de COREY : LHASA	10
1.3. Autres Systèmes-experts	15
1.4. Critique des Systèmes-experts en synthèse organique	17
2. Base de Données	19
2.1. But	19
2.2. Systèmes existants	19
2.3. Remarques à propos des systèmes analogiques existants	24
<u>CHAPITRE III : SPECIFICATIONS GENERALES D'UN OUTIL DE GESTION DE BASE DE DONNEES EN CHIMIE ORGANIQUE</u>	26
1. Mise en évidence des concepts essentiels	26
2. Proposition générale de solution	30
<u>CHAPITRE IV : ASPECTS GRAPHIQUES LIES A LA SOLUTION</u>	37
1. Géométries atomiques	37
2. Mise en évidence de quelques problèmes	39
3. Translation	40
4. Rotation	40
5. Modification d'échelle	44
6. Projection perspective	44
7. Accrochage de sous-structures	45

<u>CHAPITRE V : ANALYSE CONCEPTUELLE</u>	49
1. Eléments de l'analyse conceptuelle	49
2. Schéma conceptuel	58
3. Explication du schéma conceptuel	59
4. Contraintes d'intégrité	72
5. Spécification des fonctions de consultation	74
<u>CHAPITRE VI : CONCEPTION LOGIQUE DE LA BASE DE DONNEES</u>	77
1. But de cette étape	77
2. Schéma MAG dérivé du schéma conceptuel	78
3. Enoncé des requêtes	79
4. Spécification des fonctions utilisées par les requêtes	79
5. Algorithmes des requêtes et des fonctions d'accès	81
6. Spécification des fonctions d'introduction des données	85
<u>CHAPITRE VII : CONCLUSIONS</u>	91
<u>ANNEXE1: Description associée au schéma MAG</u>	95
<u>ANNEXE2: Architecture du système</u>	101
<u>BIBLIOGRAPHIE</u>	108

CHAPITRE I
INTRODUCTION

1. INFORMATISER LA SYNTHÈSE ORGANIQUE ?

On peut se demander quel serait l'apport de l'informatique à la synthèse organique.

Il faut envisager cet apport sous deux aspects :

a) Quantitatif :

La synthèse organique est une science récente mais très productive ; à titre d'exemple, les 24 volumes de "Synthetic Methods of Organic Chemistry" de Theilleimer répertorient pas moins de 22.000 réactions chimiques [21] et l'augmentation annuelle du nombre de références était estimé à 2000 en 1970. Cette masse d'information doit être gérée afin de pouvoir être consultée en un minimum de temps par le chimiste ; ces données font apparaître le besoin d'automatisation de la bibliographie organique.

b) Qualitatif :

Les outils informatiques envisageables pour aider le chimiste organicien vont du simple système de fichiers lui permettant de gérer ses références bibliographiques et de les consulter via des mots-clés [17], au Système-Expert [1] proposant au chimiste des schémas réactionnels originaux pour une molécule-cible choisie (par "molécule-cible", nous désignons toute molécule à synthétiser).

Entre ces deux extrêmes, il est permis d'envisager des systèmes plus souples, facilitant le travail du chimiste en lui proposant des méthodes de transformation répertoriées pour des molécules analogues à la molécule-cible. Ce travail met en évidence les problèmes liés à un tel système et aborde la conception logique d'une base de données en synthèse organique.

2. PLAN DU TRAVAIL

Dans le Chapitre II, nous présentons succinctement les différents systèmes informatiques - opérationnels ou non - en synthèse organique ; l'accent sera mis sur les domaines d'application, les avantages et les inconvénients de chacun de ces systèmes.

Le Chapitre III introduit les différents aspects d'une base de données en synthèse organique :

- présentation des concepts de base de la synthèse organique ;
- proposition d'une solution "graphique".

Le Chapitre IV aborde les principes généraux liés au traitement graphique des molécules : translation, rotation, modification d'échelle, projection et connexion de structures.

Les Chapitres V et VI s'intéressent à la conception de la base de données et des fonctions d'accès ; les traitements tels que : "affichage de graphes", "saisie de graphes" ne sont pas abordés dans ce travail.

Dans le Chapitre V, le schéma conceptuel d'une réaction chimique est proposé et les requêtes sont spécifiées.

Le Chapitre VI propose un schéma de base de données de type MAG ; une spécification précise ainsi qu'un algorithme ADL est donné pour chacune des fonctions d'accès.

CHAPITRE II

APERCU DES SYSTEMES EXISTANTS

1. SYSTEMES-EXPERTS

1.1. Aspects généraux des systèmes-experts en synthèse organique

1.1.1. Méthodes de raisonnement en synthèse organique

Comment raisonne le chimiste dans la recherche d'une voie de synthèse optimale ? A chaque étape, en effet, il a le choix parmi une multitude de réactions "possibles", conduisant chacune à un composé intermédiaire différent. "L'arbre" de synthèse est donc énorme !

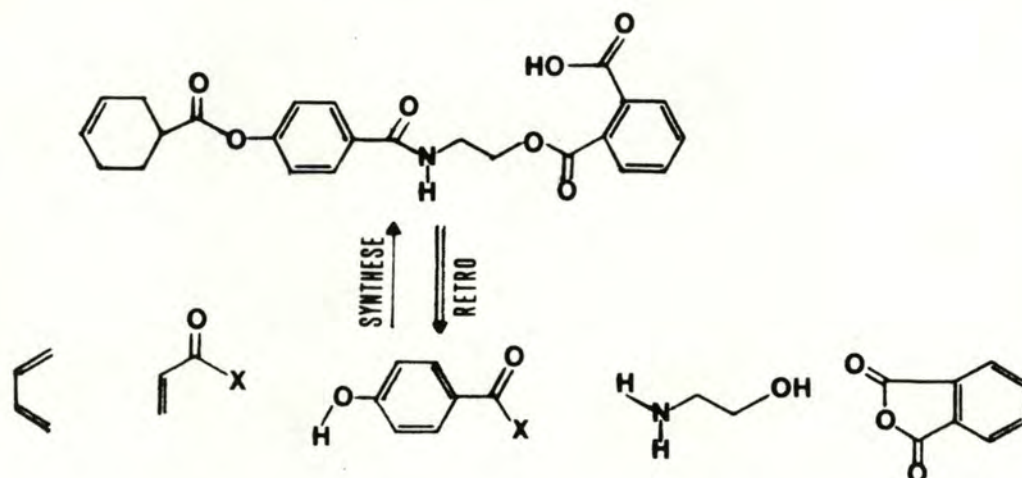
A titre d'exemple, un stéroïde - composé d'une vingtaine d'atomes autres que l'Hydrogène - pourrait être l'objet de 2^{18} voies de synthèse ! Comme le fait remarquer E.J. COREY [1], "curieusement, cet aspect du problème n'a jamais été traité de façon systématique dans la littérature chimique". En fait, le chimiste de synthèse utilise des techniques de raisonnement plus ou moins sophistiquées et intuitives en fonction du composé à synthétiser.

COREY distingue trois types de raisonnement en chimie organique [1] :

1.1.1.1. Raisonnement associatif

Cette méthode consiste à distinguer, dans le composé à synthétiser, un certain nombre de sous-structures qui peuvent être "aisément" enchaînées les unes aux autres.

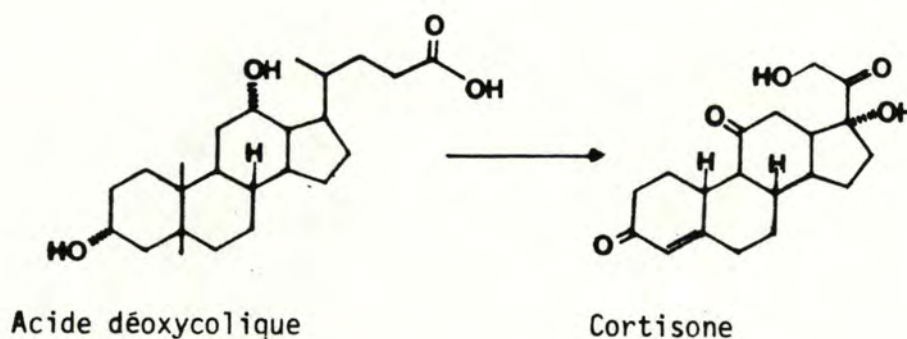
Ex.



Cette méthode ne peut s'appliquer qu'à des composés pour lesquels la perception des sous-structures de base est immédiate.

1.1.1.2. Raisonnement par "similitude"

Ce type de raisonnement consiste à repérer, dans le composé-cible, une structure essentielle et critique, et à établir la relation entre cette structure et un composé connu. Ainsi, la première synthèse de la cortisone partait d'un composé naturel connu et de structure similaire : l'acide déoxycolique.



Malgré les similitudes de structures entre les composés de départ et d'arrivée, il ne faut pas moins d'une trentaine d'étapes pour réaliser la transformation ! On a imaginé depuis des méthodes de synthèse de la cortisone beaucoup plus élégantes qui partent de composés beaucoup plus simples.

1.1.1.3. Méthode logique

Dans cette approche, le chimiste doit "analyser" la structure de la molécule-cible et rechercher les précurseurs susceptibles de conduire, en une seule étape, à cette molécule-cible.

Chaque précurseur devient alors la nouvelle molécule-cible. L'analyse de la molécule-cible se fait en plusieurs étapes :

- a) Il faut tout d'abord repérer, dans la molécule, les aspects structuraux qui sont significatifs :
 - les chaînes, les cycles, les ramifications ;
 - les groupes fonctionnels ;
 - les carbones asymétriques ;
 - les points sensibles et instables dans la molécule.

- b) L'étape suivante consiste à mettre en évidence les opérations susceptibles de réduire la complexité de la molécule-cible ; cette réduction de la complexité sera une combinaison des effets suivants :
 - simplification de la connectivité interne de la molécule par ouverture de cycles ;
 - réduction de la taille de la molécule par déconnexion de chaînes ou de ramifications ;
 - suppression de fonctionnalités ;
 - modification ou suppression de sites d'instabilité inhabituelle ;
 - simplification de la stéréochimie par suppression, par exemple, de carbone asymétrique.

- c) On peut également envisager des étapes qui ne réduisent pas la complexité, mais qui préparent une étape simplificatrice :
- échange de groupes fonctionnels ;
 - introduction de groupes fonctionnels ;
 - modification de la fonctionnalité pour contrôler le niveau de réactivité ;
 - introduction de groupes qui permettent le contrôle stéréochimique ;
 - réarrangements internes qui permettent de modifier des cycles, des chaînes ou des groupes fonctionnels.

1.1.1.4. Commentaires à propos de la méthode logique

1. Cette méthode de raisonnement montre que la conception d'un schéma de synthèse n'est pas une démarche évidente, puisque, comme le fait remarquer COREY [2] :
 - a) Une solution à un problème de synthèse ne peut être obtenue "simplement" par analogie avec des problèmes résolus précédemment.
 - b) Les produits de départ ne sont pas directement apparents.
 - c) Il faut considérer de nombreux schémas de synthèse possibles afin de sélectionner le plus simple et le plus efficace.
 - d) La complexité d'une molécule ne provient pas de sa taille, mais plutôt d'aspects structuraux tels que : cycles, stéréocentres, interactions entre groupes...

2. Toutes les transformations présentées ci-dessus peuvent être classées sur base des modifications qu'elles engendrent :

- a) modification du squelette de la molécule ;
- b) modification des groupes fonctionnels ;
- c) modification de la stéréochimie.

Seules les transformations du type (a) sont réellement simplificatrices ; les transformations des types (b) et (c) ne se justifient que si elles génèrent un composé susceptible d'accepter, lui, une transformation du type (a).

L'efficacité d'une transformation peut donc se mesurer par l'importance des simplifications de structure auxquelles elle conduit.

Ainsi, la transformation de DIELS-ALDER



peut s'analyser de la manière suivante dans le sens rétro-synthétique :

- a) nombre de cycles diminue ;
- b) nombre de carbones asymétriques diminue ;
- c) déconnexion, donc, augmentation du nombre de fragments ;
- d) simplification de la fonctionnalité.

3. Une stratégie consiste à appliquer de préférence les transformations les plus puissantes, c'est-à-dire les plus simplificatrices dans le sens rétro-synthétique. Lorsque la molécule-cible ne possède pas les fonctions qui permettraient d'appliquer une transformation simplificatrice, la stratégie consiste alors à définir les

fonctions souhaitées comme des sous-buts et essayer d'y parvenir en un minimum d'étapes. Ces étapes seront généralement non simplificatrices.

1.1.2. Pourquoi un Système-Expert ?

Le but d'un "système-expert" en Synthèse Organique est de trouver, pour une molécule-cible donnée, la - ou les - voies de synthèse les plus intéressantes.

Une approche possible est l'application de la "méthode logique" décrite au paragraphe précédent (1.1.1.3). A partir d'une molécule-cible, il s'agit d'explorer l'arbre de synthèse conduisant à cette molécule. Pour cela, le système-expert doit posséder :

- des règles de transformations qui, dans ce cas, sont la traduction directe des équations de transformations chimiques ;
- un "moteur d'inférence" qui lui permet d'appliquer ces règles à la molécule-cible et de parcourir l'arbre de synthèse ;
- des règles d'évaluation et des heuristiques qui lui permettent :
 - . d'abandonner les branches "naïves" de l'arbre ;
 - . de sélectionner les branches les plus intéressantes.

1.1.3. Moteur d'inférence

On peut a priori envisager quatre types d'algorithmes de parcours de l'arbre de synthèse.

a) "Forward Search"

On part de tous les produits de départ imaginables, et on parcourt l'arbre de synthèse jusqu'à obtenir le composé souhaité.

Comme le nombre de produits de départ à considérer est impressionnant, cette méthode n'est pas applicable.

b) "Backward Search"

Une autre approche consiste à parcourir l'arbre de synthèse à partir du composé-cible.

Cette méthode parcourt l'arbre de façon "aveugle", sans tenir compte des contraintes et des spécificités du problème.

c) "Backward Search" avec fonction d'évaluation

A chaque étape du parcours de l'arbre, on réévalue la voie de synthèse sur base de paramètres tels que : rendement, coût... Ceci permet d'abandonner des voies de synthèse peu prometteuses.

d) "Backward Search" avec heuristique

Le choix de la branche de l'arbre est basé sur des connaissances chimiques.

Ex. "Si la molécule terminale doit posséder une fonctionnalité fort sensible, il est logique de former cette fonction en dernier lieu".

1.2. Système-Expert de COREY : LHASA

1.2.1. Classification des rétro-réactions par COREY

Les règles de transformation utilisées par le système-expert sont en fait des rétro-réactions, c'est-à-dire des réactions considérées dans le sens opposé au sens de la synthèse organique.

COREY propose la classification suivante pour les transformations rétrosynthétiques [6] :

1. Transformations qui nécessitent deux groupes fonctionnels :

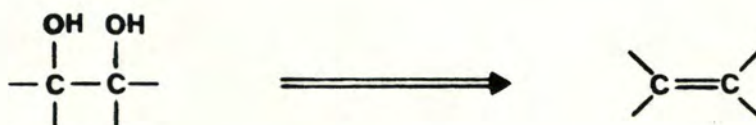
a) Déconnexion des liaisons entre les groupes :



b) Formation d'une nouvelle liaison :



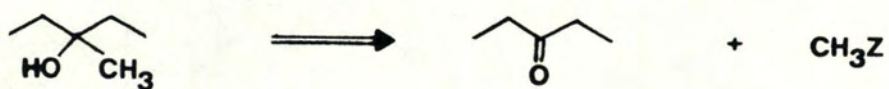
c) Modification de la fonctionnalité :



d) Réarrangements.

2. Transformations d'un seul groupe fonctionnel :

a) Déconnexion :



b) Formation d'un cycle :



c) Modifications de la fonctionnalité :

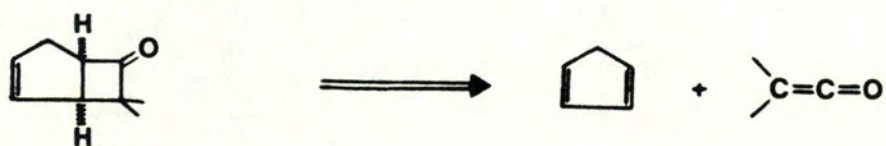


d) Réarrangement :

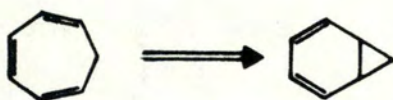


3. Transformations qui dépendent de la taille des cycles et de la stéréochimie :

a) Ouverture de cycle :



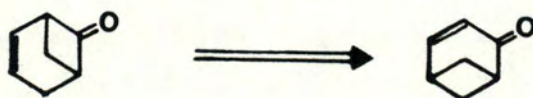
b) Formation de cycle :



c) Modification de groupes fonctionnels :



d) Réarrangements :



4. Transformations qui conduisent à l'adjonction de fonctionnalités :



Notes :

- La présence d'une double flèche (\Longrightarrow) indique que la réaction est considérée sous sa forme rétrosynthétique.
- Cette liste de classes de rétro-réactions n'est pas exhaustive.

1.2.2. Heuristique de COREY

Une heuristique est une stratégie de résolution qui simplifie la résolution d'un problème. Les heuristiques de COREY se basent essentiellement sur la notion de "liaisons stratégiques" dans les structures polycycliques pontées [3], [10] ainsi que sur les transformations de groupes fonctionnels [4], [5], [10].

1.2.3. Problème des interférences

Lorsqu'on met une molécule organique complexe en présence d'un réactif, il est rare que le réactif attaque spécifiquement tel ou tel groupe fonctionnel : si on veut assurer la spécificité de la réaction, il faut "protéger" les autres groupes fonctionnels.

Lors de l'élaboration d'un schéma de synthèse, le système doit être capable de repérer et de protéger les groupes fonctionnels susceptibles d'interférer avec la réaction désirée.

LHASA détecte les interférences probables [9] grâce à une table qui met en relation chaque réactif avec les groupes fonctionnels susceptibles d'être attaqués.

La décision de "protéger" un groupement fonctionnel lors d'une réaction n'est pas simple, car il faut prendre en considération :

- l'existence de techniques de masquage dans les conditions de réaction considérées ;
- le fait que la réaction de protection peut elle-même être compliquée par des interférences.

1.2.4. Codage et manipulation des molécules dans LHASA

Les molécules sont codées sous forme de table de connexion [8]. La reconnaissance des groupes fonctionnels d'une molécule se fait par comparaison des tables de connexion de la structure courante et de groupes fonctionnels-type. La reconnaissance des fonctions impose donc une comparaison atome par atome.

La perception des cycles dans LHASA se fait par parcours exhaustif du graphe moléculaire : la méthode consiste à choisir aléatoirement un atome de départ et à parcourir tous les chemins possibles à partir de cet atome ; il y a cycle lorsqu'on revient à un atome déjà visité.

Notons encore que LHASA utilise seulement deux tables de connexion au cours de l'analyse : la première pour "la structure-cible", la seconde pour la "structure-actuelle" ; les informations concernant les structures intermédiaires sont stockées sous forme de "blocs de modifications" par rapport à la "structure-parente" [7].

1.2.5. Interface avec l'utilisateur

LHASA fournit à l'utilisateur une interface graphique permettant l'introduction et la restitution des molécules sous forme de graphes dans lesquels les atomes de carbone et d'hydrogène ne sont pas représentés.

1.3. Autres systèmes-experts

1.3.1. SECS (wipke)

Le système SECS est dérivé de LHASA.

SECS contient environ 400 règles de production et est conçu de façon à pouvoir être aisément modifié. Tout comme LHASA, ce système est interactif et fonctionne par "pilota-ge", c'est-à-dire qu'il est guidé, au cours de la recherche, par l'utilisateur.

1.3.2. SYNCHEM (gelernter)

- Contrairement à SECS et LHASA, qui ont été conçus dans un souci d'efficacité, SYNCHEM a été conçu à titre expérimental et utilise entièrement les principes d'intelligence artificielle.
- Ce système utilise conjointement deux modes de représentation des molécules [11] :
 - a) Représentation de WISWESSER qui permet au système de tester la "disponibilité commerciale" de telle molécule, par consultation automatique du catalogue de la firme ALDRICH.
 - b) Représentation sous forme de tables de connexion pour la manipulation des molécules.

- Ce système n'est pas interactif : le chimiste n'a donc pas la possibilité de piloter la recherche. A chaque étape, SYNCHEM génère tous les précurseurs possibles et évalue le coût des différentes voies. Ce coût est une fonction de paramètres tels que la facilité, le rendement.

1.3.3. Système de WHITLOCK [12], [13]

- Le système décrit par WHITLOCK ne reconnaît qu'un ensemble limité de groupes fonctionnels et ne fonctionne que dans le cas de molécules-cibles linéaires.
- Ce système fonctionne de la façon suivante :
 1. Le système repère et classe les sites de la molécule-cible présentant un intérêt pour la synthèse : il en déduit un "modèle" de la molécule-cible.
 2. Une fonction-cible est choisie.
 3. Une séquence réactionnelle est générée pour atteindre cette fonction-cible.

1.3.4. Système "CAMEO" (SALATIN et JORGENSEN) [15]

Ce programme a été développé pour prévoir les produits de réactions chimiques à partir de réactifs, de produits initiaux (substrats) et de conditions expérimentales données. L'originalité de ce système par rapport aux systèmes présentés jusqu'ici, est que ce programme n'utilise pas de banque de réactions chimiques, mais base tout son travail sur des principes réactionnels généraux : Addition, Substitution, Elimination, Transfert de protons.

Le programme gère la compétition entre ces principes réactionnels.

CAMEO travaille dans le sens synthétique : il part des produits initiaux et applique ses principes de mécanismes. L'interaction avec l'utilisateur est graphique, et les molécules sont manipulées sous forme de tables de connexion.

1.4. Critique des systèmes-experts en synthèse organique

La synthèse organique n'est pas le seul domaine de la chimie organique où l'on a introduit les systèmes experts. En effet, un des premiers systèmes basés sur les principes d'intelligence artificielle a été DENDRAL [40], dont le but était de proposer des structures moléculaires à partir du spectre de masse et de la formule brute d'un composé. Le principe en est le suivant : DENDRAL possède une base de connaissances qui lui donne, pour chaque fragment moléculaire, les principaux pics de spectrométrie de masse. DENDRAL possède en outre un ensemble de règles qui lui permettent de prévoir les phénomènes de fragmentation et migration.

DENDRAL sélectionne une structure en quatre étapes :

1. Il commence par établir, à partir de la formule brute de la molécule, toutes les formules de structure "chimiquement possibles".
2. Sur base des pics les plus intenses du spectre réel, il retient les structures candidates concordantes.
3. Pour chaque structure candidate, il établit le spectre de masse "théorique" en tenant compte des migrations.
4. Par confrontation du spectre réel avec chacun des spectres candidats, il retient la (ou les) structures plausibles.

Il faut remarquer que, à part dans les règles de migration, il n'y a pas à proprement parler d'heuristique ; DENDRAL ne fait qu'appliquer une technique d'identification parfaitement au point.

Le problème est tout autre en synthèse organique : COREY a été le premier à essayer de formaliser la démarche du chimiste de synthèse et à proposer une heuristique. Ces règles sont empiriques, souvent intuitives et très complexes : un motif structural aura autant de comportements que d'environnements possibles !

2. BASES DE DONNEES

2.1. But

Le but d'une base de données analogique n'est pas de concevoir des schémas réactionnels, mais bien de rechercher, dans une base de données, les schémas réactionnels répondant à des caractéristiques précises : produits de départ ou d'arrivée, groupes fonctionnels formés ou rompus, réactifs.

Un tel système facilite donc le travail bibliographique du chimiste en lui fournissant, par exemple, des méthodes de synthèse pour des composés organiques analogues à ceux qu'il souhaite synthétiser.

2.2. Systèmes existants

2.2.1. Système de COHEN

2.2.1.1. Principe

- Le catalogue est constitué de réactions chimiques.
- Chaque réaction du catalogue consiste en un ensemble de paramètres décrivant les caractéristiques structurales du produit de départ, du produit d'arrivée, ainsi qu'un paramètre décrivant le type de la réaction. Un texte donne les informations essentielles concernant la réaction : conditions expérimentales et références. Ce texte peut se réduire à un numéro de fiche "off-line" ou de microfilm.
- Chaque article du catalogue contient un seul produit de départ et un seul produit d'arrivée ; dès lors, si deux ou plusieurs produits de départ ou d'arrivée contribuent à la même réaction, il est nécessaire de coder et d'introduire plusieurs fois cette réaction dans le catalogue.

2.2.1.2. Codage

Chaque produit de départ et d'arrivée est représenté par :

a) des paramètres structuraux

- Le squelette de la molécule est tout d'abord décrit par un paramètre principal : Main Structural Parameter (MSP).

Ex.



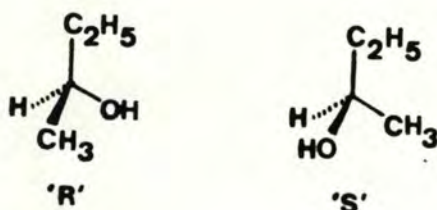
- Les substituants (alkyle, aryle...) ainsi que les groupes fonctionnels reliés au MSP sont décrits par des paramètres secondaires.

b) des paramètres de position

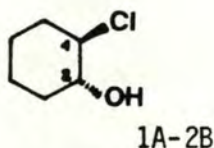
Chaque substituant ou groupe fonctionnel est suivi par un nombre représentant sa position sur le MSP ou sur un substituant primaire ou secondaire.

c) des paramètres stéréochimiques

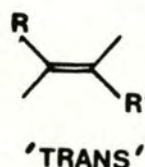
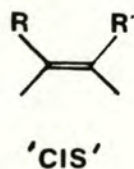
- stéréochimie absolue : "R" ou "S"



- stéréochimie par rapport à un cycle :



- stéréochimie des doubles liaisons : CIS ou TRANS :



Enfin, la transformation chimique est caractérisée par un paramètre décrivant le type de la réaction : Reaction Type Parameter :

"NC" : pas de modification du squelette carboné.

"C+" : formation d'une liaison carbone-carbone.

"C-" : rupture d'une liaison carbone-carbone.

2.2.1.3. Recherche dans la Base de Données

Le chimiste peut interroger le système de différentes manières :

- a) Liste des réactions se faisant sur un produit donné.
- b) Liste des réactions conduisant à un produit donné.
- c) Liste des réactions qui transforment un produit donné en un autre produit donné.

2.2.2. Système développé par A. MICHEL (*)

Le système développé par A. MICHEL est basé sur les possibilités offertes par des progiciels de gestion de fichiers tels que D-BASE II.

Le système propose 17 tables : chaque table correspond à une catégorie de groupe fonctionnel.

Une table se présente comme un menu : chaque option de la table est référencée par une lettre.

Le chimiste décrit la réaction de la manière suivante :

- Premier menu :

1. Liaison formée.
2. Liaison détruite.
3. Liaison inchangée.

- Deuxième menu :

Sommaire des 17 tables fonctionnelles (tables A à Q).

Pour décrire sa réaction, le chimiste revient autant de fois qu'il le désire au menu principal.

Une réaction chimique sera donc décrite par une chaîne de caractères dont chaque lettre représente un accès à une table.

2.2.3. SYNLIB

- SYNLIB (SYNthesis LIBrary) est un système interactif implémenté sur VAX 11/780 et utilisant un terminal VT 100 muni d'un light-pen.
- Ce système est constitué d'une librairie de réactions chimiques, d'un logiciel graphique pour l'interaction avec l'utilisateur ainsi que d'une série d'algorithmes destinés à la consultation de la BD.

(*) *Système développé au département de chimie des Facultés Universitaires de Namur par A.MICHEL, en collaboration avec A.KRIEF.*

2.2.3.1. Introduction d'un schéma réactionnel

L'utilisateur introduit un nouveau schéma réactionnel à l'aide du light-pen.

Les produits de départ et d'arrivée sont dessinés à l'écran : les réactifs, conditions expérimentales et références bibliographiques sont introduites au clavier.

2.2.3.2. Recherche dans la BD.

- La version actuelle de SYNLIB permet de rechercher dans la base de données les schémas réactionnels conduisant à un produit donné ainsi que les schémas réactionnels se faisant sur un produit donné.
- La recherche se fait de la manière suivante :
 1. L'utilisateur dessine à l'écran la molécule qu'il désire rechercher dans la BD.
 2. Il peut, s'il le désire, préciser sa question en marquant d'un astérisque les atomes importants sur lesquels doit porter la recherche.
 3. Lorsque la recherche est terminée, le système fournit à l'utilisateur le nombre de schémas trouvés : l'utilisateur peut à ce moment augmenter - ou au contraire diminuer - la précision de sa question par marquage, et relancer le processus de recherche.
 4. Les schémas réactionnels sont enfin sortis à l'écran ou sur imprimante.

2.3. Remarques à propos des systèmes analogiques existants

2.3.1. Codage des informations

Trois modes de représentation de l'information apparaissent dans les systèmes étudiés :

- a) Représentation d'une molécule par un code pouvant être implémenté, par exemple, sous forme d'une chaîne de caractères (COHEN).
- b) Représentation d'une réaction chimique (et non plus d'une molécule) sous forme d'une liste de pointeurs vers des tables de modifications de structures (A. MICHEL).
A noter que, dans ce cas, les structures moléculaires ne sont pas représentées : seules les modifications de structures survenant au cours de la réaction seront mémorisées.
- c) Représentation des molécules sous forme de graphes (représentation logique) et codage sous forme de tables de connexion (représentation physique) ; (SYNLIB). Cette représentation logique d'une molécule est très riche car très proche de la réalité :
 - elle contient des informations concernant la structure de la molécule (liaison, géométrie) ;
 - elle est aisément manipulable ;
 - elle est extensible à volonté.

2.3.2. Interface utilisateur/machine

L'interface doit permettre une utilisation simple, rapide et agréable du système. Cette interface devrait en outre masquer les opérations de codage des molécules.

- a) Interface de type "graphique" (SYNLIB) :
 - l'utilisateur introduit graphiquement les molécules ;
 - le codage est pris en charge par le système.

- b) Interface de type "Menu" (A. MICHEL) :
 - l'utilisateur sélectionne des entrées dans des menus ;
 - il n'y a pas d'opérations de codage.

- c) Interface de type "String" (COHEN) :
 - l'utilisateur introduit le "string" de codage du schéma réactionnel.

2.3.3. Aspect conceptuel

- Le système d'A. MICHEL travaille sur les concepts de "transformation" et de "groupe fonctionnel". L'intérêt est que l'information est pré-digérée avant d'être introduite dans le système.

- Les autres systèmes travaillent sur le concept "d'atome" : la conséquence de ce choix est que les groupes fonctionnels - qui sont essentiels en synthèse organique - ne sont pas introduits comme tels dans le système, mais doivent être repérés par recherche d'homomorphisme.

CHAPITRE III
SPECIFICATIONS GENERALES
D'UN OUTIL DE GESTION DE BASE DE DONNEES
EN CHIMIE ORGANIQUE

La première partie de ce chapitre identifie les concepts chimiques à prendre en considération lors de l'élaboration d'une base de données en chimie organique ; ces éléments seront la base de l'analyse conceptuelle (Chapitre V). La seconde partie de ce chapitre trace une ébauche de la solution.

1. MISE EN EVIDENCE DES CONCEPTS ESSENTIELS

1.1. La synthèse d'un composé organique comporte généralement plusieurs étapes consécutives ; ces étapes forment ce que l'on appelle un "schéma réactionnel". Un schéma réactionnel peut ne comporter qu'une étape (schéma réactionnel élémentaire). On considérera par la suite qu'un schéma réactionnel peut être vu comme la séquence de schémas réactionnels élémentaires. Un schéma réactionnel élémentaire décrit la transformation, sous l'action éventuelle de réactifs, de molécules initiales en molécules terminales. Une ou plusieurs de ces molécules terminales peuvent à leur tour être molécules initiales dans la transformation suivante.

1.2. Une molécule organique peut être vue comme un squelette carboné - linéaire ou cyclique - sur lequel sont greffées O, une ou plusieurs fonctions chimiques.

Ces fonctions chimiques peuvent être :

a) Conjuguées :

La réactivité chimique de la molécule dépendra de la proximité de deux ou plusieurs fonctions.

b) Compétitives :

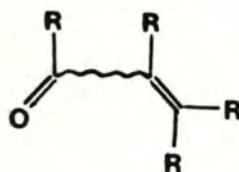
Le réactif a la possibilité de réagir sur deux ou plusieurs groupes fonctionnels ; l'orientation de la réaction dépendra des conditions expérimentales.

c) Indépendantes :

La réactivité de la molécule est la "somme" des réactivités de chaque fonction : le fait d'avoir deux ou plusieurs fonctions ne confère à la molécule aucune réactivité particulière.

1.3. Un motif fonctionnel peut être composé de fonctions chimiques disjointes : par exemple, le chimiste doit pouvoir rechercher les réactions qui font intervenir une CETONE et une OLEFINE sans préciser les positions relatives de ces fonctions dans la molécule.

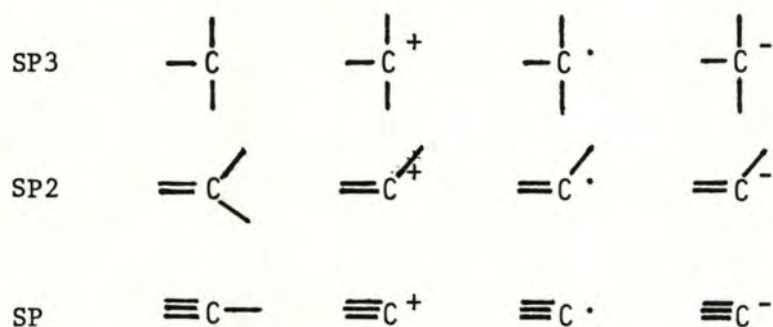
Ex.



Question : "Quelles sont les réactions chimiques qui font intervenir une cétone et une oléfine de la molécule initiale ?"

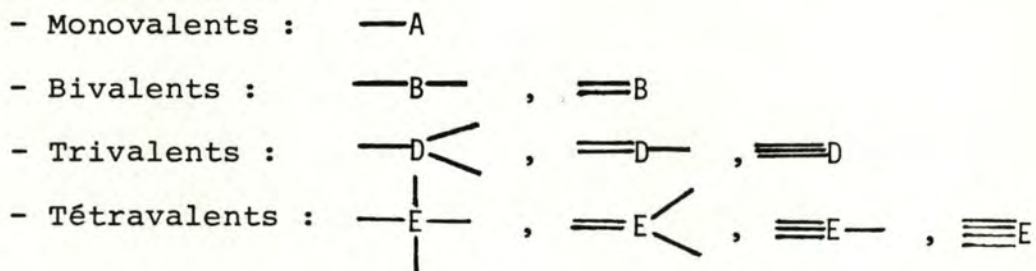
1.4. Une fonction chimique, ainsi que le squelette carboné, sont tous deux composés :

a) d'atomes de carbone sous les états d'hybridation suivants :



b) d'atomes d'Hydrogène (H).

c) d'hétéroatomes :

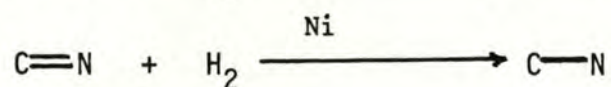
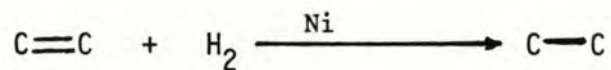


1.5. Il doit être possible de généraliser la notion de squelette carboné et d'hétéroatome : le chimiste désignera par :

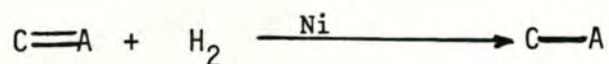
- "R" n'importe quelle chaîne carbonée ;
- "X" n'importe quel halogène.

1.6. De même, il doit être possible de généraliser la notion de FONCTION CHIMIQUE ; certaines réactions chimiques sont en effet spécifiques à un type de liaison chimique, quels que soient les atomes intervenant dans la liaison :

Ex.



Il est utile, dans ce cas, de généraliser la réaction décrite :



où A représente n'importe quel atome lié par une liaison double à un atome de carbone du produit initial.

1.7. La stéréochimie "absolue" ne doit pas nécessairement être prise en ligne de compte : il est donc inutile de prévoir une représentation spatiale des molécules si cela s'avère trop coûteux.

Par contre, la stéréochimie "relative" est importante : type d'attaque sur une double liaison...

1.8. Le langage naturel du chimiste organicien est graphique : il serait donc intéressant que le chimiste puisse communiquer graphiquement avec la base de données.

1.9. Il n'est pas impératif que tous les renseignements se rapportant à un schéma réactionnel se trouvent dans la base de données ; à chaque schéma réactionnel, correspondra dans la base de données :

- a) une représentation du schéma réactionnel ;
- b) des renseignements concernant la stéréochimie relative ;

- c) une cote représentant une "FICHE OFF-LINE" qui contiendra :
- les références bibliographiques ;
 - les conditions expérimentales de la réaction.

- 1.10. La méthode adoptée pour le codage des molécules doit être :
- simple ;
 - de préférence transparente pour le chimiste ;
 - biunivoque : à une molécule correspondra une et une seule représentation, et, inversement, à une représentation correspondra une et une seule molécule.

2. PROPOSITION GENERALE DE SOLUTION

2.1. Qualités attendues

2.1.1. Facilité d'utilisation

La facilité d'utilisation est primordiale : il convient que l'utilisateur soit totalement libéré des opérations de codification des schémas réactionnels.

2.1.2. Choix de représentation

Le choix de la représentation logique d'une molécule est essentiel ; cet aspect a été étudié au Chapitre II. Cet aspect du problème déterminera la qualité des traitements envisageables sur les données. Par exemple, la représentation d'une molécule sous forme d'un graphe autorisera l'utilisation d'algorithmes de recherche de motifs structuraux (reconnaissance de motifs). Une méthode de codage dérivée de la nomenclature IUPAC sera

extrêmement pauvre et limitée quant aux traitements possibles.

2.1.3. Performance du système

Ce travail traite la conception logique du système. Les aspects liés à la performance - quoique essentiels - ne seront pas pris en considération.

2.2. Idées conductrices de la solution

2.2.1. Principes généraux

La solution retenue présente deux aspects :

- a) Saisie et restitution graphique des schémas réactionnels : les molécules seront introduites graphiquement grâce à un crayon optique ou un écran tactile. La rapidité d'introduction des données peut être favorisée en permettant à l'utilisateur :
 - de n'introduire que des morceaux de molécules : il est inutile d'introduire et de stocker une molécule entière si, seuls, quelques atomes ou groupes d'atomes interviennent dans la réaction chimique ;
 - d'utiliser des fonctions pré-programmées pour les motifs structuraux courants : cycles, hétérocycles.

- b) Les molécules seront traitées comme des graphes ; à ce stade, deux solutions sont envisageables :
 - soit une molécule est considérée comme un graphe dont les sommets sont des atomes et les arcs des liaisons chimiques ;
 - soit une molécule est considérée comme un graphe dont les sommets sont des sous-structures (fonctions chimiques, cycles...) et les arcs des liaisons chimiques.

Ces deux solutions sont comparées ci-dessous :

Première solution :

=====

1. Aspect graphique :

- Les molécules sont construites atome par atome.
- Les opérations d'introduction de schémas réactionnels deviennent fastidieuses dans le cas de molécules complexes.
- Afin que les géométries moléculaires soient respectées, il convient que le système guide les opérations de dessin en se basant sur la géométrie des atomes sélectionnés.

2. Reconnaissance des fonctions chimiques [17,27,31,32] :

Le système considère la molécule comme un graphe ; afin de permettre la classification des schémas réactionnels sur base des transformations de fonctions chimiques, il faut prévoir une méthode de reconnaissance de groupes fonctionnels.

Il s'agit de comparer tous les fragments possibles d'une molécule avec un ensemble de fonctions chimiques "types" connues du programme.

Le principe de cette méthode est le suivant :

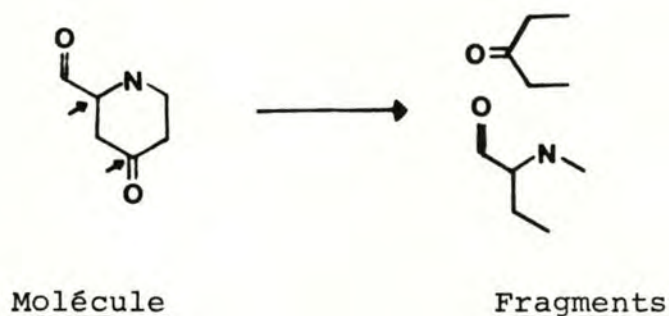
- a) On parcourt le graphe sommet par sommet et, pour chaque sommet i de degré ≥ 3 on isole le sous-graphe tel que le chemin séparant le sommet i de tout sommet j appartenant au sous-graphe est ≤ 2 (les atomes d'hydrogène ne sont pas pris en considération).
- b) On compare les sous-graphes sélectionnés, avec l'ensemble de fonctions connues du système. On peut ainsi voir, pour chaque sous-graphe, si une ou plusieurs fonctions chimiques du catalogue y sont contenues. Afin de ne pas faire inutilement une comparaison sommet par sommet entre chaque sous-graphe et chaque fonction

chimique, on peut imaginer une comparaison rapide : par exemple, avant toute recherche d'homomorphisme, vérifier que :

- 1) toute liaison de la fonction du catalogue est contenue dans le sous-graphe ;
- 2) tout atome de la fonction du catalogue est contenu dans le sous-graphe.

Il faut bien constater que, malgré toutes les optimisations que l'on pourrait y apporter, cette méthode est très lourde et peu compatible avec les exigences d'un système interactif.

Ex.

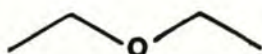


Les sommets de degré ≥ 3 sont marqués d'une flèche. Les deux fragments obtenus doivent ensuite être comparés aux fonctions du catalogue.

Le fait de considérer deux couches d'atomes centrées sur i est très utile, puisque cela permet de repérer les fonctions conjuguées (très importantes dans les réactions de synthèse).

Les seules informations qui ne seront pas détectées par la méthode seront les fonctions de degré < 3 intercalées entre deux chaînes carbonées non substituées de longueur ≥ 2 .

Ex.



3. Classement du schéma réactionnel :

Le classement du schéma réactionnel présente deux exigences :

- A/ On peut vouloir accéder à tous les schémas réactionnels se faisant sur (ou conduisant à) tel et tel groupe fonctionnel : ce type de classement peut être réalisé si tous les groupes fonctionnels ont été détectés et mémorisés.
- B/ On peut vouloir accéder à tous les schémas réactionnels se faisant sur (ou conduisant à) tel et tel produit : pour qu'une telle comparaison soit réalisable, il faut que les molécules aient été numérotées automatiquement de manière telle que :
- 2 molécules identiques présentent toujours le même ordre de numérotation ;
 - 2 molécules différentes présentent toujours une numérotation différente.

Ce problème se complique encore si l'on désire tenir compte de la stéréochimie. Les techniques de numérotation des molécules sont largement décrites dans la littérature [28], [29] ; elles sont généralement efficaces, à l'exception, pour certaines d'entre elles, des cas de molécules à haut degré de symétrie.

Ces techniques de numérotation sont évidemment très lourdes et peu envisageables dans le cas d'un système interactif tournant sur microordinateur.

Deuxième solution :

=====

Une molécule a, jusqu'à présent, été considérée comme un ensemble d'atomes. Cette définition ne tient pas compte du fait qu'une molécule est avant tout pour le chimiste un ensemble de sous-structures réactionnelles : groupes fonctionnels, chaînes carbonées, cycles.

Il serait utile de reconsidérer la solution initiale en adoptant la définition suivante pour une molécule :

- une molécule est un ensemble de sous-structures ;
- une sous-structure peut être :
 - . un groupe fonctionnel ;
 - . un cycle ;
 - . une chaîne carbonée ;
 - . un atome.

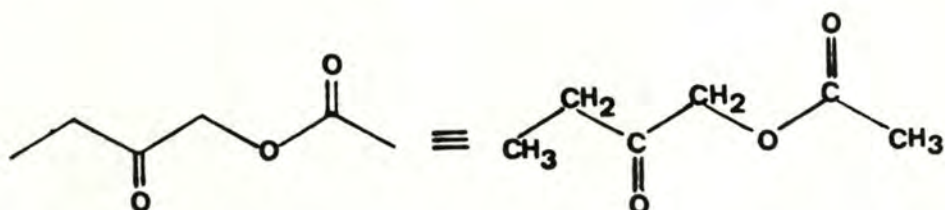
Cette approche du problème présente plusieurs avantages :

- a) le chimiste peut décrire un schéma réactionnel complexe en sélectionnant quelques groupes fonctionnels et éléments du squelette ;
- b) les sous-structures étant pré-définies, il y a moins de calcul préalable à la phase de dessin ;
- c) la classification des schémas réactionnels ne nécessite plus la reconnaissance de groupes fonctionnels.

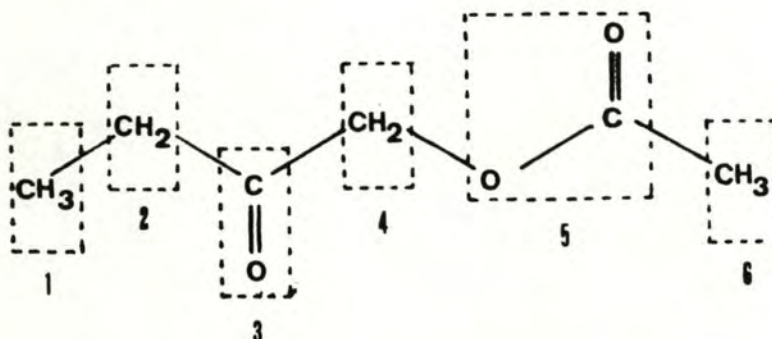
Il faut cependant constater que cette approche ne résout pas les problèmes de reconnaissance :

- de fonctions conjuguées ;
- d'identité de produits.

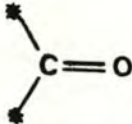
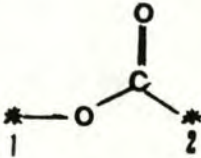
Ex.



Cette structure peut être considérée comme une séquence de sous-structures :



Ces sous-structures sont pré-dessinées dans la base de données :

- 1) $\text{-(CH}_2\text{)-}$ - utilisée pour les 1ère, 2ème, 4ème et 6ème sous-structures
 - CH₂ est complété à CH₃ par défaut.
- 2)  les deux liaisons sont équivalentes.
- 3)  les liaisons (1) et (2) ne sont pas équivalentes : le chimiste doit indiquer, lors de la sélection, par quelle liaison cette sous-structure doit être rattachée à la structure affichée à l'écran.

Une molécule sera toujours finalement représentée sous sa forme simplifiée : sans hydrogène ni carbone. Cependant, durant la phase de dessin, toutes les directions possibles de liaison seront tracées à l'écran.

CHAPITRE IV
ASPECTS GRAPHIQUES LIES A LA SOLUTION

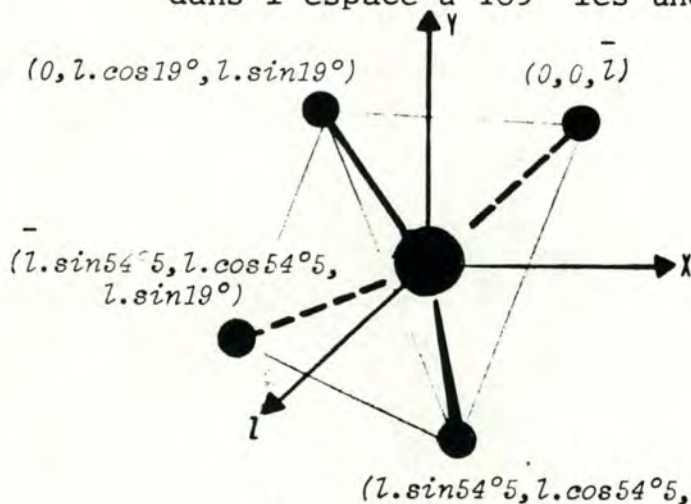
Ce chapitre développe l'aspect graphique de la solution. Nous traiterons d'abord les opérations graphiques classiques (translation, rotation, projection perspective, modification d'échelle) ; nous aborderons ensuite un problème graphique propre à notre solution : l'accrochage de sous-structures.

1. GEOMETRIES ATOMIQUES

La géométrie des structures moléculaires est déterminée par la géométrie des atomes constitutants ; en fait, les géométries atomiques possibles sont peu nombreuses : on peut distinguer trois cas de figure :

1.1. Atome SP₃

Un tel atome possède quatre directions de liaison orientées dans l'espace à 109° les unes des autres :



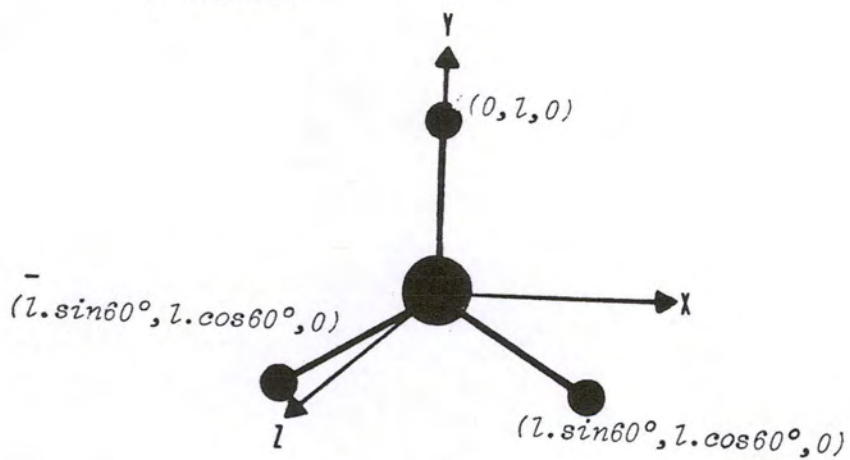
où l représente la longueur de liaison

le système devra posséder l'information suivante :

$$\text{COORD.SP}_3 = \{ (0, l \cos 19^\circ, \sin 19^\circ); (0, 0, l); (\bar{l} \sin 54,5^\circ, l \cos 54,5^\circ, l \sin 19^\circ); (l \sin 54,5^\circ, l \cos 54,5^\circ, l \sin 19^\circ) \}$$

1.2. Atome SP₂

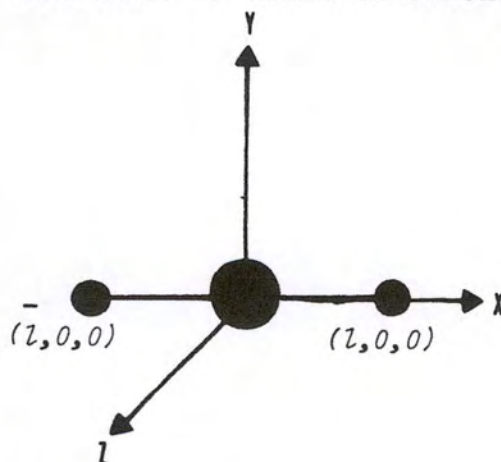
Possède 3 directions de liaison orientées à 120° l'une de l'autre.



$$\text{COORD.SP}_2 = \{(0, l, 0); (\bar{l} \sin 60^\circ, l \cos 60^\circ, 0); (l \sin 60^\circ, l \cos 60^\circ, 0)\}$$

1.3. Atome SP

Possède 2 directions de liaison orientées à 180° l'une de l'autre.



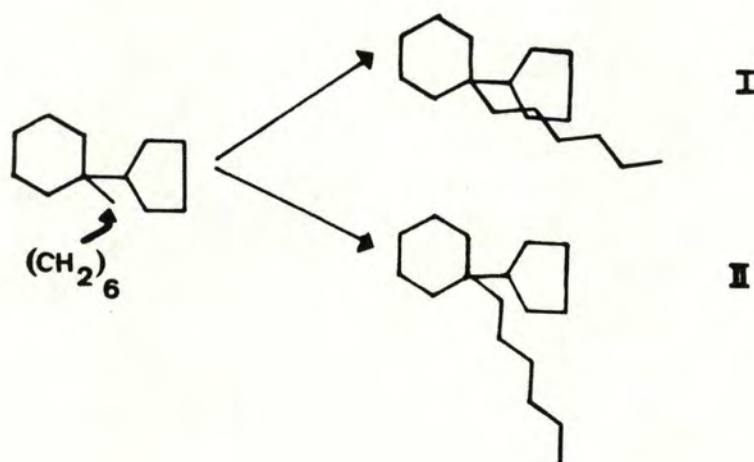
$$\text{COORD.SP} = \{(\bar{l}, 0, 0); (l, 0, 0)\}$$

2. MISE EN EVIDENCE DE QUELQUES PROBLEMES

2.1. Problème de l'encombrement

Il n'est pas toujours possible de greffer une sous-structure sans avoir un recouvrement partiel des structures :

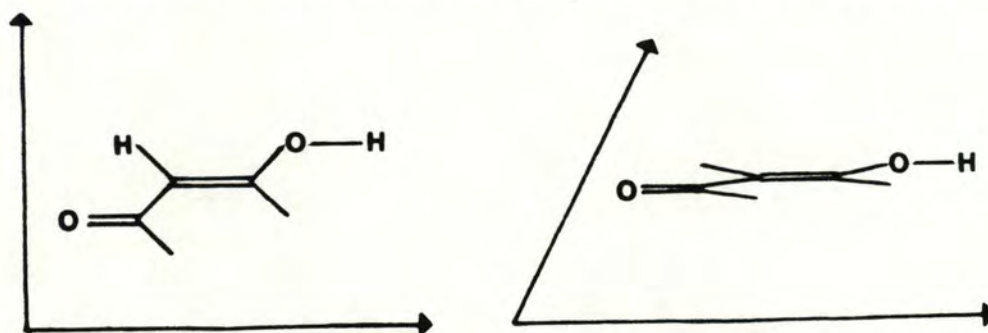
Ex. :



Il faudra pour résoudre ce problème, donner à l'utilisateur la possibilité de modifier l'orientation de sous-structures.

2.2. Problème de l'orientation des molécules

La molécule devra, dans certains cas, être réorientée par l'utilisateur, afin d'offrir un angle de vue optimal.



Tout comme le problème précédent, celui-ci sera résolu en fournissant à l'utilisateur un opérateur de rotation.

3. TRANSLATION

Cette transformation a pour effet de déplacer le point P (x,y,z) à la position P'(x+a,y+b,z+c).

La représentation matricielle de cette transformation sera :

$$(x' \ y' \ z') = (x \ y \ z) + (a \ b \ c)$$

Comme nous allons le voir, la translation est le seul opérateur à nécessiter l'addition matricielle (les autres opérateurs utilisent la multiplication matricielle). Aussi, afin de normaliser ces opérations, on utilise des "coordonnées homogènes" ; cet artifice permet d'exprimer la translation sous forme d'une multiplication matricielle.

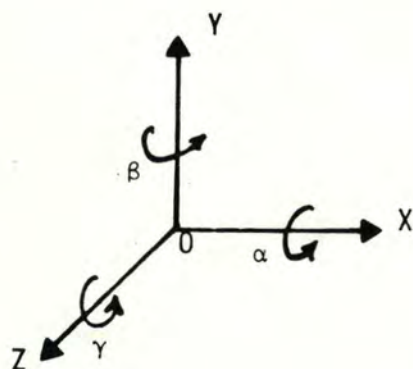
$$(x' \ y' \ z' \ 1) = (x \ y \ z \ 1) \times \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ a & b & c & 1 \end{pmatrix}$$

coordonnées homogènes : 4 coordonnées.

4. ROTATION

4.1. Rotation autour des axes x,y ou z

Le but de cette opération est de déplacer un point ou un objet d'un angle donné autour d'un des axes du système. Nous en donnons les opérateurs :



$$R_{Ox}(\alpha) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha & 0 \\ 0 & -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

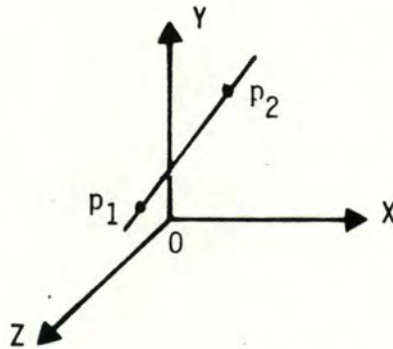
$$R_{Oy}(\beta) = \begin{pmatrix} \cos\beta & 0 & -\sin\beta & 0 \\ 0 & 1 & 0 & 0 \\ \sin\beta & 0 & \cos\beta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$R_{Oz}(\gamma) = \begin{pmatrix} \cos\gamma & \sin\gamma & 0 & 0 \\ -\sin\gamma & \cos\gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

4.2. Rotation autour d'un axe quelconque

L'opérateur dont nous aurons besoin pour effectuer des rotations de sous-structures autour de liaisons chimiques, est un opérateur de rotation autour d'un axe quelconque par rapport au référentiel.

Un tel opérateur peut être construit par composition des opérateurs de rotation et translation définis jusqu'à présent. Soit un référentiel OXYZ et un axe défini par les points $P_1(x_1, y_1, z_1)$ et $P_2(x_2, y_2, z_2)$.



On peut exprimer les équations de l'axe P_1P_2 sous la forme :

$$Y = a_1 + bX$$

$$Y = a_2 + cZ$$

$$\text{où : } b = \frac{Y_2 - Y_1}{x_2 - x_1}$$

$$c = \frac{Y_2 - Y_1}{z_2 - z_1}$$

$$a_1 = \frac{Y_1 x_2 - Y_2 x_1}{x_2 - x_1}$$

$$a_2 = \frac{Y_1 z_2 - Y_2 z_1}{z_2 - z_1}$$

Avant d'effectuer la rotation γ autour de l'axe P_1P_2 , il s'agit de faire coïncider l'axe P_1P_2 et un axe du référentiel (par convention OX).

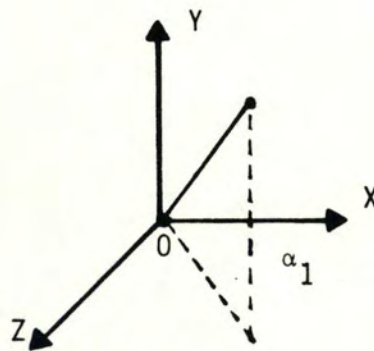
Les étapes successives sont :

- a) Effectuer une translation de l'axe P_1P_2 de façon à le faire passer par l'origine du référentiel.

La matrice de translation s'écrit :

$$T_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & a_1 & a_2 & 1 \end{pmatrix}$$

- b) Amener l'axe de rotation dans le plan OXY ; pour cela, il faut effectuer une rotation α_1 de l'axe P_1P_2 autour de l'axe OX.

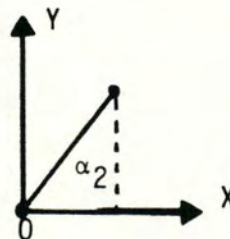


$$\alpha_1 = \text{Arctg} \left(\frac{Z}{X} \right)$$

et la matrice de rotation s'écrit :

$$T_2 = \begin{pmatrix} \cos\alpha_1 & 0 & -\sin\alpha_1 & 0 \\ 0 & 1 & 0 & 0 \\ \sin\alpha_1 & 0 & \cos\alpha_1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- c) Amener l'axe de rotation sur l'axe OX en effectuant une rotation α_2 de l'axe P_1P_2 autour de OZ.



$$\alpha_2 = \text{Arctg} \left(\frac{Y}{X} \right)$$

et la matrice de rotation s'écrit :

$$T_3 = \begin{pmatrix} \cos\alpha_2 & \sin\alpha_2 & 0 & 0 \\ \sin\alpha_2 & \cos\alpha_2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

d) On effectue enfin la rotation désirée (γ) autour de P_1P_2 :

$$R = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\gamma & \sin\gamma & 0 \\ 0 & -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

e) On rétablit le système d'axes initial en effectuant les transformations opposées que nous noterons : T'_1, T'_2, T'_3 .

L'opérateur global peut donc s'exprimer par la composition matricielle de ces différents opérateurs :

$$R_\gamma = T'_1 \times T'_2 \times T'_3 \times R \times T_3 \times T_2 \times T_1$$

5. MODIFICATION D'ECHELLE

Il s'agit de multiplier les coordonnées x, y, z de chaque point respectivement par a, b, c .

La matrice de transformation s'écrit : $S = \begin{pmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

Il faut remarquer qu'une augmentation d'échelle provoquera un déplacement de l'image vers le coin supérieur droit, alors qu'une diminution d'échelle provoquera un déplacement vers le coin inférieur gauche.

Il s'agira donc d'appliquer une correction de translation.

6. PROJECTION PERSPECTIVE

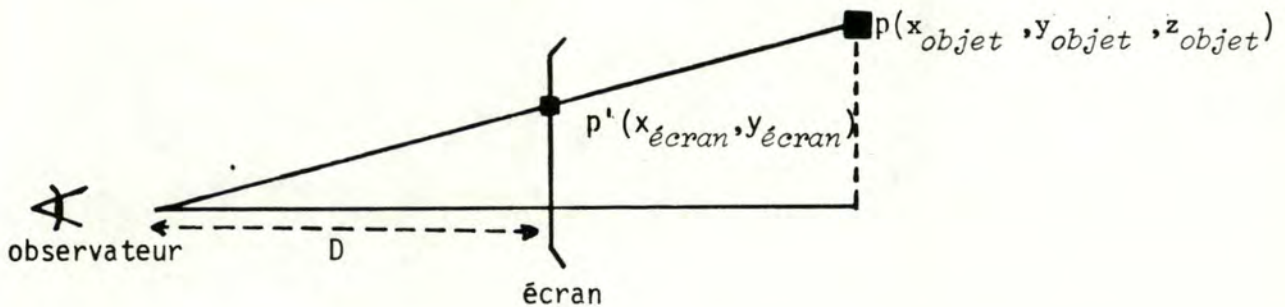
Il reste à décrire l'opération graphique qui permet de projeter une image en trois dimensions sur un écran.

L'effet de perspective dépend en fait de la distance observateur/écran.

Cette transformation est du type :

$$X_{\text{écran}} = \frac{D \cdot X_{\text{objet}}}{Z_{\text{objet}}} \quad (\text{I})$$

$$Y_{\text{écran}} = \frac{D \cdot Y_{\text{objet}}}{Z_{\text{objet}}}$$



Les relations (I) sont obtenues immédiatement par similitude des triangles.

La matrice de transformation aura donc la forme :

$$P = \begin{pmatrix} \frac{D}{z} & 0 & 0 \\ 0 & \frac{D}{z} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

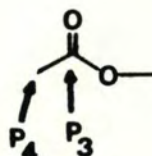
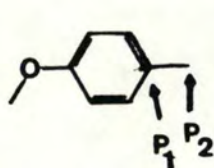
7. ACCROCHAGE DE SOUS-STRUCTURES

L'accrochage d'une nouvelle sous-structure sur un produit se fait en deux étapes :

- 1°) Accrochage de la sous-structure au produit ;
- 2°) Rotation de la sous-structure afin de faire coïncider les liaisons.

Ces deux étapes sont détaillées ci-dessous.

Soit la situation suivante :



$$\begin{aligned} P_1 & (x_1, y_1, z_1) \\ P_2 & (x_2, y_2, z_2) \\ P_3 & (x_3, y_3, z_3) \\ P_4 & (x_4, y_4, z_4) \end{aligned}$$

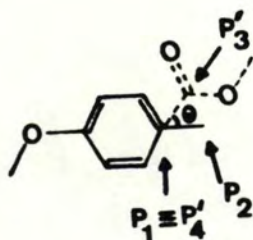
Etape 1 :

Effectuer une translation de la sous-structure de façon à faire coïncider P_1 et P_4 .

L'opérateur matriciel est :

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ x_1 - x_4 & y_1 - y_4 & z_1 - z_4 & 1 \end{pmatrix}$$

La nouvelle situation est :



$$\begin{aligned} P_1 & (x_1, y_1, z_1) \\ P_2 & (x_2, y_2, z_2) \\ P'_3 & (x'_3, y'_3, z'_3) \\ P'_4 & (x'_4, y'_4, z'_4) \end{aligned}$$

Etape 2 :

Il s'agit maintenant :

- 1°) de trouver l'équation de l'axe de rotation (perpendiculaire en P_1 au plan défini par P_1, P_2, P'_3).
- 2°) d'effectuer une rotation θ de la sous-structure autour de cet axe.

- Equation de l'axe de rotation :

- . $P_1 (x_1, y_1, z_1)$ est un point de l'axe ;
- . un second point de l'axe sera : $P (x, y, 0)$

avec :

$$x = \frac{(v_1 - \frac{\mu_1}{\mu_2} v_2) z_1}{(\lambda_1 - \frac{\mu_1}{\mu_2} \lambda_2)} + x_1$$

$$y = \frac{(v_1 - \frac{\lambda_1}{\lambda_2} v_2) z_1}{(\mu_1 - \frac{\lambda_1}{\lambda_2} \mu_2)} + y_1$$

$$\begin{aligned} \text{et : } \lambda_1 &= x_2 - x_1 ; & \lambda_2 &= x'_3 - x_1 \\ \mu_1 &= y_2 - y_1 ; & \mu_2 &= y'_3 - y_1 \\ v_1 &= z_2 - z_1 ; & v_2 &= z'_3 - z_1 \end{aligned}$$

- Calcul de l'angle θ :

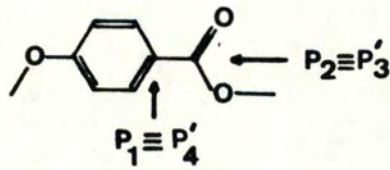
$$\theta = 2 \text{ Arc sin } \left(\frac{D}{2l} \right)$$

avec :

$$D = \sqrt{(x'_3 - x_2)^2 + (y'_3 - y_2)^2 + (z'_3 - z_2)^2}$$

. l = longueur d'une liaison chimique.

Nous obtenons finalement la situation représentée ci-dessous :



CHAPITRE V

ANALYSE CONCEPTUELLE

Le premier point de ce chapitre explique de façon intuitive les différents éléments (ENTITES, RELATIONS) du schéma conceptuel.

Une description plus formelle de ces éléments, ainsi qu'une mise en évidence systématique des contraintes d'intégrité liées au problème, sera abordée aux points 2, 3 et 4. Enfin, au point 5, les spécifications des fonctions de consultation seront abordées.

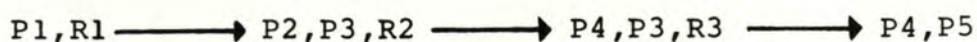
1. ELEMENTS DE L'ANALYSE CONCEPTUELLE

1.1. Schéma réactionnel

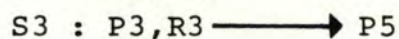
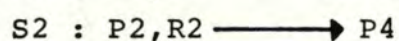
Un SCHEMA REACTIONNEL décrit la transformation d'un ou plusieurs PRODUITS INITIAUX en un ou plusieurs PRODUITS FINAUX sous l'action - éventuelle - d'un ou plusieurs PRODUITS REACTIFS. (On admet qu'une transformation de produits peut se dérouler sans l'intervention de PRODUITS REACTIFS : ce sera le cas lors d'une transformation chimique se faisant sous l'action unique d'une modification des conditions expérimentales : température, pression, solvant...).

Un SCHEMA REACTIONNEL peut être constitué de plusieurs étapes : un schéma réactionnel peut donc être décomposé en un ou plusieurs schémas réactionnels. Il est nécessaire, dans ce cas, de prendre en considération l'ordonnement des schémas.

Ainsi, le schéma réactionnel suivant :



sera représenté par trois schémas réactionnels :

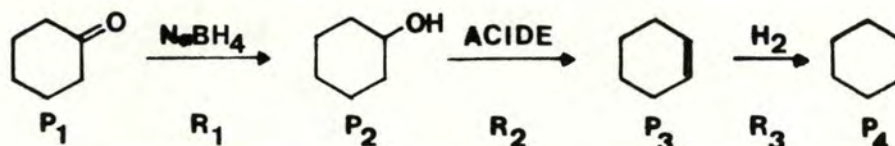


où - S1, S2, S3 identifient les schémas réactionnels et introduisent la notion d'ordre ;

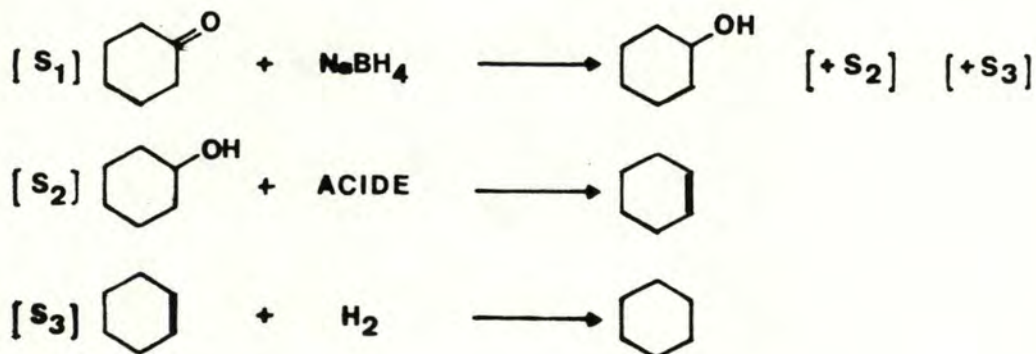
- P1, P2, P3, P4, P5 identifient les différents composés organiques ;

- R1, R2, R3 identifient les différents réactifs.

Ce cas extrême - deux composés P2, P3 transformés successivement en deux autres composés P4, P5 - est rare ; le cas le plus fréquent est illustré par l'exemple ci-dessous :



sera , en fait , représenté par :



1.2. Produit

a) Aspect chimique

Un même produit peut intervenir dans des réactions chimiques distinctes avec des rôles divers :

PRODUIT INITIAL ;

PRODUIT FINAL ;

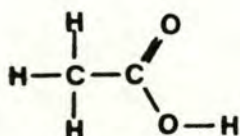
PRODUIT REACTIF.

Pour le chimiste organicien, la structure d'un PRODUIT INITIAL ou FINAL est primordiale ; par contre, un REACTIF désignera un PRODUIT dont la structure - ou les modifications de structure - ne présentent pas d'intérêt. Ainsi, l'acide acétique peut intervenir dans une première réaction en tant qu'acide - c'est-à-dire en tant que générateur de protons -, et dans une seconde réaction en tant que SUBSTRAT ayant à subir des modifications de structure.

Dans le premier cas, la structure de l'acide acétique n'intéresse pas le chimiste ; on représentera le REACTIF par :

HAc. (qui est une chaîne de caractères).

Dans le second cas, la structure est primordiale : l'acide acétique sera représenté par sa formule développée :



(qui est un graphe)

Un produit intervenant en tant que réactif n'aura pas de structure associée.

b) Aspect graphique

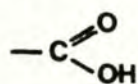
Un PRODUIT INITIAL ou TERMINAL sera constitué d'un ensemble de sous-structures (cfr. définition plus loin) ordonnées. Un tel produit peut donc être considéré comme un GRAPHE décomposable en niveaux dont les sommets seraient des sous-structures et les arêtes des liaisons joignant les sous-structures. (Ceci est vrai à condition que tout cycle appartenant au produit soit considéré comme sous-structure : un produit ne contiendra donc jamais de "cycle de sous-structures").

Précisons qu'un produit doit être considéré dans un espace tridimensionnel : à chaque sommet du graphe sera donc associé l'ensemble de ses coordonnées (x,y,z).

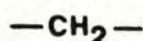
1.3. Sous-structures

On définit comme SOUS-STRUCTURE tout ensemble d'atomes sémantiquement riche pour le chimiste organicien : ce peut être une fonction, une chaîne, un cycle, un cycle ponté, un polycycle, un cycle modifié, ou, plus simplement, un atome.

Quelques exemples de sous-structures sont donnés ci-dessous :



: Fonction



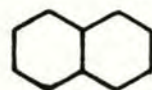
: Chaîne carbonée



: Cycle modèle



: Cycle modifié (obtenu par modification de :



: Polycycle



: Cycle ponté

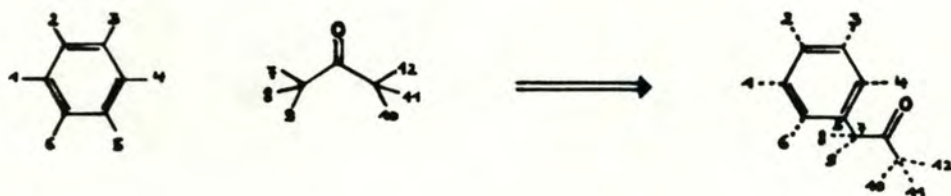


: Atome

Une SOUS-STRUCTURE est donc un ensemble ordonné d'ATOMES pris dans une GEOMETRIE particulière.

A une sous-structure correspond un certain nombre de PROLONGEMENTS distincts permettant de la relier à ses suivantes.

Ex. :

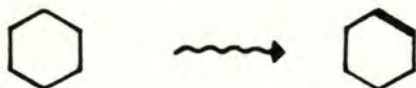


Une sous-structure est elle-même un graphe non décomposable en niveaux, c'est-à-dire pouvant contenir des cycles.

L'agencement des atomes constitutifs de la sous-structure doit être considéré dans un espace à trois dimensions : à chaque sommet du graphe sera associé l'ensemble de ses coordonnées.

Une sous-structure pourra être construite à partir de sous-structures pré-existantes ; cette possibilité sera particulièrement utile lors de la construction d'hétérocycles et de polycycles.

Ex. :



1.4. Notion de COMPOSANT

Une même sous-structure peut intervenir dans la description de plusieurs PRODUITS et peut même participer plusieurs fois à la description d'un même PRODUIT.

Dans chacune de ses participations, la sous-structure :

1. fera intervenir des prolongements différents : ces prolongements sont donc associés à une occurrence d'une sous-structure et non à la sous-structure elle-même ;
2. sera positionnée et orientée différemment : chaque occurrence d'une sous-structure possèdera ses coordonnées globales ainsi qu'un angle d'orientation.

Chaque occurrence d'une sous-structure sera considérée comme un COMPOSANT d'un PRODUIT.

A chaque COMPOSANT sera associé l'ensemble des PROLONGEMENTS associé à la SOUS-STRUCTURE "mère".

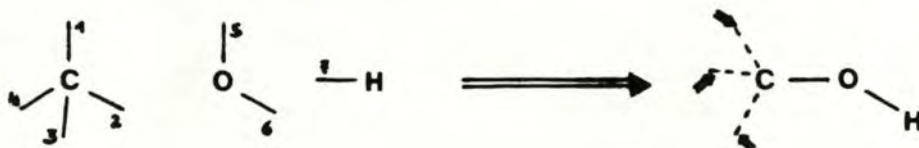


(C1 et C2 sont deux occurrences distinctes - c'est-à-dire deux composants - d'une même sous-structure).

1.5. Notion de PROLONGEMENT

Un PROLONGEMENT est une LIAISON CHIMIQUE - libre ou non - associée à une occurrence d'une sous-structure (c'est-à-dire à un COMPOSANT). L'ensemble des prolongements associé à un composant est en fait l'ensemble des liaisons des atomes constitutifs de ce composant.

L'ensemble des prolongements LIBRES associé à une sous-structure constitue l'ensemble des liaisons pouvant participer à la formation d'un PRODUIT.



1.6. Notion d'ATOME

L'ATOME est le constituant de base de toute sous-structure. Un atome est caractérisé soit par le SYMBOLE chimique associé, soit par son NUMERO ATOMIQUE.

Notons qu'un atome peut également être caractérisé par la PERIODE et la FAMILLE auxquelles il appartient ; cette caractérisation est celle adoptée dans le Tableau Périodique des Eléments Chimiques de Mendeleev ; elle présente l'avantage de donner implicitement des informations primordiales sur les atomes :

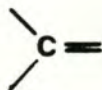
Ex. : tout atome appartenant à la famille "7a" possède 7 électrons sur sa couche externe, possède donc un électron non apparié et est donc très électronégatif.
Tous les atomes d'une même famille présentent des propriétés similaires.

Un atome peut exister sous diverses GEOMETRIES, selon l'étage d'oxydation auquel il se trouve.

Ex. : l'atome de CARBONE (symbole : C ; numéro atomique : 6) peut exister sous les géométries suivantes :



SP3



SP2



SP

On peut même étendre la notion de "géométrie atomique" de façon à englober la notion de charge.

Ainsi, C^+ , C^- , C^\cdot ne seraient que des cas particuliers de géométries atomiques.

1.7. Notion de SOMMET

A un atome et une géométrie donnée correspond un ensemble de SOMMETS.

A chaque sommet correspond l'ensemble de ses coordonnées spatiales ; ces coordonnées sont exprimées en prenant pour convention que l'ATOME se trouve au point de coordonnées (0,0,0). Graphiquement, chaque sommet sera relié à son atome par une LIAISON simple ou multiple ; le "type" associé au sommet indique la multiplicité de la liaison.

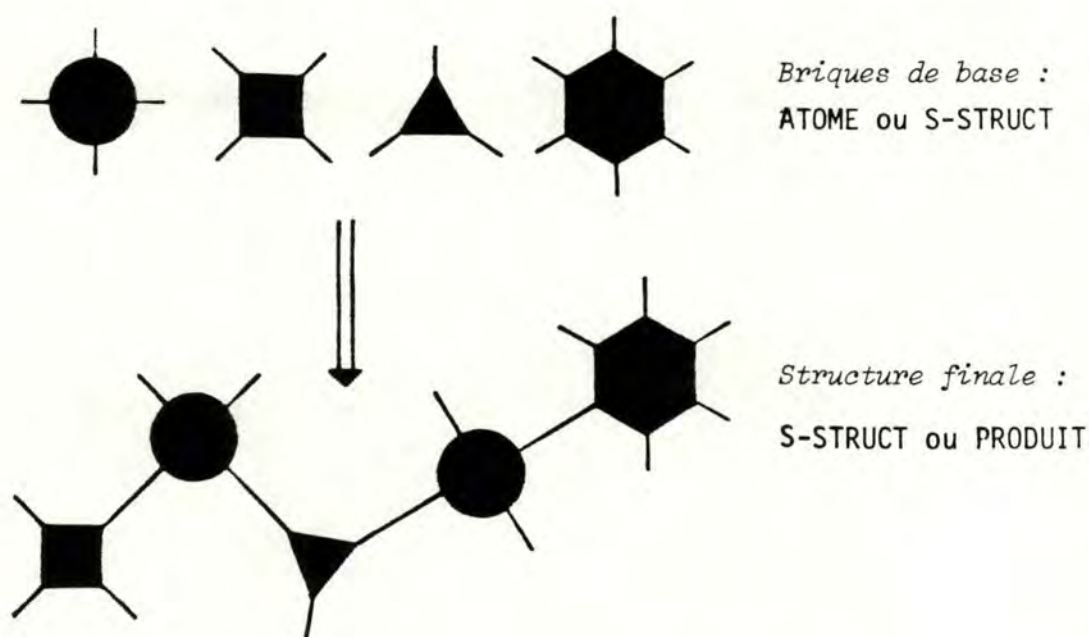
1.8. Notion de COMPOSANT ATOMIQUE

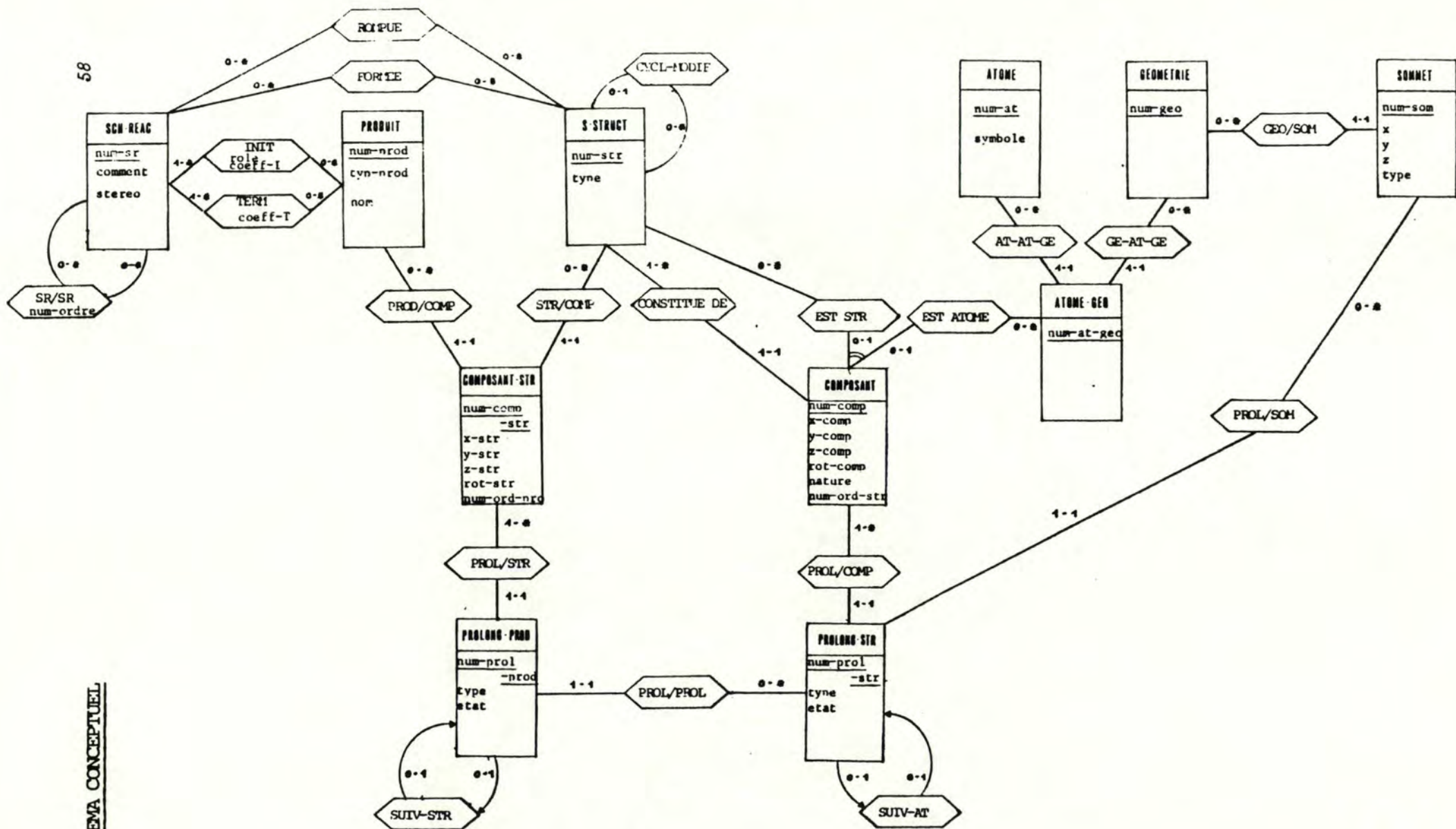
Un même atome peut, sous une géométrie donnée, intervenir dans plusieurs SOUS-STRUCTURES, et même plusieurs fois dans la même sous-structure. Dans chacune de ses participations, l'atome peut faire intervenir des prolongements différents. Un ensemble de prolongements est donc associé à chaque occurrence d'un ATOME. Chaque occurrence d'un atome est un COMPOSANT ATOMIQUE. Un composant atomique sera donc une occurrence d'un atome, dans une géométrie particulière, participant à une occurrence de SOUS-STRUCTURE.

Il faut remarquer la symétrie du problème :

- un PRODUIT est composé d'occurrences de SOUS-STRUCTURES reliées entre elles par certains de leurs PROLONGEMENTS ;
- une sous-structure est composée d'ATOMES et/ou de SOUS-STRUCTURES reliés entre eux par certains de leurs prolongements.

Ainsi, le schéma suivant représente aussi bien un PRODUIT constitué de SOUS-STRUCTURES, qu'une SOUS-STRUCTURE constituée d'ATOMES et de SOUS-STRUCTURES :





3. EXPLICATION DU SCHEMA CONCEPTUEL

Afin de simplifier l'explication liée au schéma conceptuel, considérons 4 BLOCS :

BLOC_1 :

- Lié à la notion de SCHEMA REACTIONNEL.
- Comprend - l'entité "SCH-REAC" ;
 - les associations "SR/SR"
 - "INIT"
 - "TERM"
 - "ROMPUE"
 - "FORMEE"

BLOC_2 :

- Lié à la notion de PRODUIT.
- Comprend - les entités "PRODUIT"
 - "COMPOSANT-STR"
 - "PROLONG-PROD"
- les associations "PROD/COMP"
 - "STR/COMP"
 - "PROL/STR"
 - "PROL/PROL"
 - "SUIV/STR"

BLOC_3 :

- Lié à la notion de SOUS-STRUCTURE.
- Comprend - les entités "S-STRUCT"
 - "COMPOSANT"
 - "PROLONG-STR"

- les associations "CYCLE/MODIF"
- "CONSTITUE DE"
- "EST STR"
- "EST ATOME"
- "PROL/COMP"
- "SUIV/AT"
- "PROL/SOM"

BLOC_4 :

- Lié à la notion d'ATOME.
- Comprend - les entités "ATOME"
- "ATOME-GEO"
- "GEOMETRIE"
- "SOMMET"
- les associations "AT/AT/GE"
- "GE/AT/GE"
- "GEO/SOM"

3.1. Bloc 1

3.1.1. SCH-REAC (type d'entité)

Toute occurrence de SCH-REAC représente un schéma-réactionnel. SCH-REAC comprend les attributs :

- NUM-SR (qui est un numéro IDENTIFIANT) ;
- COMMENT (qui est un texte de commentaires liés au schéma réactionnel) ;
- STEREO (qui donne la stéréochimie de la réaction). Cette façon de traiter la stéréochimie de la réaction chimique peut être complétée par une représentation tridimensionnelle des molécules.

3.1.2. SR/SR (type d'association)

Ce type d'association traduit le fait qu'un schéma réactionnel peut être composé de plusieurs autres schémas réactionnels.

Attribut : NUM-ORDRE qui donne l'ordre de succession des schémas réactionnels "cibles" dans le schéma réactionnel "origine".

Toute occurrence de SR/SR représente l'association d'une occurrence de SCH-REAC à une autre occurrence de SCH-REAC. Cette association n'est obligatoire ni pour l'entité origine, ni pour l'entité cible.

Cette association est "many-to-many" : une occurrence de SCH-REAC peut être :

- origine de plusieurs associations ;
- cible de plusieurs associations.

3.1.3. INIT (type d'association)

Associe, à un schéma réactionnel, tous ses produits initiaux. Toute occurrence de INIT associe une occurrence de SCH-REAC à une occurrence de PRODUIT.

INIT est une association "many-to-many". L'association INIT est obligatoire pour le type d'entité SCH-REAC et facultative pour le type d'entité PRODUIT.

Attributs :

- ROLE : - attribut obligatoire ;
 - indique si le produit intervient comme réactif (R) ou comme composé organique (C). (Les valeurs R et C sont les seules permises).
- COEFF-I :
 - attribut obligatoire ;
 - indique le nombre de molécules du produit intervenant dans l'équation moléculaire ;
 - la valeur par défaut sera "1".

3.1.4. TERM (type d'association)

Associe, à un schéma réactionnel, tous ses produits terminaux.

Toute occurrence de TERM associe une occurrence de SCH-REAC à une occurrence de PRODUIT.

TERM est une association "many-to-many".

TERM est une association obligatoire pour SCH-REAC et facultative pour PRODUIT.

Attribut : COEFF-T : indique le nombre de molécules de PRODUIT intervenant dans l'équation moléculaire. La valeur par défaut de COEFF-T est "1".

3.1.5. ROMPUE (type d'association)

Associe, à un schéma réactionnel, toutes les sous-structures rompues lors d'une réaction chimique.

Toute occurrence de ROMPUE associe une occurrence de SCH-REAC à une occurrence de S-STRUCT.

Cette association est "many-to-many" ; elle est facultative pour SCH-REAC et S-STRUCT.

3.1.6. FORMEE (type d'association)

Associe, à un schéma réactionnel, toutes les sous-structures formées.

Toute occurrence de FORMEE associe une occurrence de SCH-REAC à une occurrence de PRODUIT.

Cette association est "many-to-many" et est facultative pour PRODUIT et SCH-REAC.

3.2. Bloc 2

3.2.1. PRODUIT (type d'entité)

Toute occurrence de PRODUIT représente un produit chimique.

Attributs :

- NUM-PROD : numéro identifiant l'occurrence de PRODUIT (attribut obligatoire).
- NOM : chaîne de caractères représentant le nom du produit (attribut facultatif).
- TYPE-PROD : indique si le PRODUIT possède une structure (S) ou se limite à un nom (N) ; (attribut obligatoire).

3.2.2. COMPOSANT-STR (type d'entité)

Représente une occurrence d'une sous-structure participant à la composition d'un produit.

Attributs :

- NUM-COMP-STR :
 - . numéro identifiant une occurrence de COMPOSANT-STR ;
 - . attribut obligatoire.
- X-STR, Y-STR, Z-STR :
 - . translation que doit subir l'occurrence de COMPOSANT-STR lors de son "accrochage" au produit ;
 - . attributs obligatoires.
- ROT-STR :
 - . rotation que doit subir l'occurrence de COMPOSANT-STR lors de son "accrochage" au produit ;
 - . attribut obligatoire.
- NUM-ORD-PROD :
 - . numéro de séquence du composant dans le produit ;
 - . attribut obligatoire.

3.2.3. PROLONG-PROD (type d'entité)

Une occurrence de PROLONG-PROD représente un prolongement - c'est-à-dire une "valence" chimique libre ou saturée - appartenant à une occurrence de COMPOSANT-STR.

Attributs :

- NUM-PROL-PROD :
 - . numéro identifiant une occurrence de PROLONG-PROD ;
 - . attribut obligatoire.
- TYPE :
 - . indique le type de prolongement ;
 - . attribut obligatoire ;
 - . les types possibles sont "simple", "double", "triple".
- ETAT :
 - . indique si le prolongement est "libre" ou "saturé" ;
 - . attribut obligatoire.

3.2.4. PROD/COMP (type d'association)

Associe, à un PRODUIT, tous ses COMPOSANTS ; (l'ordre des composants dans le graphe du produit est donné par l'attribut "num-ord-pro" de COMPOSANT-STR).

Cette association est "one-to-many" : à une occurrence de PRODUIT peuvent correspondre plusieurs occurrences de composant-str, mais une occurrence de composant-str n'appartient qu'à un PRODUIT.

Cette association est :

- facultative pour PRODUIT (un produit peut ne pas posséder de structure) ;
- obligatoire pour COMPOSANT-STR.

3.2.5. STR/COMP (type d'association)

Associe, à une sous-structure, toutes ses occurrences.

Une occurrence de STR/COMP associe une occurrence de COMPOSANT-STR à une occurrence de S-STRUCT.

STR/COMP est une association "one-to-many" : une occurrence de S-STRUCT peut être associée à plusieurs occurrences de COMPOSANT-STR. Cette association est obligatoire pour COMPOSANT-STR et facultative pour S-STRUCT.

3.2.6. PROL/STR (type d'association)

Associe, à un composant, tous ses prolongements.

Une occurrence de PROL/STR associe une occurrence de PROLONG-PROD à une occurrence de COMPOSANT-STR.

PROL/STR est une association "one-to-many" : une occurrence de COMPOSANT-STR peut posséder plusieurs occurrences de PROLONG-PROD. PROL/STR est obligatoire pour COMPOSANT-STR et pour PROLONG-PROD.

3.2.7. PROL/PROL (type d'association)

Associe, à un prolongement de sous-structure, toutes ses occurrences.

Une occurrence de PROL/PROL associe une occurrence de PROLONG-PROD à une occurrence de PROLONG-STR.

PROL/PROL est une association "one-to-many" : une occurrence de PROLONG-STR peut être associée à plusieurs occurrences de PROLONG-PROD. PROL/PROL est obligatoire pour l'entité PROLONG-PROD et facultative pour l'entité PROLONG-STR.

3.2.8. SUIV-STR (type d'association)

L'association SUIV-STR donne la SEQUENCE des sous-structures intervenant dans la composition d'un produit.

Une occurrence de SUIV-STR associe une occurrence de PROLONG-PROD à UNE AUTRE occurrence de PROLONG-PROD.

SUIV-STR est une association "one-to-one" facultative pour l'entité-origine comme pour l'entité-cible.

3.3. Bloc 3

3.3.1. S-STRUCT (type d'entité)

Toute occurrence de S-STRUCT représente une sous-structure.

Attributs :

- NUM-STR :
 - . numéro identifiant une occurrence de S-STRUCT ;
 - . attribut obligatoire.
- TYPE :
 - . représente le type de sous-structure ;
 - . attribut obligatoire ;
 - . valeurs autorisées : "cycle", "cycle modifié", "poly-cycle", "chaîne", "atome".

3.3.2. COMPOSANT (type d'entité)

Une occurrence de COMPOSANT représente une occurrence de S-STRUCT ou d'ATOME participant à la composition d'une S-STRUCT.

Attributs :

- NUM-COMP :
 - . numéro identifiant une occurrence de COMPOSANT ;
 - . attribut obligatoire.
- X-COMP, Y-COMP, Z-COMP :
 - . translation que doit subir l'occurrence de COMPOSANT lors de son "accrochage" au produit ;
 - . attributs obligatoires.
- ROT-STR :
 - . rotation que doit subir l'occurrence de COMPOSANT lors de son "accrochage" au produit ;
 - . attribut obligatoire.
- NATURE :
 - . indique le type de composant : "atome" ou "sous-structure" ;
 - . attribut obligatoire.
- NUM-ORD-STR :
 - . numéro de séquence du composant dans la sous-structure ;
 - . attribut obligatoire.

3.3.3. PROLONG-STR (type d'entité)

Une occurrence de PROLONG-STR représente un prolongement - c'est-à-dire une "valence" chimique libre ou saturée - appartenant à une occurrence de COMPOSANT.

Attributs :

- NUM-PROL-STR :
 - . numéro identifiant une occurrence de PROLONG-STR ;
 - . attribut obligatoire.
- TYPE :
 - . indique le type de prolongement ;
 - . attribut obligatoire ;
 - . les types possibles sont "simple", "double", "triple".
- ETAT :
 - . indique si le prolongement est "libre" ou "saturé" ;
 - . attribut obligatoire.

3.3.4. CONSTITUE_DE (type d'association)

Association donnant accès à tous les COMPOSANTS d'une SOUS-STRUCTURE ; (l'ordre des composants dans le graphe de la sous-structure est donné par l'attribut "num-ord-str" de l'entité COMPOSANT.

Cette association est :

- "one-to-many" ;
- obligatoire pour les deux types d'entité.

3.3.5. EST_STR (type d'association)

Associe un COMPOSANT à la SOUS-STRUCTURE dont il est une occurrence. Une occurrence de EST STR associe, à une occurrence de COMPOSANT, une occurrence de SOUS-STRUCTURE.

Cette association est "many-to-one" : une même occurrence de S-STRUCT peut être associée à plusieurs occurrences de COMPOSANT.

L'association EST STR est facultative pour les entités S-STRUCT et COMPOSANT.

3.3.6. EST_ATOME (type d'association)

Associe un composant à l'atome dont il est une occurrence. Une occurrence de EST ATOME associe une occurrence de COMPOSANT à une occurrence de ATOME-GEO.

L'association EST ATOME est "many-to-one" : une même occurrence de ATOME-GEO peut être associée à plusieurs occurrences de COMPOSANT.

Cette association est facultative pour les entités ATOME-GEO et COMPOSANT.

3.3.7. CYCLE-MODIF (type d'association)

Une occurrence de l'association CYCLE-MODIF associe, à une occurrence de S-STRUCT de type "cycle modifié", l'occurrence de S-STRUCT de type "cycle" dont il est issu.

L'association CYCLE-MODIF est "one-to-many" : plusieurs occurrences de S-STRUCT de type "cycle modifié" peuvent être issues de la même occurrence de S-STRUCT de type "cycle".

L'association est facultative pour les entités S-STRUCT "origines" ainsi que pour les entités S-STRUCT "cibles".

3.3.8. PROL/COMP (type d'association)

Associe, à un COMPOSANT, tous ses PROLONGEMENTS.

Une occurrence de PROL/COMP associe une occurrence de COMPOSANT à une occurrence de PROLONG-STR.

L'association PROL/COMP est "one-to-many" : une occurrence de COMPOSANT peut être associée à plusieurs occurrences de PROLONG-STR.

L'association PROL/COMP est obligatoire pour les entités COMPOSANT et PROLONG-STR.

3.3.9. PROL/SOM (type d'association)

Une occurrence de PROL/SOM associe une occurrence de PROLONG-STR à une occurrence de SOMMET.

L'association PROL/SOM est "many-to-one" : à une occurrence de SOMMET peuvent correspondre plusieurs occurrences de PROLONG-STR.

L'association PROL/SOM est obligatoire pour l'entité PROLONG-STR et facultative pour l'entité SOMMET.

3.3.10. SUIV-AT (type d'association)

L'association SUIV-AT donne la SEQUENCE des sous-structures et atomes intervenant dans la composition d'une sous-structure.

Une occurrence de SUIV-AT associe une occurrence de PROLONG-STR à UNE AUTRE occurrence de PROLONG-STR.

SUIV-AT est une association "one-to-one" facultative pour l'entité-origine comme pour l'entité-cible.

3.4. Bloc 4

3.4.1. ATOME (type d'entité)

Une occurrence de ATOME représente un atome et est identifiée par NUM-AT.

Attributs :

- NUM-AT : numéro atomique ;
- SYMBOLE : symbole chimique de l'atome.

3.4.2. GEOMETRIE (type d'entité)

Une occurrence de GEOMETRIE représente un type de géométrie d'atome.

Attribut :

- NUM-GEO : numéro identifiant une occurrence de GEOMETRIE.

3.4.3. ATOME-GEO (type d'entité)

Une occurrence de ATOME-GEO représente une occurrence de GEOMETRIE associée à une occurrence de ATOME.

Attribut :

- NUM-AT-GEO : numéro identifiant de ATOME-GEO.

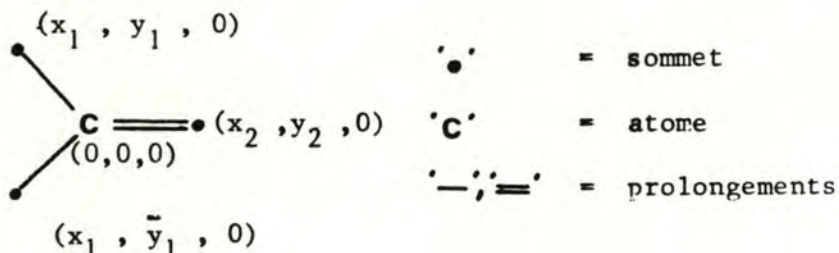
3.4.4. SOMMET (type d'entité)

Une occurrence de SOMMET représente l'extrémité d'un prolongement associé à une GEOMETRIE.

Attributs :

- NUM-SOM : numéro identifiant un SOMMET ;
- X,Y,Z : coordonnées associées à un sommet, c'est-à-dire à l'extrémité d'un prolongement ;
- TYPE : indique le type de prolongement : simple, double, triple.

Schéma explicatif :



3.4.5. AT/AT/GE (type d'association)

Une occurrence de AT/AT/GE associe une occurrence de ATOME-GEO à une occurrence de ATOME.

AT/AT/GE est une association "many-to-one" : plusieurs occurrences de ATOME-GEO peuvent être associées à la même occurrence de ATOME.

L'association AT/AT/GE est facultative pour ATOME et obligatoire pour ATOME-GEO.

3.4.6. GE/AT/GE (type d'association)

Une occurrence de GE/AT/GE associe une occurrence de ATOME-GEO à une occurrence de GEOMETRIE.

L'association GE/AT/GE est "many-to-one" : plusieurs occurrences de ATOME-GEO peuvent être associées à la même occurrence de GEOMETRIE.

L'association GE/AT/GE est obligatoire pour l'entité ATOME-GEO et facultative pour l'entité GEOMETRIE.

3.4.7. GEO/SOM (type d'association)

Une occurrence de GEO/SOM associe une occurrence de GEOMETRIE à une occurrence de SOMMET.

GEO/SOM est une association "one-to-many" : une occurrence de GEOMETRIE peut être associée à plusieurs occurrences de SOMMET.

L'association GEO/SOM est obligatoire pour l'entité SOMMET et facultative pour l'entité GEOMETRIE.

4. CONTRAINTES D'INTEGRITE

Les contraintes "d'exclusion" n'ont pas été prises en charge par le schéma conceptuel ; elles sont donc définies dans le paragraphe suivant.

1. Une occurrence de PROLONG-PROD participant en tant que "précédant" à une occurrence de l'association SUIV-STR ne peut participer en tant que "suivant" ni à cette occurrence de SUIV/STR, ni à aucune autre occurrence de SUIV-STR.
2. Inversément, une occurrence de PROLONG-PROD participant en tant que "suivant" à une occurrence de l'association SUIV-STR ne peut participer en tant que "précédant" ni à cette occurrence de SUIV-STR, ni à aucune autre occurrence de SUIV-STR.
3. Une occurrence de S-STRUCT participant en tant que "père de" à une occurrence de l'association CYCLE-MODIF ne peut participer en tant que "fils de" ni à cette occurrence de CYCLE-MODIF, ni à aucune autre occurrence de CYCLE-MODIF.
4. Inversément, une occurrence de S-STRUCT participant en tant que "fils de" à une occurrence de l'association CYCLE-MODIF ne peut participer en tant que "père de" ni à cette occurrence de CYCLE-MODIF, ni à aucune autre occurrence de CYCLE-MODIF.
5. Une occurrence de SCH-REAC ne peut participer simultanément en tant que "cible" et "origine" à la même association "SR/SR".
6. Si une occurrence de SCH-REAC est "origine" d'une autre occurrence de SCH-REAC, alors elle ne peut être "cible" de cette même occurrence.

7. Une occurrence de COMPOSANT participant à une occurrence de l'association EST STR, ne peut participer à une occurrence de EST ATOME.

8. Une occurrence de PROLONG-STR participant en tant que "précédant" à une occurrence de l'association SUIV-AT ne peut participer en tant que "suivant" ni à cette occurrence de SUIV-AT, ni à aucune autre occurrence de SUIV-AT.

9. Inversément, une occurrence de PROLONG-STR participant en tant que "suivant" à une occurrence de l'association SUIV-AT ne peut participer en tant que "précédant" ni à cette occurrence de SUIV-AT, ni à aucune autre occurrence de SUIV-AT.

10. Une occurrence de PROLONG-STR participant à une occurrence de SUIV-AT ne peut participer à une occurrence de PROL/PROL.

11. Une occurrence de PRODUIT dont l'attribut TYP-PROD = 'N' ne peut participer à aucune association PROD/COMP.

12. Une occurrence de PROLONG-PROD est nécessairement du même TYPE que l'occurrence de PROLONG-STR à laquelle elle est associée.

13. Une occurrence de PROLONG-STR est nécessairement du même TYPE que l'occurrence de SOMMET à laquelle elle est associée.

5. SPECIFICATION DES FONCTIONS DE CONSULTATION

5.1. Limitations du modèle

Une limitation essentielle du système est la détection automatique des groupes fonctionnels "rompus" ou "formés". Il apparaît en effet que cette détection pose de sérieux problèmes : qu'entend-on par "rompus" et "formés" ?

Considérons le schéma réactionnel suivant :



L'examen de ce schéma peut donner lieu à trois interprétations totalement différentes :

1ère interprétation :

Il n'y a aucun changement dans la nature et le nombre de sous-structures ; il n'y a donc aucun groupe fonctionnel FORME ou ROMPU.

2ème interprétation :

Il y a un groupe fonctionnel FORME : Cl

Il y a un groupe fonctionnel ROMPU : Cl

3ème interprétation :

Il y a un groupe fonctionnel FORME : C=O

Il y a un groupe fonctionnel ROMPU : C=O

Laquelle de ces interprétations choisir ?

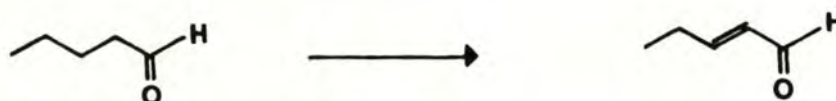
- La première interprétation implique des définitions un peu simplistes des notions "rompues" et "formées" :
 - . un groupe fonctionnel "FORME" existe dans un produit terminal alors qu'il n'existait dans aucun des produits initiaux ;
 - . un groupe fonctionnel "ROMPU" n'existe dans aucun produit terminal alors qu'il existait dans un produit initial.

- Les deuxième et troisième interprétations impliquent les définitions suivantes :
 - . un groupe fonctionnel "FORME" est :
 - .. soit un groupe qui existe dans un produit terminal alors qu'il n'apparaissait dans aucun des produits initiaux ;
 - .. soit un groupe qui n'occupe pas la même position dans le produit initial et terminal ;
 - . un groupe fonctionnel "ROMPU" est :
 - .. soit un groupe qui n'existe dans aucun produit terminal, alors qu'il apparaissait dans un produit initial ;
 - .. soit un groupe qui n'occupe pas la même position dans le produit initial et terminal.

Malheureusement, cette définition n'est pas satisfaisante ; en effet :

- 1) Appliquée à l'exemple précédent, cette définition ne permet pas de déduire la nature des groupements "formés" ou "rompus" : Cl ou C=O ?

- 2) Prenons le cas d'un schéma réactionnel modifiant la structure du squelette moléculaire :



La fonction aldéhyde n'a pas été modifiée ; il est cependant impossible d'affirmer que sa position est inchangée.

5.2. Solution envisageable

En conclusion, il apparaît qu'une détection "intelligente" des modifications de groupes fonctionnels ne peut être réalisée que par le chimiste.

Cette solution laisse au chimiste le soin d'indiquer, lors de l'introduction d'un nouveau schéma réactionnel, quels sont les groupes "formés" et "rompus".

Cette méthode présente l'inconvénient d'être un peu plus contraignante pour l'utilisateur du système.

5.3. Spécification des requêtes

1ère requête :

Afficher les schémas réactionnels qui conduisent à la formation de telles et telles sous-structures.

2ème requête :

Afficher les schémas réactionnels qui conduisent à la rupture de telles et telles sous-structures.

CHAPITRE VI
CONCEPTION LOGIQUE DE LA BASE DE DONNEES

1. BUT DE CETTE ETAPE

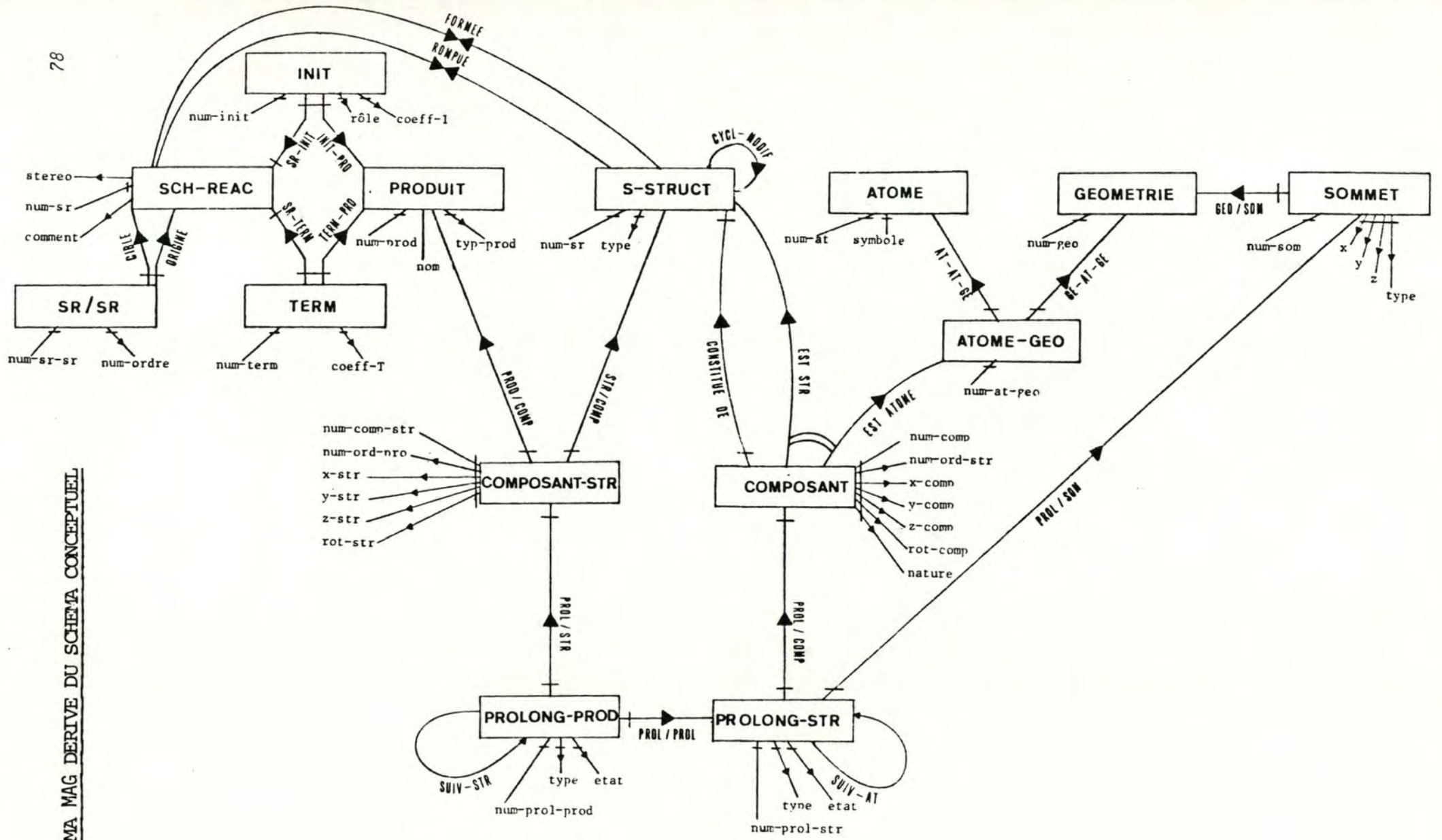
Le but est de fournir une première implémentation :

- a) des structures de données pour un SGBD virtuel de type MAG [37] , [38] ;
- b) des fonctions d'accès dans un langage abstrait : ADL [39] .

Ces structures de données et ces algorithmes pourront, lors d'une étape ultérieure, être traduits pour un SGBD et un langage de programmation réels. ' .

Nous donnons dans les pages suivantes le schéma de la base de données MAG ainsi que les spécifications et les algorithmes des requêtes d'accès.

La description des structures de données du schéma MAG est présentée en annexe.



2. SCHEMA MAG DERIVE DU SCHEMA CONCEPTUEL

3. ENONCE DES REQUETES

3.1. Première requête

Consulter tous les schémas réactionnels qui conduisent à la formation de la sous-structure identifiée par "x".

3.2. Deuxième requête

Consulter tous les schémas réactionnels qui conduisent à la rupture de la sous-structure identifiée par "x".

4. SPECIFICATION DES FONCTIONS UTILISEES PAR LES REQUETES

4.1. CONSULTE-SR (sr)

- Consulter tous les produits initiaux liés au schéma réactionnel référencé par "sr" ;
- consulter tous les produits terminaux liés au schéma réactionnel référencé par "sr" ;
- consulter tous les schémas réactionnels liés au schéma réactionnel référencé par "sr".

4.2. CONSULTE-PRO (p)

- Fournir le nom et le type du produit ;
- consulter le premier COMPOSANT-STR du PRODUIT référencé par p ; (COMPOSANT-STR dont NUM-ORD-PRO=1).
(* les COMPOSANT-STR suivants seront consultés par récursivité sur CONSULTE-COMP-STR (cstr) *).

4.3. CONSULTE-COMP-STR (cstr)

- Fournir les valeurs : x-str , y-str , z-str , rot-str ;
- consulter la sous-structure (S-STRUCT) liée à l'occurrence de COMPOSANT-STR référencée par cstr;
- pour chaque occurrence de PROLONG-PROD (plg) liée à l'occurrence de COMPOSANT-STR référencée par cstr et dont l'ETAT est "saturé" :
 - . Accéder au PROLONG-PROD (plg') qui lui est connecté;
 - . Consulter le COMPOSANT-STR (cstr') auquel appartient plg'. (* récursivité *)

4.4. CONSULTE-STR (str)

- Consulter le premier COMPOSANT de la sous-structure (S-STRUCT) référencée par str; (COMPOSANT dont NUM-ORD-STR = 1);
- (* les COMPOSANTS suivants seront consultés par récursivité sur CONSULTE-COMP (c). *)

4.5. CONSULTE-COMP (c)

- Fournir les valeurs : x-comp , y-comp , z-comp , rot-comp;
- si la NATURE de c est "atome", consulter l'occurrence de ATOME-GEO liée à c;
- si la NATURE de c est "sous-structure", consulter l'occurrence de S-STRUCT liée à c;
- pour chaque PROLONG-STR (plg) lié au COMPOSANT référencé par c et dont l'ETAT est "saturé" :
 - . Accéder au PROLONG-STR (plg') qui lui est connecté;
 - . Consulter le COMPOSANT (c') auquel appartient plg'. (* récursivité *)

4.6. CONSULTE-AT-GEO (at-geo)

- Accéder à l'ATOME "at" lié à at-geo;
- fournir le symbole de l'atome référencé par "at";
- consulter la GEOMETRIE liée à at-geo.

4.7. CONSULTE-GEOM(ge)

- Accéder à chaque occurrence de SOMMET liée à la GEOMETRIE référencée par ge;
- fournir x,y,z,type.

5. ALGORITHMES DES REQUETES ET DES FONCTIONS D'ACCES.

5.1. Première requête

```

input (x) ;
for sr := SCH-REAC (FORMEE : S-STRUCT (:NUM-STR=x)) do
  CONSULTE-SR (sr);
endfor ;

```

5.2. Deuxième requête

```

input (x) ;
for sr := SCH-REAC (ROMPUE : S-STRUCT (:NUM-STR=x)) do
  CONSULTE-SR (sr);
endfor ;

```

5.3. CONSULTE-SR (sr) ;

```

EXPORTE (stereo,comment) ;
n := 1 ; (* initialisation de la variable de boucle
          d'accès aux occurrences de SCH-REAC. *)
(* consultation des produits initiaux : *)
for in := INIT (SR-INIT :SCH-REAC(:NUM-SR = sr)) do
  EXPORTE (rôle,coeff-I) ;
  for pi := PRODUIT(INIT-PRO: INIT(:NUM-INIT=in))do
    CONSULTE-PRO(pi);
  endfor;
endfor;

```

```

(* consultation des produits terminaux : *)
for te := TERM(SR-TERM: SCH-REAC(:NUM-SR=sr)) do
  EXPORTE(coeff-T);
  for pf := PRODUIT(TERM-PRO: TERM(:NUM-TERM=te)) do
    CONSULTE-PRO(pf);
  endfor;
endfor;

(* consultation des schémas intermédiaires : *)
for sr' := SCH-REAC(CIBLE: SR/SR((ORIGINE: SCH-REAC
(:NUM-SR = sr)) and (:NUM-ORDRE=n))) do
  CONSULTE-SR(sr');
  n := n + 1;
endfor;

```

- 5.4. CONSULTE-PRO (p);
 EXPORTE(nom, typ-prod);
 cstr := COMPOSANT-STR(PROD/COMP: PRODUIT(:NUM-PROD = p)
 and (:NUM-ORD-PRO = 1));
 CONSULTE-COMP-STR(cstr);
- 5.5. CONSULTE-COMP-STR (cstr);
 EXPORTE (x-str, y-str, z-str, rot-str);
 str := S-STRUCT(STR/COMP: COMPOSANT-STR(:NUM-COMP-STR=cstr));
 CONSULTE-STR(str);
 for prol := PROLONG-PROD(PROL/STR: COMPOSANT-STR
 (:NUM-COMP-STR = cstr)) do
 EXPORTE(type, etat);
 if ETAT(:PROLONG-PROD(:NUM-PROL-PROD=prol)) = saturé then
 prol' := PROLONG-PROD(SUIV-STR: PROLONG-PROD
 (:NUM-PROL-PROD = prol));
 cstr' := COMPOSANT-STR(PROL/STR : PROLONG-PROD
 (:NUM-PROL-PROD = prol'));
 CONSULTE-COMP-STR(cstr');
 endif;
 endfor;

```

5.6. CONSULTE-STR (str) ;
    c := COMPOSANT (CONSTITUE DE : S-STRUCT (:NUM-STR=str)
                    and (:NUM-ORD-STR = 1)) ;
    CONSULTE-COMP (c) ;
    if TYPE (COMPOSANT(NUM-COMP=c)) = cycle modifié then
        str' := S-STRUCT(CYCL-MODIF:S-STRUCT(:NUM-STR=str)) ;
        CONSULTE-STR (str') ;
    endif ;

5.7. CONSULTE-COMP (c) ;
    EXPORTE (x-comp, y-comp, z-comp, rot-comp) ;
    if nature (c) = "atome" then
        at-geo := ATOME-GEO (EST-ATOME:COMPOSANT (:NUM-COMP=c)) ;
        CONSULTE-AT-GEO (at-geo) ;
    else
        str := S-STRUCT (EST-STR:COMPOSANT (:NUM-COMP=c)) ;
        CONSULTE-STR (str) ;
    endif ;
    for prolg := PROLONG-STR (PROL/COMP:COMPOSANT (:NUM-COMP
                                                    =c)) do
        prolg' := PROLONG-STR (SUIV-AT:PROLONG-STR
                               (:NUM-PROLONG-STR=prolg)) ;
        if EXISTE-prolg then
            c' := COMPOSANT (PROL/COMP:PROLONG-STR
                              (:NUM-PROLONG-STR=prolg')) ;
            CONSULTE-COMP (c') ;
        endif ;
    endfor ;

5.8. CONSULTE-AT-GEO (at-geo) ;
    at := ATOME (AT-AT-GE : ATOME-GEO (:NUM-AT-GEO=at-geo)) ;
    EXPORTE (symbole (:at)) ;
    ge := GEOMETRIE (GE-AT-GE : ATOME-GEO (:NUM-AT-GEO=at-geo)) ;
    CONSULTE-GEOM (ge) ;

5.9. CONSULTE-GEOM (ge) ;
    for s := SOMMET (GEO/SOM : GEOMETRIE (:NUM-GEO=ge)) do
        EXPORTE (x,y,z,type) ;
    endfor ;

```


5.10. Remarques

1. Toutes les données consultées sont exportées vers un module de plus haut niveau qui se charge de l'affichage des schémas réactionnels ; la fonction EXPORTE (résultats) a donc pour effet de renvoyer les résultats au module demandeur.
2. Etant donné la complexité du schéma MAG et la présence de contraintes d'exclusion sur les chemins inter-articles, il a été décidé de décomposer directement les algorithmes prédictifs.
3. La fonction INPUT (argument) se fera par sélection, sur l'écran tactile, de la sous-structure désirée.
4. Dans les spécifications, des raccourcis ont été volontairement pris - lorsque cela était possible - afin d'alléger le "style" des spécifications.
Par exemple :
"Accéder à l'occurrence de ATOME liée à l'occurrence de AT-GEO référencée par "at-geo" a été volontairement simplifié en :
"Accéder à l'ATOME lié à "at-geo".

6. SPECIFICATION DES FONCTIONS D'INTRODUCTION DES DONNEES

6.1. CRE-SOM (num-geo, x, y, z, type) : num-som ;

- Créer occurrence de SOMMET identifiée par "num-som" ;
- attribuer à ce sommet :
 - . les coordonnées "x", "y", "z" ;
 - . le type de liaison "type" ;
- associer ce sommet à l'occurrence de GEOMETRIE référencée par "num-geo".

6.2. CRE-GEO (num-geo) ;

(* "num-geo" est fourni par l'utilisateur, car ce n'est pas un identifiant quelconque, mais une référence aux géométries chimiques SP32, SP34...*).

- Créer une occurrence de GEOMETRIE identifiée par "num-geo".

6.3. CRE-AT (num-at, symbole) ;

- Créer une occurrence de ATOME identifiée par "num-at" ;
- attribuer à cet atome le symbole atomique "symbole".
(* à noter que "symbole" est également identifiant *).

6.4. CRE-AT-GEO (symbole, num-geo) : num-at-geo ;

- Créer une occurrence de ATOME/GEO identifiée par "num-at-geo";
- associer cette occurrence de ATOME/GEO à l'occurrence de ATOME identifiée par "symbole", ainsi qu'à l'occurrence de GEOMETRIE identifiée par "num-geo".

6.5. CRE-PROLONG-STR (num-comp, num-som) : num-prol-str ;

- Créer une occurrence de PROLONG-STR et lui attribuer un identifiant "num-prol-str" ;
- mettre à "libre" l'item ETAT de cet article ;
- associer cette occurrence de PROLONG-STR à l'occurrence de SOMMET référencée par "num-som" ;
- associer cette occurrence de PROLONG-STR à l'occurrence de COMPOSANT référencée par "num-comp".

6.6. CRE-COMPOSANT (num-str, num-ord-str, x,y,z, rot, nature) : num-comp ;

- Créer une occurrence de COMPOSANT, lui attribuer l'identifiant num-comp et l'associer à la sous-structure "num-str" ;
- attribuer à ce composant :
 - . les coordonnées "x,y,z" ;
 - . l'orientation "rot" ;
 - . la nature "nature" ;
 - . le numéro d'ordre "num-ord-str" ;

6.7. CRE-STR : num-str ;

(* la création de sous-structures de type "cycle", "cycle modifié", "cycles juxtaposés", "cycle ponté" devra être traitée séparément *)

1. Créer une occurrence S-STRUCT et lui attribuer l'identifiant "num-str" ;
2. saisir la première occurrence de COMPOSANT (* sous-structure ou atome *) ;
3. créer une occurrence de COMPOSANT par la fonction :
 CRE-COMPOSANT (num-str, num-ord-str, o,o,o,o, nature) :
 num-comp ;

4. Si nature = "atome", alors

- CONSULT-AT-GEO (at-geo) ;
- Associer l'occurrence de COMPOSANT identifiée par "num-comp" à l'occurrence de ATOME/GEO identifiée par "at-geo".
- Pour chaque occurrence de SOMMET associée à l'occurrence de ATOME/GEO identifiée par "at-geo" :
 - . créer une occurrence de PROLONG-STR ;
 - . associer cette occurrence de PROLONG-STR à l'occurrence de COMPOSANT identifiée par "num-comp".

Si nature = "sous-structure", alors

- Associer l'occurrence de COMPOSANT identifiée par "num-comp" à l'occurrence de S-STRUCT identifiée par "str" ;
- Prendre une copie de "str" ; (* c'est-à-dire copier toutes les occurrences de COMPOSANT et PROLONG-STR ainsi que toutes les occurrences des sets EST STR, PROL/COMP, PROL/SOM et SUIV-AT dans lesquels ils interviennent *).

5. Saisir l'occurrence suivante de COMPOSANT (num-comp) et de PROLONG-STR (num-prol-str) (* c'est-à-dire le prolongement à greffer *).

6. Tant qu'il y a des occurrences de COMPOSANT,

- Saisir l'occurrence de PROLONG-STR (num-prol-str') (* c'est-à-dire le prolongement devant recevoir la greffe *).
- Créer une occurrence de COMPOSANT par la fonction : CRE-COMPOSANT (x,y,z,rot,nature) (* où x,y,z,rot sont les coordonnées et l'orientation de "num-prol-str'" *).
- Effectuer le point (4).
- Créer une occurrence de SUIV-AT entre les deux occurrences des PROLONG-STR identifiées par "num-prol-str" et "num-prol-str'".

- Saisir l'occurrence suivante de COMPOSANT (num-comp'') et de PROLONG-STR (num-prol-str'').

6.8. CRE-PROLONG-PROD (num-comp-str, num-prolong-str, état) :
num-prol-prod ;

- Créer une occurrence de PROLONG-PROD ;
- associer cette occurrence de PROLONG-PROD à l'occurrence de PROLONG-STR référencée par "num-prolong-str", ainsi qu'à l'occurrence de COMPOSANT-STR référencée par "num-comp-str".

6.9. CRE-COMPOSANT-STR (num-prod, num-ord-pro, num-str, x,y,z, rot) : num-comp-str ;

- Créer une occurrence de COMPOSANT-STR ;
- lui attribuer l'identifiant "num-comp-str" ;
- l'associer à l'occurrence de S-STRUCT référencée par "num-str" ainsi qu'à l'occurrence de PRODUIT référencé par "num-prod" ;
- attribuer à ce COMPOSANT-STR :
 - . les coordonnées "x,y,z" ;
 - . l'orientation "rot" ;
 - . le numéro d'ordre "num-ord-pro" ;
- CONSULTE-STR (num-str) ;
- pour chaque occurrence de PROLONG-STR dont l'état est "libre" et qui est associée à num-str via les sets (CONSTITUE DE.PROL/COMP), faire :

CRE-PROLONG-PROD (num-comp-str, num-prolong-str,
libre) : num-prolong-prod ;

6.10. CRE-PRODUIT (nom, type) : num-prod ;

- Créer une occurrence de PRODUIT ;
- Si type < > "réactif" alors :
 - saisie de la première occurrence de S-STRUCT ;

- CRE-COMPOSANT-STR (num-prod, l, num-str, o,o,o,o) :
num-comp-str ;
- saisie de l'occurrence suivante de S-STRUCT et de l'occurrence de PROLONG-STR qui va servir à la greffe de la sous-structure ;
- saisie de l'occurrence de PROLONG-PROD sur laquelle il faut greffer (num-prol-prod) ;
- CRE-COMPOSANT-STR (num-prod, num-ord-pro, num-str, x,y,z, rot) : num-comp-str ;
(* . qui fait apparaître les occurrences de PROLONG-PROD associées à chaque occurrence PROLONG-STR ;
. x,y,z, rot sont les valeurs associées à l'occurrence de PROLONG-PROD référencée par num-prol-prod *) ;
- créer une occurrence de SUIV-STR entre les occurrences de PROLONG-PROD référencées par num-prol-prod et num-prol-prod'
(* num-prol-prod' étant une "copie" du prolongement de S-STRUCT servant à la greffe *)).

6.11. CRE-INIT (num-s, num-prod) : num-init ;

- Créer une occurrence de INIT ;
- associer cette occurrence de INIT à l'occurrence de SCH-REAC référencée par num-sr ainsi qu'à l'occurrence de PRODUIT référencée par num-prod.

6.12. CRE-TERM (num-sr, num-prod) : num-term ;

- Créer une occurrence de TERM ;
- associer cette occurrence de TERM à l'occurrence de SCH-REAC référencée par num-sr ainsi qu'à l'occurrence de PRODUIT référencée par num-prod.

6.13. CRE-SCH-REAC : num-sr ;

- Créer une occurrence de SCH-REAC ;
- saisie de la première occurrence de PRODUIT ;
- tant qu'il y a des produits initiaux :
 - . si le produit n'existe pas, CRE-PRODUIT : num-pro ;
 - . CRE-INIT (num-sr, num-pro) : num-init ;
 - . saisie de l'occurrence suivante de PRODUIT ;
- tant qu'il y a des produits terminaux :
 - . si le produit n'existe pas, CRE-PRODUIT : num-pro ;
 - . CRE-TERM (num-sr, num-pro) : num-term ;
 - . saisie de l'occurrence suivante de PRODUIT ;
- tant qu'il y a des schémas intermédiaires :
 - . créer une occurrence de SCH-REAC : num-sr' :
 - . pour chaque occurrence de PRODUIT liée à l'occurrence de SCH-REAC référencée par num-sr via le set composé (SR-TERM.TERM-PRO), faire :
 - CRE-INIT (num-sr', npro) : num-init ;
 - (* les produits "terminaux" du sous-schéma précédent deviennent produits "initiaux" dans le sous-schéma actuel *) ;
 - . saisie de l'occurrence suivante de PRODUIT ;
 - . tant qu'il y a des produits initiaux :
 - si le produit n'existe pas, CRE-PRODUIT / num-pro ;
 - CRE-INIT (num-sr', num-pro) : num-init ;
 - saisie de l'occurrence suivante de PRODUIT.
 - . tant qu'il y a des produits terminaux :
 - . si le produit n'existe pas, CRE-PRODUIT : num-pro ;
 - . CRE-TERM (num-str', num-pro) : num-term ;
 - . saisie de l'occurrence suivante de PRODUIT.

CHAPITRE VII

CONCLUSIONS

1. PERFORMANCES D'UN TEL SYSTEME

L'aspect performance n'a pas été abordé au cours de ce travail.

Bien que l'étude n'ait pas été menée jusqu'à la conception physique du système, il est possible, moyennant quelques hypothèses, d'évaluer les performances d'accès à la base de données.

Les hypothèses que nous avons adoptées sont les suivantes :

- a) Le SGBD est de type relationnel ;
- b) Seuls les fichiers "ROMPUS", "FORMEE", "INIT", "TERM" sont accédés par index ; les autres fichiers sont consultés en "accès direct".

La sélection des schémas réactionnels répondant à tel critère ne pose aucun problème de performance. En effet, la sélection des critères de recherche se fait par pointage des sous-structures (à former ou à rompre) sur l'écran tactile ; ce pointage a pour effet de renvoyer les numéros de sous-structures au module d'accès à la base de données. Les schémas réactionnels sont alors sélectionnés par accès indexé aux fichiers "ROMPUS" ou "FORMES".

C'est la consultation des schémas réactionnels qui est très coûteuse en nombre d'accès à la base de données : il apparaît en effet que l'affichage d'un graphe moléculaire nécessite, en moyenne, 10 accès à la base de données par atome (autre que l'Hydrogène) !

On peut proposer quelques solutions pour améliorer les performances de consultation :

1°/ Il apparait que certains fichiers sont de taille fixe et limitée ; c'est le cas pour :

- a) "SOMMET" : le nombre de types de sommets est 9 :
- 4 sommets pour SP_3 ;
 - 3 sommets pour SP_2 ;
 - 2 sommets pour SP .
- b) "GEOMETRIE" : le nombre de géométries envisageables est inférieur à 15. Les plus courantes sont SP_{34} , SP_{32} , SP_{31} , SP_{24} , SP_{22} , SP ; à cette liste, on peut ajouter les géométries correspondant aux formes chargées et radicalaires des atomes. Ce sont les géométries : SP_{3R} , SP_{3+} , SP_{3-} ...
- c) "ATOME" : le fichier complet contient 103 occurrences d'atomes. Cependant, seules quelques occurrences seront fréquemment utilisées ; il s'agit de :
- C, H, O, N,
 - P, S, Se,
 - F, Cl, Br, I,
 - B, Al, Li, Na, K, Mg, Ca.
- d) "ATOME/GEO" : pour chaque occurrence de ces atomes, il faut considérer ses "géométries" les plus fréquentes.

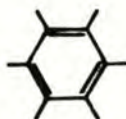
Par exemple :

ATOME GEOMETRIE	C	H	O	N
SP_{34}	°			
SP_{33}				°
SP_{32}			°	
SP_{31}		°		
SP_{24}	°			
SP_{23}				°
SP_{22}			°	
SP_{14}	°			
SP_{13}				°

Chaque point du tableau représente une occurrence de ATOME/GEO.

Il ne semble pas déraisonnable d'envisager le stockage en mémoire centrale de ces fichiers, dans leur entièreté ou en partie ; ceci permettrait de ramener à 4 par atome le nombre d'accès à la mémoire secondaire.

2°/ L'affichage graphique de chaque sous-structure nécessite l'accès à tous ses "COMPOSANTS" via ses prolongements "PROLONG-STR" ; par exemple, l'affichage d'un noyau "benzène" nécessiterait une vingtaine d'accès à la mémoire secondaire et ce, malgré les améliorations envisagées au point (1°) :



Il serait souhaitable de stocker les sous-structures les plus fréquentes sous une forme plus "groupée" : par exemple, l'utilisation de "tables de connexions" permettrait de consulter une telle sous-structure en un seul accès à la mémoire secondaire.

2. ASPECTS NON DEVELOPPES

2.1. Certains aspects du problème n'ont pas été traités. Il s'agit essentiellement de la spécification :

- des fonctions d'affichage graphique ;
- des fonctions de saisie "tactile" ;
- des fonctions de recherche d'isomorphisme (identité de deux motifs structuraux) ;
- des fonctions de recherche d'homomorphisme (inclusion d'un motif structural dans un autre) ;

- du traitement de sous-structures particulières telles que : -- cycles modifiés ;
 - cycles pontés ;
 - cycles juxtaposés.

2.2. D'autres aspects du problème ont été fortement - trop - simplifiés ; c'est le cas :

- de la détection des groupes fonctionnels "FORMES" et "ROMPUS", qui n'a pu être automatisée (cf. Chapitre V, 5.1.) ;
- de la visualisation tridimensionnelle des cycles : il serait intéressant de pouvoir définir interactivement de nouveaux cycles. Cependant, le calcul des coordonnées spatiales des sommets d'un cycle est très complexe : aussi, la solution adoptée consiste à introduire "manuellement" les coordonnées de chaque sommet.

ANNEXE 1Description associée au schéma MAG

database SYNTORG

file	F-SCH-REAC	contains	SCH-REAC
	F-PRODUIT		PRODUIT
	F-INIT		INIT
	F-TERM		TERM
	F-SR/SR		SR/SR
	F-S-STRUCT		S-STRUCT
	F-COMPOSANT-STR		COMPOSANT-STR
	F-PROLONG-PROD		PROLONG-PROD
	F-COMPOSANT		COMPOSANT
	F-PROLONG-STR		PROLONG-STR
	F-ATOME		ATOME
	F-GEOMETRIE		GEOMETRIE
	F-ATOME-GEO		ATOME-GEO
	F-SOMMET		SOMMET

record-type SCH-REAC
 identifier NUM-SR
 2 NUM-SR
 2 STEREO
 2 COMMENT

record-type PRODUIT
 identifier NUM-PROD
 2 NUM-PROD
 2 TYP-PROD
 2 NOM

record-type INIT

identifier NUM-INIT
identifier within SR-INIT, INIT-PRO
2 NUM-INIT
2 rôle
2 COEFF-I

record-type TERM

identifier NUM-TERM
identifier within SR-TERM, TERM-PRO
2 NUM-TERM
2 COEFF-T

record-type SR/SR

identifier NUM-SR-SR
identifier within CIBLE, ORIGINE
2 NUM-SR-SR
2 NUM-ORDRE

record-type S-STRUCT

identifier NUM-STR
2 NUM-STR
2 TYPE

record-type COMPOSANT-STR

identifier NUM-COMP-STR
2 NUM-COMP-STR
2 NUM-ORD-PRO
2 X-STR
2 Y-STR
2 Z-STR
2 ROT-STR

record-type PROLONG-PROD

 identifier NUM-PROL-PROD

 2 NUM-PROL-PROD

 2 TYPE

 2 ETAT

record-type COMPOSANT

 identifier NUM-COMP

 2 NUM-COMP

 2 NUM-ORD-STR

 2 X-COMP

 2 Y-COMP

 2 Z-COMP

 2 ROT-COMP

 2 NATURE

record-type PROLONG-STR

 identifier NUM-PROL-STR

 2 NUM-PROL-STR

 2 TYPE

 2 ETAT

record-type ATOME

 identifier NUM-AT

 identifier SYMBOLE

 2 NUM-AT

 2 SYMBOLE

record-type GEOMETRIE

 identifier NUM-GEO

 2 NUM-GEO

record-type ATOME-GEO

 identifier NUM-AT-GEO

 2 NUM-AT-GEO

record-type SOMMET

 identifier NUM-SOM

 2 NUM-SOM

 2 X

 2 Y

 2 Z

 2 TYPE

access-path-type ORIGINE

 from one SCH-REAC

 to many SR/SR total

 order sorted by NUM-ORDRE

access-path-type CIBLE

 from many SR/SR total

 to one SCH-REAC

access-path-type SR-INIT

 from one SCH-REAC total

 to many INIT total

 order last

access-path-type INIT-PRO

 from many INIT total

 to one PRODUIT

access-path-type SR-TERM

 from one SCH-REAC total

 to many TERM total

 order last

access-path-type TERM-PRO

 from many TERM total

 to one PRODUIT

access-path-type FORMEE
from many S-STRUCT
to many SCH-REAC
order sorted by NUM-SR

access-path-type ROMPUE
from many S-STRUCT
to many SCH-REAC
order sorted by NUM-SR

access-path-type PROD/COMP
from one PRODUIT
to many COMPOSANT-STR total
order sorted by NUM-ORD-PRO

access-path-type STR/COMP
from many COMPOSANT-STR total
to one S-STRUCT

access-path-type PROL/STR
from one COMPOSANT-STR total
to many PROLONG-PROD total
order sorted by NUM-PROL-PROD

access-path-type SUIV-STR
from one PROLONG-PROD
to one PROLONG-PROD

access-path-type CONSTITUE DE
from one S-STRUCT total
to many COMPOSANT total
order sorted by NUM-ORD-STR

access-path-type CYCL-MODIF
from many S-STRUCT
to one S-STRUCT

access-path-type PROL/COMP
from one COMPOSANT total
to many PROLONG-STR total
order sorted by NUM-PROL-STR

access-path-type EST STR
from many COMPOSANT
to one S-STRUCT

access-path-type EST ATOME
from many COMPOSANT
to one ATOME-GEO

access-path-type AT-AT-GE
from many ATOME-GEO total
to one ATOME

access-path-type GE-AT-GE
from many ATOME-GEO total
to one GEOMETRIE

access-path-type GEO/SOM
from one GEOMETRIE
to many SOMMET total
order sorted by NUM-SOM

access-path-type PROL/SOM
from many PROLONG-STR total
to one SOMMET

access-path-type SUIV-AT
from one PROLONG-STR
to one PROLONG-STR

access-path-type PROL/PROL
from many PROLONG-PROD total
to one PROLONG-STR

ANNEXE 2

Architecture du système.

L'objectif final du travail est d'implémenter le système décrit sur un micro-ordinateur de type HP150 doté d'un "Touch-Screen".

Nous proposons dans les pages suivantes une ébauche d'architecture du système complet. Nous ne reviendrons pas sur l'aspect "base de données" -qui a été largement développé aux chapitres 5 et 6 ; par contre, nous aborderons succinctement les aspects liés à l'utilisation de l'écran graphique, ainsi qu'à l'introduction des données via le "touch-screen".

Outre les modules "Base de Données" ,le système proposé comportera les modules suivants :

1. Module "COORDINATEUR"

Le rôle de ce module est d'assurer la séquence correcte des fonctions des autres modules (appel de fonctions avec passage d'arguments et retour des résultats) .

Les principales fonctions associées à ce module seront:

INTRO-SCHEMA-REACTIONNEL : assurer la succession des fonctions permettant à l'utilisateur d'introduire un nouveau schéma réactionnel (succession des écrans, gestion de la BD, gestion du touch-screen) .

SUPPRESS-SCHEMA-REACTIONNEL : succession des fonctions permettant la suppression d'un schéma réactionnel.

RECHERCHE-SSTR : succession des fonctions permettant la recherche de schémas réactionnels conduisant à la rupture (ou à la formation) de telle et telle sous-structure .

RECHERCHE-PRODUIT : succession des fonctions permettant la recherche de schémas réactionnels conduisant à la formation (ou à la rupture) de tel et tel produit .

Des fonctions analogues peuvent être imaginées pour les autres éléments de la base de données : PRODUITS , SOUS-STRUCTURES, ATOMES , GEOMETRIES, SOMMETS.

2. Module "OPERATIONS-LISTES"

Son rôle est d'effectuer des opérations particulières sur des listes fournies par le module COORDINATEUR. Les fonctions associées à ce module seront:

INTER-LISTE (l_1, l_2) : l_3 : l_3 sera le résultat de l'intersection des listes l_1 et l_2 .

UNION-LISTE (l_1, l_2) : l_3 : l_3 sera le résultat de l'union des listes l_1 et l_2 .

3. Module "RECONNAISSANCE DE MOTIFS"

Son rôle sera de vérifier l'identité de deux PRODUITS (isomorphisme) ou morceaux de PRODUITS (homomorphisme). Les fonctions associées à ce module seront:

IDENTITE-PRODUIT (p_1, p_2) : result : "result" prend la valeur "vrai" si p_1 et p_2 sont identiques , et "faux" sinon.

IDENTITE-S-STRUCT (str ,p) : result : "result" prend la valeur "vrai" si str est contenu dans p, et "faux" sinon.

4. Module "SAISIE"

Le rôle de ce module sera de gérer les écrans de saisie: écran graphique , touch-screen ,écran alphanumérique.

5. Module "GRAPHIQUE"

Ce module se chargera de l'affichage de sous-structures à l'écran graphique. La fonction associée à ce module sera :

AFFICH-STR (num-str , x,y,z,rot) : status : affiche la sous-structure "num-str" en lui appliquant la translation "x,y,z" et la rotation "rot". "status" a la valeur 1 en cas d'échec , et 0 en cas de succès.

6. Module "OPERATIONS GRAPHIQUES"

Ce module se chargera des manipulations de molécules dans l'espace ; les fonctions associées à ce module seront:

SCALE (x,y,z,s) : (x'y'z' status) : applique le facteur d'échelle "s" au point de coordonnées (x,y,z) ; le résultat est fourni dans x' ,y' ,z' . "status" prend la valeur 1 en cas d'échec et 0 en cas de succès.

TRANS (x,y,z,a,b,c) : (x'y'z' status) : applique la translation "a,b,c" au point de coordonnées "x,y,z" ; le résultat est fourni dans x' ,y' , z' . "status" prend la valeur 1 en cas d'échec et 0 en cas de succès.

ROTATION (G, u,v,w, u'v'w' ,x,y,z) : (x'y'z' status) : applique la rotation G ,autour de l'axe défini par les points de coordonnées u,v,w, et u'v'w' , au point de coordonnées x,y,z ; le résultat est fourni dans x'y'z' ; "status" prend la valeur 1 en cas d'échec,et 0 en cas de succès.

PROJECTION (x,y,z,D) : (x'y' ,status) :effectue la projection du point de coordonnées x,y,z , sur l'écran; D représente la distance observateur-écran. Le résultat est fourni dans x'y' ; "status" prend la valeur 1 en cas d'échec et 0 en cas de succès.

CONVERSION-ALPHA-GRAPH (x,y) : (x'y',status) :effectue le passage des coordonnées du TOUCH-SCREEN (x , y) aux coordonnées de l'écran GRAPHIQUE (x'y') . "status" prend la valeur 1 en cas d'échec et 0 en cas de succès.

CONVERSION-GRAPH-ALPHA (x,y) : (x'y' ,status) :effectue le passage des coordonnées de l'écran GRAPHIQUE (x , y) aux coordonnées du TOUCH-SCREEN (x'y') . "status" prend la valeur 1 en cas d'échec et 0 en cas de succès.

CONTROLE-FENETRE (a,b,a',b',x,y):status : "status" prend la valeur 0 si le point de coordonnées (x,y) se trouve à l'intérieur de la fenêtre définie par le coin supérieur gauche (a,b) et le coin inférieur droit (x'y') ; "status" prend la valeur 1 sinon.

7. Module "MATRICE"

Ce module offrira des fonctions d'opérations matricielles. Seule la multiplication matricielle sera nécessaire pour les opérations graphiques.

MULTIP(m₁,m₂,dim) : (m₃ ,status) : "m₁" et "m₂" sont les deux matrices à multiplier ; "dim" est la dimension de ces matrices ; le résultat est fourni dans "m₃" ; "status" prend la valeur 0 en cas de réussite et 1 en cas d'échec.

8. Module "GRAPH-SCREEN"

Ce module permettra l'utilisation de l'écran graphique du HP150.

Les fonctions de ce module seront :

CLEAR-GRAPH-SCREEN :status	:assure le nettoyage de l'écran graphique.
ON-GRAPH-SCREEN :status	:rend actif l'écran graphique.
OFF-GRAPH-SCREEN :status	:fait passer l'écran en mode alphanumérique.
UP-PEN:status	:lève le crayon.
DOWN-PEN:status	:abaisse le crayon.
MOVE-ABS(x,y):status	:déplace le curseur jusqu'au point (x,y).
MOVE-REL(x,y):status	:effectue une translation (x,y) du curseur par rapport à sa position initiale.
SET-ORIGINE(x,y):status	:positionne l'origine des axes de l'écran au point de coordonnées (x,y).

Pour toutes ces fonctions, "status" prend la valeur 0 en cas de succès et 1 en cas d'échec.

9. Module "TOUCH-SCREEN"

Ce module aura pour rôle la définition et la lecture de champs tactiles à l'écran.

Les fonctions de ce module seront :

DEFINE-TOUCH(x,y,x',y',beep,cursor,brill-off,brill-on,
long,point):status

x,y sont les coordonnées du coin supérieur droit du champ
à définir;

x',y' sont les coordonnées du coin inférieur gauche;

"beep" indique si un signal sonore doit accompagner la
lecture du champ ;

"cursor" indique si le curseur doit se positionner sur le
champ lors de la lecture de celui-ci;

"brill-off" indique la brillance du champ avant lecture;

"brill-on" indique la brillance du champ après lecture;

"long" indique la longueur du string à afficher sur
le champ lors de la lecture de celui-ci;

"point" est un pointeur vers ce string.

EFFACER-TOUCH(ligne,colonne):status :effacer le champ
tactile auquel le point défini par "ligne" et "colonne"
appartient.

ON-TOUCH-SCREEN:status :allume l'écran tactile et le
rend actif.

OFF-TOUCH-SCREEN :status :éteint l'écran tactile et le
rend inactif.

PREMIERE-PAGE :status :affiche la première page de
l'écran tactile.

SECONDE-PAGE :status :affiche la seconde page de
l'écran tactile.

10. Module "ALPHA-SCREEN"

Ce module fournira les fonctions nécessaires à la gestion de l'écran alphanumérique.

Les fonctions associées à ce module seront:

ALPHA-MODE :status :renvoie le status de l'écran
(1:alphanumérique ;2:graphique;)

CADRE(x,y,x',y'):status :trace un cadre passant par les points de coordonnées x,y et x',y'.

GOTO(x,y):status :positionne le curseur au point de coordonnées x,y.

ECRITURE(texte) : status :écriture du texte à partir de la position du curseur.

EFFACER-ECRAN:status :efface l'écran alphanumérique.

Pour toutes ces fonctions , "status" prend la valeur 0 en cas de succès et 1 en cas d'échec.

11. Module "ACCES-BD"

Le rôle de ce module est de cacher les primitives réelles du système de gestion de fichiers; on y trouvera des primitives telles que :

ACCES-SCH-REAC(num-sr):status :accès direct au schéma réactionnel identifié par "num-sr".

ACCES-PREM-ROMPU(nstr):status :accès au premier "ROMPU" référencé par "nstr".

BIBLIOGRAPHIE

- [1] E.J. COREY, W.T. WIPKE.
"Computer-assisted design of complex organic syntheses".
SCIENCE, 166, (1969), p. 179.
- [2] E.J. COREY.
"Computer-assisted analysis of complex synthetic problems".
CHEMICAL SOCIETY REVIEW - Centenary Lecture 1972.
- [3] E.J. COREY, W. JEFFREY HOWE, H.W. ORF, D.A. PENSAK, G. PETERSON.
"General methods of synthetic analysis. Strategic bond disconnections for bridged polycyclic structures".
Journal of the American Chemical Society, 97, 21, (1975) p. 6116.
- [4] E.J. COREY, W.L. JORGENSEN.
"Computer-assisted synthetic analysis. Generation of synthetic sequences involving sequential functional group interchanges".
J. Amer. Chem. Soc., 98, 1, (1976), p. 202.
- [5] E.J. COREY, W.L. JORGENSEN.
"Computer-assisted synthetic analysis. Synthetic strategies based on appendages and the use of reconnection transforms."
J. Amer. Chem. Soc., 98, 1, (1976), p. 189.
- [6] E.J. COREY, R.D. CRAMER, W.J. HOWE.
"Computer-assisted synthetic analysis for complex molecules. Methods and procedures for machine generation of synthetic intermediates".
J. Amer. Chem. Soc., 94, 2, (1972), p. 440.
- [7] E.J. COREY, W.T. WIPKE, R.D. CRAMER, W.J. HOWE.
"Computer-assisted synthetic analysis. Facile man-machine communication of chemical structure by interactive computer graphics".
J. Amer. Chem. Soc., 94, 2, (1972), p. 421.
- [8] E.J. COREY, W.T. WIPKE, R.D. CRAMER, W.J. HOWE.
"Techniques for perception by a computer of synthetically significant structural features in complex molecules".
J. Amer. Chem. Soc., 94, 2, (1972), p. 431.

- [9] E.J. COREY, H.W. ORF, D.A. PENSAK.
"Computer-assisted synthetic analysis. The identification and protection of interfering functionality in machine-generated synthetic intermediates".
J. Amer. Chem. Soc., 98, 1, (1976), p. 210.
- [10] M. BERSOHN, A. ESACK.
"Computers and Organic Synthesis".
Chemical Reviews, 76, 2, (1976), p. 269.
- [11] H.L. GELERNTER, A.F. SANDERS, D.L. LARSEN, K.K. AGARWAL, R.H. BOIVIE, G.A. SPRITZER, J.E. SERLEMAN.
"Empirical Explorations of Synchem".
SCIENCE, 197, 4308, (1977), p. 1041.
- [12] P.E. BLOWER, H.W. WHITLOCK.
"An Application of Artificial Intelligence to Organic Synthesis".
J. Amer. Chem. Soc., 98, 6, (1976), p. 1499.
- [13] H.W. WHITLOCK.
"A heuristic solution to the functional group switching problem in organic synthesis".
J. Amer. Chem. Soc., 98, 11, (1976), p. 3224.
- [14] B.J. COHEN.
"User-oriented approach to a computerized organic reaction catalog".
J. Chem. inf. Comput. Sci., 22, (1982), p. 195.
- [15] T.D. SALATIN, W.L. JORGENSEN.
"Computer-assisted mechanistic evaluation of organic reactions. 1. Overview".
J. Org. Chem., 45, 11, (1980), p. 2043.
- [16] D. CHODOSH, W.L. MENDELSON.
"Synthesis Library. Graphics at the bench".
Drug Information Journal, 17, (1983), p. 231.
- [17] D. SAMYN.
"Mise au point d'outils en synthèse assistée sur ordinateur".
Mémoire de fin d'études. FUN 1977.
- [18] J.B. HENDRICKSON.
"A systematic characterization of structures and reactions for use in organic synthesis".
J. Amer. Chem. Soc., 93, 25, (1971), p. 6847.
- [19] J.B. HENDRICKSON, E. BRAUN-KELLER.
"Systematic synthesis design. Generation of reaction sequences".
Journal of Computational Chemistry, 1, 4, (1980), p. 323.

- [20] J.B. HENDRICKSON.
"Systematic synthesis design. IV. Numerical codification of construction reactions".
J. Amer. Chem. Soc., 97, 20, (1975), p. 5784.
- [21] O. SCHIER, W. NÜBLING, W. STEIDLE, J. WALLS.
"A system for the documentation of chemical reactions".
Angew. Chem. internat. edit., 9, 8, (1970), p. 599.
- [22] E.J. COREY, G.A. PETERSON.
"An algorithm for machine perception of synthetically significant rings in complex cyclic organic structures".
J. Amer. Chem. Soc., 94, 2, (1972), p. 460.
- [23] E.J. COREY, W.J. HOWE, D.A. PENSAK.
"Computer-assisted synthetic analysis. Methods for machine generation of synthetic intermediates involving multistep Look-Ahead".
J. Amer. Chem. Soc., 96, 25, (1974), p. 7724.
- [24] E.J. COREY, A.P. JOHNSON, A.K. LONG.
"Computer-assisted synthetic analysis. Techniques for efficient long-range retrosynthetic searches applied to the Robinson annulation process".
J. Org. Chem., 45, 11, (1980), p. 2051.
- [25] W.H. NEWMAN, R.F. SPROULL.
"Principles of Interactive Computer Graphics".
Mc Graw-Hill.
- [26] J.B. HENDRICKSON.
"Systematic Synthesis Design. III. The scope of the problem".
J. Amer. Chem. Soc., 97, 20, (1975), p. 5763.
- [27] J.E. DUBOIS.
"Ordered chromatic graph and limited environment concept".
J. Chem. Soc., 13, 8, (1973).
- [28] W.C. HERNDON.
"Canonical labelling and linear notation for chemical graphs".
Chemical Applications of topology and graph theory, 28, p. 231, Elsevier.
- [29] J.B. HENDRICKSON, A.G. TOCZKO.
"Unique numbering and cataloging of molecular structures".
J. Chem. inf. Comput. Sci., 23, 4, (1983), p. 169.

- [30] M.M. CONE, and Al.
"Molecular structure comparison program for the identification of maximal common substructures".
J. Amer. Chem. Soc., 93, 23, (1977), p. 7668.
- [31] J.J. Mc GREGOR.
"Backtrack search algorithms and the maximal common subgraph problem".
Software - practice and experience, 12, p. 23, (1982).
- [32] J.J. Mc. GREGOR.
"Use of a maximal common subgraph algorithm in the automatic identification of the ostensible bond changes occurring in chemical reactions".
J. Chem. inf. Comput. Sci., 21, (1981), p. 137.
- [33] M.F. LYNCH and Al.
"Computers storage and retrieval of generic chemical structures in patents. 1. Introduction and general strategy".
J. Chem. inf. Comput. Sci., 21, (1981), p. 148.
- [34] M.F. LYNCH and Al.
"Computers storage and retrieval of generic chemical structures in patents. 2. GENSAL, a formal language for the description of generic chemical structures".
J. Chem. inf. Comput. Sci., 21, (1981), p. 151.
- [35] M.F. LYNCH and Al.
"Computers storage and retrieval of generic chemical structures in patents. 3. Chemical grammars and their role in the manipulation of chemical structures".
J. Chem. inf. Comput. Sci., 21, (1981), p. 161.
- [36] M.F. LYNCH and Al.
"Computers storage and retrieval of generic chemical structures in patents. 4. An extended connection table representation for generic structures".
J. Chem. inf. Comput. Sci., 22, (1982), p. 160.
- [37] J.L. HAINAUT.
"Un modèle de bases de données au niveau organique : le modèle d'accès".
(Notes de cours ; 1981).
- [38] J.L. HAINAUT.
"Cadre de référence pour la conception de bases de données".
(Notes de cours ; 1983).

- [39] J.L. HAINAUT.
"ADL : un langage de désignation de données".
(Notes de cours ; 1983).
- [40] B.G. BUCHANAN, E.A. FEIGENBAUM.
"DENDRAL and META-DENDRAL".
Artificial Intelligence, 11, (1978), pp. 5-24.