République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

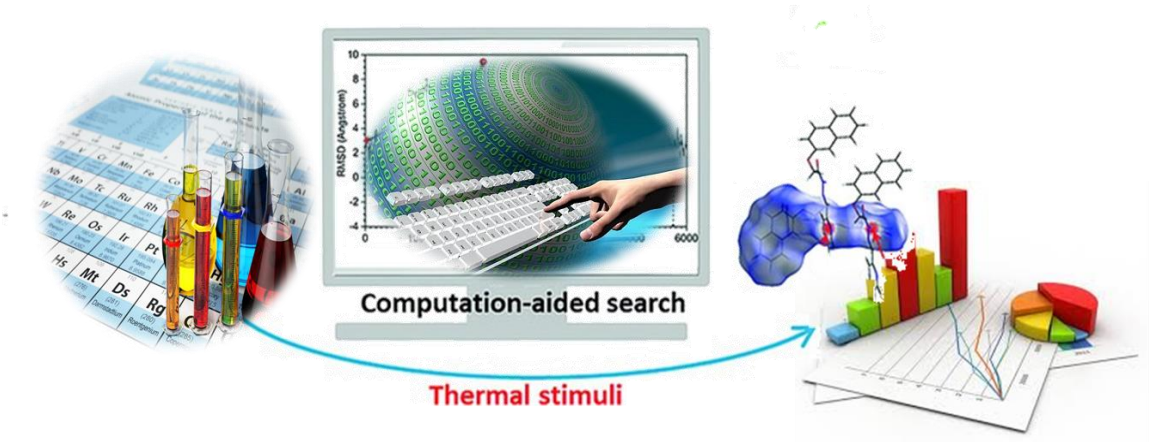
Université Mohamed khider – Biskra - Algérie

Sciences exactes et des sciences de la nature Et de la vie

Département des sciences de la matière

Polycopié de Cours

Informatique pour la chimie



Cours et travaux pratiques

Cours destinés aux étudiants de 3ème année licence chimie analytique

Dr. Hanane DJOUAMA

Laboratoire de Chimie Appliquée-LCA

Courriel : h.djouama@univ-biskra.dz

2019/2020

Avant-propos

Ce polycopié est un support de cours destiné aux étudiants de la 3ème année chimie analytique. La structure et le contenu des chapitres et leur application (TP) de ce document sont synchronisés avec le nouveau contenu du programme établi dans le canevas de l'offre de formation.

Le polycopié est subdivisé en deux parties :

Partie 1: le cours s'articule autour de sept chapitres.

L'objectif premier chapitre de ce cours est de présenter les notions de base dans le domaine d'informatique, de manipuler les outils de bureautique. Le deuxième chapitre aborde l'ntroduction aux systèmes d'exploitation type Unix/Linux. Le troisième chapitre a pour but d'initier les étudiants aux principes de base de la statistique. Ce chapitre a pour but de présenter les principes de base d'une analyse statistique de données. Le quatrième chapitre traitent l'étude de banques de données chimiques indexées par structure. Ce chapitre permet d'initier l'étudiant aux bases de données d'une manière aussi simple et claire que possible vu son importance pour la chimie, tout en essayant de donner une vision sur ce que c'est une base de données chimiques indexées par structure. Le cinquième chapitre présente la méthodologie de la recherche d'informations en chimie. Le sixième chapitre est dédié aux applications locales; représentation de la structure 3D. Le dernier chapitre est réservé aux initiation à la modélisation moléculaire.

Partie 2: les travaux pratiques

Les séances de travaux pratiques (TP) sont complémentaires à tout enseignement théorique. Notons que les travaux pratiques se déroulent sur des micro-ordinateurs fonctionnant sous le système d'exploitation Windows de Microsoft. Les logiciels installés sur ces machines permettent de se familiariser avec le traitement de texte, les tableurs et différents logiciels spécialisés dans la modélisation moleculaire.

J'ai clôturé ce polycopié par une bibliographie qui englobe les ressources mentionnées dans l'ouvrage.

Sommaire

Cours

Objectifs de cours

La place de ce cours dans le future métier des étudiants:

-Découvrir la signification du terme chimie informatique.

-Comprendre les fondements de la Statistique.

-Analyse des données (d'étudier un phénomène à partir de données).

-Visualisation et dessin des molécules à partir de données structurales.

- Utilisation de banques de données pour identifier les systèmes moléculaire .

-Comparer les méthodes de modélisation quantiques et classiques.

Initiation aux outils informatiques appliqués au domaine de la chimie.

1. Outils informatiques utilisés en chimie

L'informatique joue un rôle croissant dans la recherche en chimie. Des secteurs très variés de la recherche fondamentale ou appliquée nécessitent des spécialités du traitement informatique, de l'information chimique, de la modélisation moléculaire ou de la chimie théorique.

La chimie se prête à un traitement informatique car elle est complexe et nécessite des capacités d'acquisition, de traitement et d'archivage considérables. D'importantes bases de données se constituent à travers le monde pour permettre aux chercheurs de suivre quasiment en temps réel l'avancement de la chimie.

2. Informatique

L'informatique est une branche qui s'occupe du domaine du traitement automatique de l'information.

Le terme informatique est composé de deux mots : information et automatique. Vient des mots information automatique donc l'informatique est le domaine d'activité scientifique, technique et industriel concernant le traitement automatique de l'information par des machines.

L'informatique a pour rôle :

-La conception et la construction des ordinateurs,

-Le fonctionnement et la maintenance des ordinateurs,

-Leur exploitation (utilisation des ordinateurs dans les différents domaines d'activités).

3. Système d'information

Le système informatique constitue l'infrastructure technique du système d'information de l'organisation. Ce qu'on appelle système d'information de l'organisation comporte, outre le système informatique, un ensemble organisé d'autres ressources, humaines, organisationnelles et immatérielles, comme des méthodes, des règles, des procédures, etc. Le système d'information est

destiné à faciliter le fonctionnement de l'organisation en lui fournissant les informations utiles pour atteindre ses objectifs.

L'information est un renseignement sur une personne, un objet ou un événement. Il peut être présenté sous différente forme : texte, image, son, etc. Le traitement d'information est une suite logique d'actions qui permettent de transformer des données en résultats.

4. Le système informatique

Un système informatique est le composé de deux parties : Matériels (Hardware) et logiciels (Software).

4.1. Matériels (Hardware)

L'aspect matériel se rapporte aux composants physiques de l'ordinateur : écran, souris, clavier, disque dur, etc. Le matériel informatique est un ensemble de composants électroniques destiné à (voir Figure 1) :

-Acquérir des informations extérieures (les unités d'entrée).

-Faire des traitements (l'unité de traitement).

-Mémoriser des informations (les unités de stockage).

-Fournir le résultat des traitements effectués (les unités de sortie).

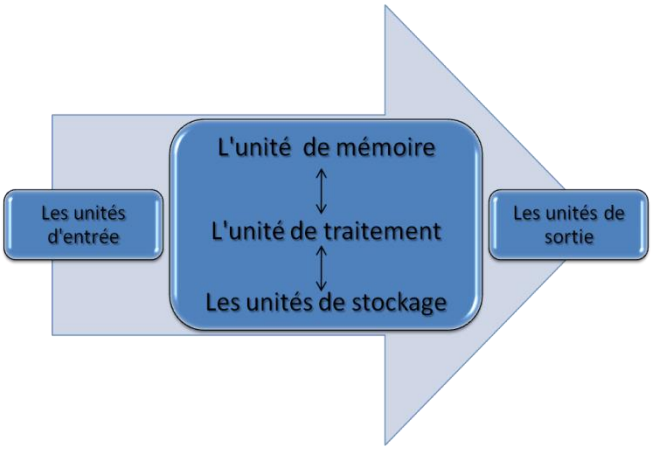


***Figure 1**: Le schéma général du matériel informatique.*

4.1.1. Brève description d'un ordinateur

La figure 2 présente l'organisation générale d'un ordinateur. On y trouve deux parties principales:

-le processeur comprenant les modules mémoire centrale, processeur central (microprocesseur), les unités d'échange et le bus de communication entre ces différents modules;

-les périphériques avec lesquels dialogue le processeur au travers des unités d'échange (ou contrôleurs). On distingue en général :

- les périphériques d'entrée tels que le clavier ou la souris;
- les périphériques de sortie tels que les imprimantes et les écrans de visualisation;
-les périphériques d'entrée et de sortie tels que les disques magnétiques ou les modems pour accéder aux réseaux de communication.

Globalement le processeur permet l'exécution d'un programme. Chaque processeur dispose d'un langage de programmation (les instructions machine) spécifique. Ainsi résoudre un problème avec un processeur consiste à exprimer ce problème comme une suite de ses instructions machine. La solution à un problème est donc spécifique de chaque processeur. Le programme machine et les données qui sont manipulées par les instructions machine sont placés dans la mémoire centrale.

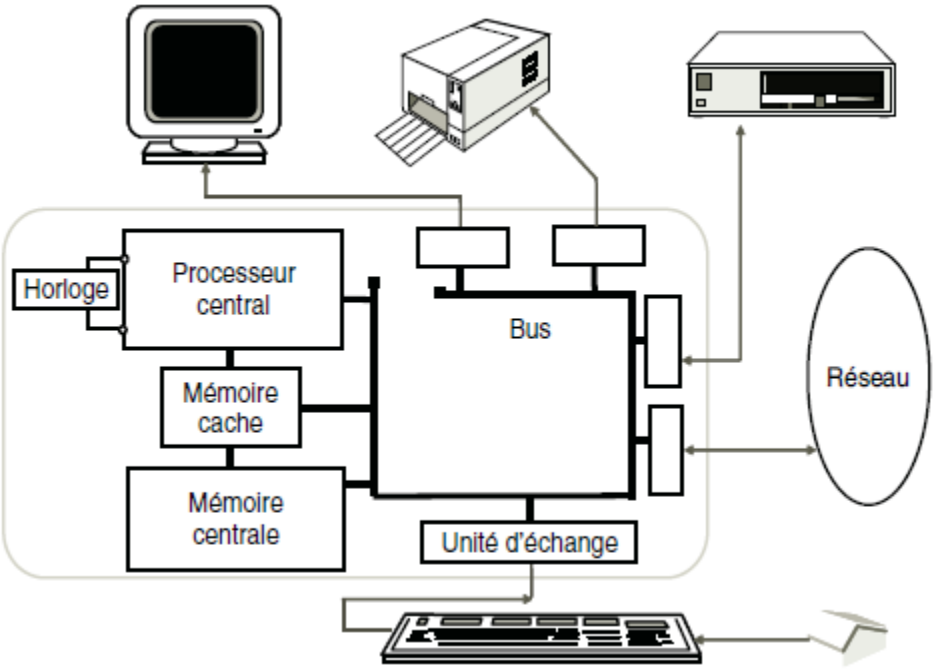


Figure 2 : *Structure matérielle générale.*

4.1. 2. Fonctionnement d'un Ordinateur

Tous les traitements réalisé par un ordinateur se fait via l'exécution d'un programme au niveau de CPU (Central Process Unit). Cette exécution se fait en suivant les étapes suivantes :

-Un programme, avant son exécution, sera tout d'abord chargé au niveau de la mémoire centrale ; (un programme est constitué de deux parties : *données* et *instructions*)

-Le microprocesseur récupère la première instruction du programme, réalise son décodage et l'exécute. Cette exécution, peut éventuellement récupérer des données de la mémoire central où écrire des données sur cette mémoire.

-Le microprocesseur réalise la même chose pour la deuxième instruction, et ainsi de suite, jusqu'à la dernière instruction du programme.

-Une fois le programme est terminé (l'exécution de la dernière instruction), l'espace de la mémoire central occupé par ce programme sera libérée.

4.2. Logiciels (Software)

L'aspect logiciel se rapport aux programmes qui sont installés dans l'ordinateur. Ce sont les logiciels qui font de l'ordinateur une machine différente des autres, c'est-à-dire une machine intelligente.

4.2.1. Définition

Un logiciel est un ensemble de programmes qui vont être exécutés par la machine pour réaliser une tâche.

4.2.2. Différents types de logiciels

On distingue trois types de logiciels : les applications, les langages de programmation et les systèmes d'exploitation.

- Les applications : C'est un type de logiciels qui peuvent être utilisés après le démarrage du système d'exploitation. Il existe plusieurs types, nous citons :

- Traitement de texte : Microsoft Office Word,

-Tableur : Microsoft Office Excel,

- Outils de présentation : Microsoft Office PowerPoint,

- Les bases de données : Microsoft Office Access.

- Les langages de programmation : Un programme informatique est une liste d'ordres indiquant à un ordinateur ce qu'il doit faire. On appelle « langage informatique » un langage destiné à décrire l'ensemble des actions consécutives qu'un ordinateur doit exécuter.

- Les systèmes d'exploitation : Ce sont des programmes qui permettent d'exploiter les ressources de la machine et de gérer la communication entre les différents périphériques. Le logiciel de base le plus important est le système d'exploitation. Parmi les systèmes d'exploitation, on peut citer : MS-DOS, Windows, Unix, Linux.

Introduction aux systèmes d'exploitation type Unix/Linux

Chapitre 2

1. Systèmes d'exploitation

Le système d'exploitation (noté SE ou OS, abréviation du terme anglais Operating System), est un ensemble de programmes responsables de la liaison entre les ressources matérielles d'un ordinateur et les applications informatiques de l'utilisateur (traitement de textes, vidéo,...). Il fournit aux programmes applicatifs des points d'entrées génériques pour les périphériques. Deux tâches :

-Fournir à l'utilisateur une machine étendue ou virtuelle, plus simple à programmer.

-Gestion des ressources.

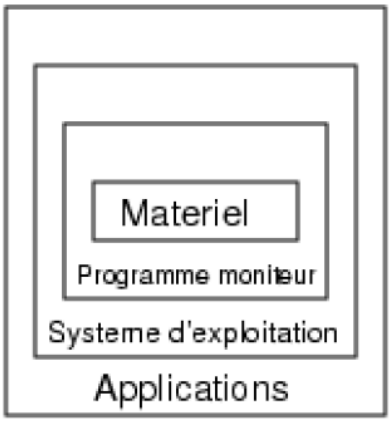


Figure 1: *Structure en couche d'un SE.*

Un système d'exploitation assure la liaison entre les utilisateurs, applications et les ressources matérielles de l'ordinateur ;

- il permet la gestion, la sauvegarde et l'organisation des informations au moyen d'une interface utilisateur compréhensible par l'homme (textes, icônes, images, graphes, *etc.*). Les informations manipulables sont enregistré sous forme de fichiers (contenu) que l'on range habituellement dans des dossiers (contenant) ;

– le système d'exploitation est sauvegardé sue le disque dur de l'ordinateur ;

– comme tous les programmes, le système d'exploitation s'exécute dans la mémoire vive.

2. Rôle d'un système d'exploitation

Le système d'exploitation contrôle et coordonne l'utilisation du matériel: Il met à la disposition des utilisateurs, les ressources matérielles de l'ordinateur :

-Gestion des ressources matérielles: le système gère de manière équitable et efficace les ressources matérielles (mémoire, processeur, périphériques, …).

Gestion du processeur: le système d'exploitation gère l'allocation du processeur entre les différents programmes. Pour l'utilisateur, les différents programmes fonctionnent parallèlement.

-Gestion de la mémoire: le système d'exploitation gère l'espace mémoire alloué à chaque application et à chaque utilisateur. Il la partage entre tous les programmes. En cas d'insuffisance de mémoire physique, le système d'exploitation peut créer une zone mémoire sur le disque dur, appelée «mémoire virtuelle», qui permet d'exécuter des applications nécessitant plus de mémoire qu'il n'y a de mémoire vive disponible sur le système.

-Sécurité / Accès aux données : Accès aux périphériques: écran, imprimante, disque dur, réseau.

Le système d'exploitation s'assure que les programmes puissent les utiliser de façon standard.

3. Unix/ Linux

3.1. Bref historique d'Unix

UNIX est créé au Laboratoire BELL (AT&T), USA, en 1969. Il est conçu par Ken Thompson et Dennis Ritchie, et inspiré du système Multics (MULTiplexed Information and Computing Service ou service multiplexé d'information et de calcul) créé en 1965 au le MIT (Massachusetts Institute of Technology). Il constitue le premier système d'exploitation multitâche et multiutilisateur.

• Initialement nommé Unics (Uniplexed Information and Computing Service)

• En 1973, le système est réécrit en langage C (langage développé par Dennis Ritchie) ce qui l'a rendu simple à porter sur de nouvelles plateforme ce qui lui a donné un véritable succès.

• Depuis la fin des années 70, deux grandes familles d'UNIX.

- Une version développée essentiellement par l'université de Berkeley (Californie), et nommée UNIX BSD (Berkeley Software Distribution).

- Une version nommée UNIX Système V commercialisé par AT&T.

• Projet GNU (1983) : objectif de développer un SE libre.

• Linux (1991) : un noyau UNIX libre développé par Linus Torvald (étudiant à l'université d'Helsinki)

Premier OS complet GNU/Linux libre. Linux est à la base d'une réécriture de Minix (1987: Andrew Tanenbaum, professeur à l'université libre d'Amsterdam a créé le système d'exploitation Minix). La

version 1.0 en 1994, qui donne naissance à la distribution d'un système d'exploitation entièrement libre, GNU/Linux.

Pourquoi utiliser Linux en chimie ?

- Manipuler des fichiers nombreux et volumineux.
- Réaliser des calculs nécessitant des serveurs puissants accessibles à distance (nombreux logiciels développés spécifiquement pour les systèmes UNIX/Linux).

3.2. Système d'exploitation UNIX

UNIX est un système d'exploitation multi-tâches et multi-utilisateurs. Il est:

Ouvert, c'est-à-dire il n'y a pas de code propriétaire (seules certaines implémentations sont propriétaires).

Portable, c'est-à-dire le code est indépendant de l'architecture (très peu de codes qui dépendent de l'architecture matériel de l'ordinateur).

Disponible sur différentes plateformes. La grande majorité des serveurs sur Internet fonctionnent sous UNIX.

Aujourd'hui, UNIX est très utilisé en informatique scientifique, et pour les serveurs réseaux.

3.3. Caractéristiques du système UNIX

Unix est un système d'exploitation multi-tâches (multithreaded en anglais): plusieurs processus (process en anglais), également appelées « tâches », peuvent être exécutées simultanément.

A chaque instant, le processeur ne traite qu'un seul processus (programme lancé), la gestion des processus est effectuée par le système.

Unix est un système d'exploitation multi-utilisateurs (multi-user): plusieurs utilisateurs peuvent utiliser le système en même temps (les ressources sont réparties entre les différents utilisateurs). Chaque utilisateur dispose de l'ensemble des ressources du système.

Le système Unix se charge de contrôler et de gérer l'utilisation et l'attribution des ressources entre les différents utilisateurs.

Unix présente une interface utilisateur interactive et simple à utiliser: le shell. Cette interface fournit des services de haut niveau. Elle intègre un langage de commandes très puissant (scripts shell).

Sous Unix, du point de vue utilisateur, il n'y a pas de notion de disque physique (partition, disque externe, …) contrairement à MS-DOS, en effet sous Unix, tout est fichier. L'utilisateur ne voit qu'une seule arborescence de fichiers hiérarchiques.

Les périphériques sont aussi représentés par des fichiers, ce qui rend le système indépendant du matériel et par conséquent assure la portabilité; l'accès aux périphériques est donc identique à l'accès aux fichiers ordinaires.

La gestion de la mémoire virtuelle : un mécanisme d'échange entre la RAM et le disque dur permet de pallier au manque de RAM.

Processus réentrants : les processus exécutant le même programme utilisent une seule copie de celui-ci en RAM.

Exemple: deux utilisateurs qui utilisent l'éditeur « vi », dans ce cas une seule copie de « vi » qui sera chargée en RAM.

3. 4. Architecture du système Unix

Un système informatique sous Unix/Linux est conçu autour d'une architecture en couche:

-La couche physique (hardware): c'est la couche la plus interne: ressources matérielles

-Au centre le noyau (en anglais kernel): le noyau UNIX est chargé en mémoire lors du démarrage de l'ordinateur. Il gère les tâches de base du système à savoir: la gestion de la mémoire, des processus, des fichiers, des entrées-sorties principales, et des fonctionnalités de communication.

-Fonctions systèmes : bibliothèque standard d'appels système. L'interpréteur de commandes. Son rôle est d'analyser la commande et envoie des appels au noyau en fonction des requêtes des utilisateurs. C'est l'interface utilisateur-Système. C'est le premier langage de commandes développé sur Unix par Steve Bourne.

-Utilitaires : éditeurs, compilateurs, gestionnaire de fenêtres et de bureau, etc.

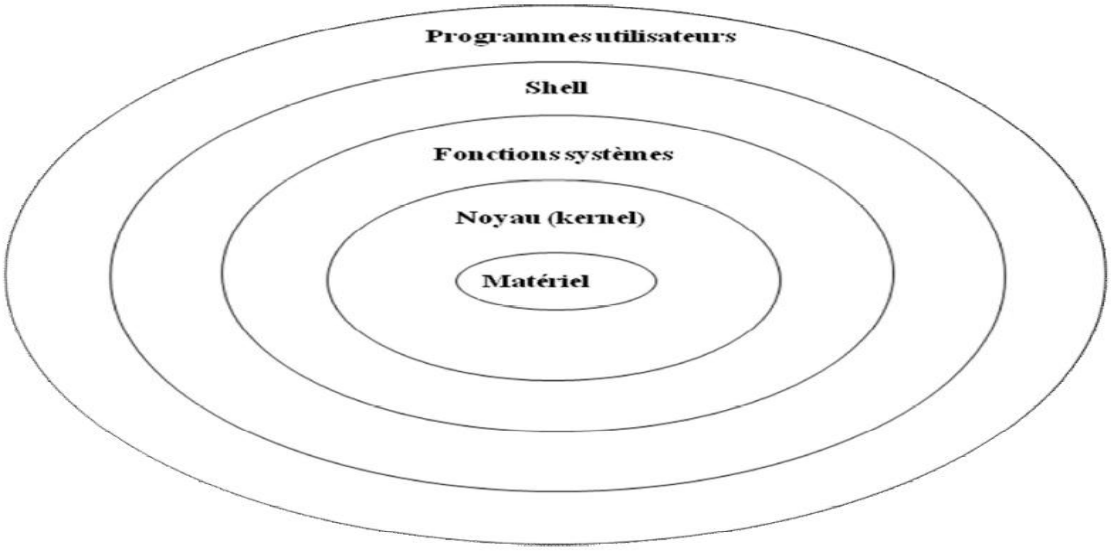


Figure 2: Architecture du système Unix

3.5. Connexion et déconnexion

Puisque Unix est un système multi-utilisateurs, alors il comporte les mécanismes d'identification et de protection permettant d'éviter toute interférence entre les différents utilisateurs. On distingue deux types d'utilisateurs: les administrateurs systèmes et les utilisateurs normaux:

-L'administrateur système appelé aussi « root », utilisateur privilégié ou super utilisateur (super user). Il dispose de tous les droits sur la machine et le système Unix. Il s'occupe de l'administration du système, en particulier il crée les comptes des utilisateurs.

-L'utilisateur normal dispose des droits réduits qui sont définis par l'administrateur système.

Unix associe à chaque utilisateur (un compte):

-Un nom d'utilisateur ou nom de connexion (appelé « login »).

-Un mot de passe (password en anglais),

-Un Home Directory (répertoire de l'utilisateur ou répertoire de connexion),

-Un langage de commandes (shell).

Donc à chaque connexion, le système demande aux utilisateurs leur login et leur mot de passe pour pouvoir travailler sur la machine. Si les deux sont valides alors Unix initialise l'environnement et ouvre une session de travail.

- Connexion

Pour se connecter à la machine et ouvrir une session de travail (pour pouvoir travailler sur la machine) il faut s'identifier. Pour cela, il faut:

- Entrer le nom de connexion après le message «login»

Login : <on tape ici le nom d'utilisateur>

- Entrer mot de passage après le message «password»

Password : <on tape ici le mot de passe>

Une fois connecté, l'utilisateur de trouve alors dans son propre répertoire de connexion (home directory) correspondant à son login (home directory).

Remarque: Pour des raisons de sécurité, les caractères du mot de passe sont cachés, et la vérification se fait après avoir tapé le login et le mot de passe. Si le login ou le mot de passe et incorrecte, un message d'erreur est alors affiché: « Invalid login name »

Attention: Unix fait la différence entre les minuscules des MAJUSCULES.

- Déconnexion

Pour terminer la session de travail, la méthode de déconnexion dépend de l'environnement de travail:

- Dans le cas d'un terminal, la commande de déconnexion est: « exit » ou « ctrl-D (^D) »

- Dans le cas d'environnement graphique, la méthode de déconnexion dépend l'interface graphique.

4. Système de fichiers (file system)

Notion de fichier et de répertoire

Le fichier est la plus petite entité logique de stockage permanent sur un disque ou d'autres supports physiques. Il peut contenir du texte, des données, programmes images. ou des programmes stockés sur un disque. Les fichiers sont classés dans des répertoires (catalogues). Chaque répertoire peut contenir d'autres sous-répertoires, formant ainsi une organisation arborescente.

Nomenclature des fichiers

Le nom d'un fichier sous Unix est une suite de caractères, dont la taille peut aller jusqu'à 255 caractères. La plupart des caractères sont acceptés, y compris l'espace (très déconseillé).

Cependant quelques caractères sont à éviter * & ; ( ) ~ <espace> \ | ` ? - (en début de nom)

Les différents types de fichiers

Pour l'utilisateur sous Unix, il n'existe pas la notion de disques physiques (tout est fichier).

L'utilisateur ne voit qu'une seule arborescence formée de répertoire et de fichiers. On distingue:

-Les fichiers ordinaires : Ce sont soit des fichiers contenant du texte, soit des exécutables (ou binaires), soit des fichiers de données. Le système n'impose aucun format particulier aux fichiers et il les traite comme des séquences d'octets. Contrairement au système MSDOS, on ne peut pas connaitre, à priori, les types des fichiers. Pour connaitre les types des fichiers on utilise par exemple la commande: « file ».

-Les répertoires (les fichiers répertoires): c'est un ensemble de fichiers ou d'autres répertoires (sous-répertoires) de manière récursive. Ils permettent une organisation hiérarchique.

-Les fichiers spéciaux : Ce sont des fichiers qui servent d'interface pour les divers périphériques (terminaux, disques dur, clavier, …). Les opérations de lecture/écriture sur ces fichiers sont directement dirigées vers le périphérique associé. Les fichiers correspondant aux périphériques sont stockés dans le répertoire « /dev » (devices).

Répertoires Unix

Un système de fichiers (File System en anglais), appelé aussi système de gestion de fichiers, est une structure de donnée qui définit l'organisation d'un disque (ou partition d'un disque). Il offre à l'utilisateur une vision homogène et structurée des données et des ressources : disques, mémoires, périphériques. Sous système Unix, tout est fichier, il n'y a pas de notions de disques, partition de disques, périphériques, ….

Les fichiers sont regroupés dans des répertoires et les répertoires contiennent soit des fichiers, soit d'autres répertoires.

Une telle organisation génère une hiérarchie de répertoires et de fichiers organises en arbre:

- La racine est désignée par « / » (slash): « / » désigne est le répertoire racine.
- Les noeuds sont les répertoires non vides

- Les feuilles sont les fichiers ou les répertoires "vides".

Sous Unix plusieurs systèmes de fichiers peuvent être rattachés au système de fichiers principal.

Chaque système de fichiers peut correspondre physiquement à :

- une partition ou à la totalité d'un disque physique.
- un périphériques (un DVD, un disque externe, …)

Par contre, sous Windows, les partitions, les périphériques, les disques externes, … sont vus comme des lecteurs indépendants (C:, D:, ...).

Sou Unix, on a:

- Un seul arbre général.
- Sa racine est désignée par « / ».
- Chaque répertoire peut contenir des fichiers ou des sous-répertoires.
- Un disque logique (partition, disque externe, clé USB, …) est vu comme un sous arbre qui se rattache à l'arbre principal. Le rattachement du sous arbre se fait automatiquement ou par l'utilisateur avec la commande « mount »

5. Les liens

Puisqu'un fichier est identifié par son numéro d'inode et non pas par son nom de fichier, il est possible de donner plusieurs noms à un même fichier grâce à la notion de lien: ceci permet d'accéder au même fichier à différents endroits de l'arborescence.

Avantage:

- Possibilité d'accéder au même fichier depuis des endroits et des noms différents

Une seule copie sur le disque et plusieurs façons d'y accéder.

- Si l'un des fichiers est modifié, la même modification est prise en compte par l'autre fichier.
- Simplifier l'accès à des fichiers dont les noms (chemin) sont difficiles à retenir.

Types de liens

On distingue deux types de liens : les liens symboliques« symbolic links » et les liens dur (ou lien physique) « hard links ».

Lien physique (hard link)

Un lien dur permet de donner plusieurs noms de fichiers qui ont le même inode (c'est dire des fichiers qui partagent le même contenu). Ceci peut se réaliser en rajoutant de nouveaux noms, dans la table des catalogues, associés au même inode associé au fichier source. Dans ce cas le même fichier physique (inode) est pointé par différents noms de fichiers. Le fichier source et le fichier lien pointent directement sur les données résidant sur le disque (l'information ne réside qu'une seule fois sur le disque mais elle peut être accédée par deux noms de fichiers différents).

Les droits du fichier source ne sont pas modifiés.

Lien symbolique

Un lien symbolique ne rajoute pas une entrée dans la table catalogue, mais c'est un fichier texte spécial, rajouté dans la table des inodes, qui contient un lien (sorte d'alias) vers un autre fichier ou répertoire. Donc son numéro d'inode est différent du fichier source mais qui pointe sur le fichier source pour permettre d'accéder aux informations sur le disque. Toute opération sur ce fichier (lecture, écriture, …) s'effectue sur le fichier référencé.

6. Les processus

Le système Unix/Linux est multi-tâche car plusieurs programmes peuvent être en cours d'exécution en même temps sur une même machine. Mais à chaque instant, le processeur ne traite qu'un un seul programmes lancés (un seul processus). La gestion des processus est effectuée par le système.

Les processus correspondent à l'exécution de tâches. Plusieurs définitions existent, on peut citer :

- Un processus est une tâche en train de s'exécuter. Il est doté d'un espace d'adressage (ensemble d'adresses) dans lesquelles le processus peut lire et écrire

- Un processus est l'activité résultant de l'exécution d'un programme par un processeur. C'est l'image de l'état du processeur et de la mémoire au cours de l'exécution du programme.

C'est donc l'état de la machine à un instant « t ».

Quelques données concernant les processus

• le PID (Process IDentifier): chaque processus Unix est identifié par une valeur numérique unique (PID). Le PID du premier processus lancé par le système est 1, c'est le processus « init » père de tous les processus.

7. Les entrées/sorties

Lors de l'exécution d'une commande, un processus est créé. Celui-ci va alors ouvrir trois canaux de communication:

- L'entrée standard,
- La sortie standard,
- La sortie d'erreur standard.

A chacun des trois canaux est affecté un nom de fichier et un numéro :

• Le fichier « stdin »: le processus lit les données en entrée à partir du fichier « stdin ». Il est ouvert avec le numéro logique 0. Par défaut, il est associé associé au clavier.

• Le fichier « stdout »: le processus écrit les sorties qu'il produit dans le fichier « stdout ». Il est ouvert avec le numéro logique 1. Par défaut, il est associé à l'écran.

• Le fichier « stderr » : le processus écrit les messages d'erreur dans le fichier « stderr ». Il est ouvert avec le numéro logique 2. Par défaut, il est associé à l'écran.

8. Les commandes de base du Système Unix

8.1. Introduction à l'interpréteur de commandes: le shell

Le shell est un programme (un fichier exécutable) qui a la charge d'analyser et d'exécuter les commandes:

- Il lit et interprète les commandes.
- Il transmet ces commandes au système.
- Il retourne le résultat.

Le shell sert d'interface entre le noyau (le système d'exploitation) et l'utilisateur. Toutes les commandes sont envoyées au noyau à travers le shell;

- soit en ligne de commande,
- soit via une interface graphique.

Principe du fonctionnement du shell en mode ligne de commande

En mode ligne de commande le shell affiche dans un terminal ou dans une console virtuelle, une chaine de caractères appelée « prompt » (ou invite de commande) et attend la saisi d'une commande. Quand on tape une commande suivi de la touche « Entrée, Return: la touche ↵ », le shell exécute cette commande et ensuite réaffiche le prompt et reste en attente sur une nouvelle commande. En générale la convention pour le prompt: $ ou % pour l'utilisateur normal # pour root (super-utilisateur ou administrateur) dans tous les shells

On peut personnaliser le prompt, par exemple on peut avoir un prompt comme ceci: smi-s3 >

Remarque :

- Pour lancer un terminal depuis l'interface graphique: on utilise le menu du bureau.
- Pour se connecter à une console virtuelle (un écran noir avec une invite de commande), depuis l'interface graphique: on utilise la combinaison des touches « Ctrl+Alt+FN », où N

est un chiffre de 1 à 6 (il y a 6 consoles virtuelles désignées par « tty1 » , … « tty6 ».

- Pour revenir au mode graphique depuis une console virtuelle, on utilise la combinaison des touches «Ctrl+ALT+F7 ».

N.B. Le mode graphique est désigné par « tty7 ».

Les principaux shells

Il existe plusieurs shells sous Unix :

• Le bourne shell; sh (/bin/sh): ancêtre de tous les shells. Il est disponible sur toute plateforme UNIX.

• Le Bourne again shell; bash (/bin/bash): version améliorée de « sh ». Disponible dans le domaine public (Fourni le plus souvent avec Linux).

• Le Korn shell; ksh (/bin/ksh): Bourne Shell étendu par l'AT&T.

• Le C shell (C veut dire California); csh (/bin/csh): Il est développé par BSD. Il offre plus de facilité (rappel des commandes avec les flèches, gérer l'historique des commandes, etc).

• Le Tenex C shell (TC Shell ou le C shell amélioré); tcsh (/bin/tcsh): c'est une extension du C Shell (csh).

Le Bourne Shell, le Korn Shell et le Bash Shell sont compatibles entre eux. Le C Shell et le TC Shell sont compatibles entre eux. Par contre, ces deux familles ne sont pas compatibles entre elles. Il est toutefois possible d'exécuter des procédures Bourne Shell alors que le shell de login est le C Shell

- Pour connaitre tous les shells que l'utilisateur peut utiliser, on consulte le fichier « /etc/shells » en exécutant par exemple la commande. % cat /etc/shells

- Pour connaitre le shell utiliser, par exemple, on tape la commande % echo $SHELL

8.2. La documentation Unix

Pour connaitre les différentes options sur les commandes, on peut utiliser la commande «man», qui permet d'accéder au manuel en ligne où toutes les commandes sont documentées.

% man commande.

L'affiche des pages du manuel se fait par page. Pour se déplacer dans le manuel, on utilise

- La touche « ENTRÉE ou RETURN » pour avancer d'une ligne ;

- La touche « ESPACE » pour avancer d'une page ;

- Le caractère « b » pour reculer d'une page ;

- Pour recherche un mot, on utilise le caractère « / » suivi du mot à rechercher ce mot. En suite pour rechercher l'occurrence suivante du même mot, on tape tapez le caractère « n » pour «next ».

- Pour quitter on tape le caractère « q ».

On peut aussi chercher la signification d'un élément, en utilisant la commande « whatis »

Exemple :

% whatis kill

On peut aussi utiliser la commande « help », quand c'est possible, qui donne plus de détail sur la commande, mais moins d'informations que la commande « man ».

8.3. Les commandes liées à l'environnement

8.3.1. La commande « mkdir » (make directory)

La commande « mkdir » permet de créer des répertoires.

Syntaxe :

mkdir [-p] [repertoire]

- Les arguments de « mkdir » sont les noms des répertoires à créer.

- Si le chemin n'est pas spécifié, le répertoire est créé dans le répertoire courant.

- Si le chemin est spécifié, la commande crée le répertoire dont le nom et le chemin sont spécifiés en argument de la commande. Le chemin peut être relatif ou absolu.

Exemple:

% mkdir projet crée le répertoire « projet » dans le répertoire courant.

% mkdir ../exam crée le répertoire « examen » dans le répertoire père du répertoire courant (chemin relatif).

% mkdir /home/etudiant/SM/tp crée le répertoire « tp» dans le répertoire « SM » qui est un sous-répertoire du répertoire de connexion « etudiant » (chemin absolu).

Remarque: Lorsqu'un répertoire est créé, il possède automatiquement deux sous répertoires à savoir « . » et « .. » .

8.3.2. La commande « rmdir » (remove directory)

La commande « rmdir » permet de supprimer des répertoires.

Syntaxe:

rmdir [repertoire]

- Les arguments de « rmdir » sont les noms des répertoires existants.

- Si le chemin n'est pas spécifié, le répertoire à supprimer est situé dans le répertoire courant.

- Si le chemin est spécifié, la commande supprime le répertoire dont le nom et le chemin sont spécifiés en argument de la commande. Le chemin peut être relatif ou absolu.

Exemple:

% rmdir tp

Supprime le répertoire « tp » situé dans le répertoire courant.

% rmdir ../exam

Supprime le répertoire « exam» situé dans le répertoire père du répertoire courant.

% rmdir /home/etudiant/SM/tp

Supprime le répertoire « tp» situé dans le répertoire « SM » qui est un sous-répertoire du répertoire de connexion «etudiant».

Attention: Les répertoires à supprimer doivent être vides (ils ne contiennent que « . » et « .. »).

En cas de plusieurs répertoires à supprimer, la commande « rmdir » supprime les répertoires dans l'ordre dans lesquels ils ont été précisés sur la ligne de commande. Par conséquent, on doit faire attention à l'ordre des arguments.

Exemple:

% rmdir SM SM/info SM/info/cours

Affiche un message d'erreur: « echec de suppression « SM »: le dossier est non vide ».

Il faut respecter l'ordre en commençant par le répertoire le plus interne pour être sûr que le répertoire est vide: % rmdir SM/info/cours SM/info SM

8.3.3. Les commandes de copies et de déplacements de fichiers

8.3.3.1. La commande cp (copy)

La commande « cp » permet de copier un ou plusieurs fichiers vers un autre fichier ou vers un répertoire.

Syntaxe :

cp [option] fich-source fich_destination

Options de la commande cp

• L'option « -i » : l'utilisateur doit confirmer avant d'écraser un fichier existant.

• L'option « -p » : préservation des permissions, dates d'accès de modification.

• L'option « -r » : récursif. Si le fichier source est un répertoire, alors on copie les fichiers et les sous-répertoires.

Exemple:

- Le fichier destination est un répertoire: copie du fichier «image.jpg» dans le répertoire «SM »

% cp image.jpg SM ou % cp image.jpg SM/

- Un seul fichier source et le fichier destination n'est pas un répertoire: copie du fichier «cv.txt » dans le fichier «cv_back.txt»

% cp cv.txt cv_back.txt

Plusieurs fichiers sources séparés par « espace », la destination est un répertoire: les fichiers sont copiés dans le répertoire «SM».

% cp cv.txt cv_back.txt SM/

8.3.3.2. La commande mv (move)

Elle permet de renommer et/ou déplacer un fichier. Sa syntaxe est similaire à la commande « cp ».

Syntaxe :

mv [option] fich-source fich_destination

Exemples:

- Un seul fichier source.

% mv cv.txt cv_old.txt

Renomme le fichier source « cv.txt » en lui donnant le nouveau nom « cv_old.txt ».

- Plusieurs fichiers source, la destination doit être un répertoire.

Déplacement des fichiers « image.jpg », « cv_old.txt » et du répertoire «cours» dans le répertoire « documents »

N.B. L'option « -i » : l'utilisateur doit confirmer avant d'écraser un fichier existant.

8.3.4. La commande « ls »

La commande « ls » liste le contenu d'un répertoire.

Syntaxe :

ls [options] [arguments]

- La commande « ls » sans arguments, liste le contenu du répertoire courant sauf les fichiers cachés (les fichiers commençant par un point).

- Si l'argument de la commande « ls » est un nom de fichier alors elle liste ce fichier s'il existe. Sinon elle affiche une erreur.

Exemple:

%ls image.eps

image.eps

Les options les plus utilisées de la commande « ls ».

- L'option « -l »: liste les fichiers avec des informations supplémentaires sur chaque fichier (le type de fichier, les protections, le nombre de liens , le propriétaire, le groupe, la taille en octets, la date de la dernière modification).

- L'option « -a »: affiche tous les fichiers y compris les fichiers cachés (fichiers dont le nom commence par « . » .

- L'option « -i » : affiche les numéros des inodes des fichiers.

- L'option « -s » : affiche la taille en Ko de chaque fichier.

- L'option « -F »: ajoute un « / » après le nom de chaque répertoire, un «*» après chaque fichier possédant le droit d'exécution et un « @ » après chaque fichier lien.

- L'option « -R »: liste les fichiers et les répertoires de façon récursive

- L'option « -d »: si le paramètre est un nom de répertoire, il vérifie s'il existe et ne descend pas dans un répertoire.

8.3.5. Effacement d'un fichier ou un répertoire - Commande rm

La commande « rm »(remove) permet de supprimer des fichier ou des répertoires, à condition que les permissions le permettent.

Syntaxe :

rm [options] [argument]

- Si l'argument est un nom de fichier ordinaire alors, il sera supprimé.
- Si l'argument est un répertoire alors il faut rajouter l'option « -r ».
- Si plusieurs arguments sont spécifiés, alors ils seront tous supprimés.

Les options:

« -r » (récursif): efface le répertoire ainsi que son contenu (fichiers et sous-répertoires).

« -i » (interactive): demande une confirmation (y ou n) sur chaque fichier à effacer.

« -f » (force) : supprime le fichier, même s'il est protégé (sans tenir compte des protections du fichier), il ne vérifie que le propriétaire. Vous pouvez donc effacer vos fichiers, même s'ils sont protégés.

Attention : il faut utiliser les options « -r » et « -f » avec précaution.

8.3.6. La commande ln

La commande « ln » permet de créer des lien sur un fichier (fichier existant). Si l'un des fichiers, source ou lien, est modifié, les autres le sont aussi.

Syntaxe:

ln [options] fichier_source fichier_lien

- fichier_source: c'est le fichier (ou répertoire) sur lequel on crée un lien.
- fichier_lien: c'est le nom du fichier lien. On peut accéder au fichier source « fichier_source » via le fichier lien « fichier_lien ».

Remarque : Si la destination est un répertoire, alors on crée un lien dans ce répertoire qui a le même nom que le fichier source.

8.3.7. La commande cd

La commande cd permet le positionnement sur un répertoire

Syntaxe :

cd [répertoire]

Pour cette commande, il y a des options fréquentes :

cd / positionnement à la racine

cd .. positionnement sur le répertoire parent

8.3.8. Commandes de manipulation de fichiers

Les trois commandes cat, more et less permet de lire un fichier text. La différence entre ces trois commandes est la méthode pour lire le fichier, soit page par page, soit ligne par ligne ou même avec le pourcentage % du fichier entier qui apparait à la fin du terminal.

cat

Syntaxe :

cat fichier fichier1 [fichier2]

more ou less

Avec ces deux commandes, on peut affichent le contenu d'un fichier page par page, ou bien ligne par ligne utilisant la touche entrée, page par page avec "Espace", on quitte l'affichage avec "q". Des fois le pourcentage % apparaît sur la dernière ligne du terminal, ce dernier indique la position actuelle par rapport à la totalité du fichier.

Chapitre 3

Traitement statistique et graphique de données.

1. Statistique

La statistique est l'étude de la collecte de données, leur analyse, leur traitement, l'interprétation des résultats et leur présentation afin de rendre les données compréhensibles par tous. C'est à la fois une science, une méthode et un ensemble de techniques. L'analyse des données est utilisée pour d'écrire les phénomènes étudiés, faire des prévisions et prendre des décisions à leur sujet. En cela, la statistique est un outil essentiel pour la compréhension et la gestion des phénomènes complexes.

Les données étudiées peuvent être de toute nature, ce qui rend la statistique utile dans tous les champs disciplinaires et explique pourquoi elle est enseignée dans toutes les filières universitaires.

La statistique consiste à :

– Recueillir des données.

– Présenter et résumer ces données.

– Tirer des conclusions sur la population étudiée et d'aider à la prise de décision.

1.1. Statistique Descriptive

Ensemble des méthodes et techniques permettant de présenter, de d'écrire, de résumer des données nombreuses et variées. Il faut d'abord préciser l'ensemble étudie, appelé population statistique, dont les éléments sont des individus, ou unités statistiques. Il est fréquent qu'on ne puisse observer toute la population statistique.

-**Statistique Descriptive univariée** : La Statistique Descriptive univariée étudie un seul caractère statistique, et ne s'intéresse donc pas aux liens éventuels entre plusieurs caractères.

- **Statistique Descriptive bivariée** : La Statistique Descriptive bivariée concerne l'extraction d'information sur deux caractères statistiques, et leurs liens éventuels.

- **Statistique Descriptive multivariée** : La Statistique Descriptive multivariée analyse un nombre k(> 2) de variables mesurées ou observées simultanément sur les mêmes individus. Elle permet de mettre en évidence le type de lien existant éventuellement entre ces variables.

1.2.. Statistique inférentielle : est l'ensemble des méthodes permettant, a partir d'un échantillon, d'estimer des paramètres d'une population statistique et/ou de tester des hypothèses sur cette population. A l'inverse de la statistique descriptive, la statistique inférentielle fait appel `a la théorie des probabilités `a travers les notions de précision statistique et de risque d'erreur décisionnel.

2. Généralités sur la statistique descriptive

2.1. Épreuve statistique

Les statistiques visent à étudier les caractéristiques d'un ensemble d'observations comme les mesures obtenues lors d'une expérience. Donc, l'épreuve statistique est une expérience que l'on provoque.

2.2. Population

En statistique, le terme de population s'applique à tout objet statistique étudié, qu'il s'agisse d'étudiants, de ménages ou de n'importe quel autre ensemble sur lequel on fait des observations statistiques.

Une population statistique est un ensemble d'éléments sur lesquels porte une étude.

Exemple :

– On considère l'ensemble des étudiants de la section A. On s'intéresse aux nombre de frères et sœurs de chaque étudiant. Dans ce cas

Population = ensemble des étudiants

2.3. Individu (unité statistique)

Une population est composée d'individus. Les individus qui composent une population statistique sont appelés unités statistiques.

2.4. Échantillon

L'échantillon est un sous ensemble fini de la population. La taille de l'′echantillon est le nombre d'éléments sélectionnes pour constituer l'échantillon.

Exemple :

– Dans l'exemple indiqué ci-dessus, un individu est tout étudiant de la section.

2.5. Modalités

Les modalités d'une variable statistique sont les différentes valeurs que peut prendre celle-ci.

Exemple :

Variable est " Rendement"

Modalités sont " Faible, Moyen, Elevé"

2.6. Caractère (variable statistique)

Variable statistique (ou caractère statistique) : propriété (ou valeur) distinctive observée ou mesurée sur l'individu statistique. Elle peut être qualitative ou quantitative.

Exemple :

Concentration, température, couleur …

2.6.1.Caractère qualitatif

Un caractère statistique est qualitatif si ses valeurs, ou modalités, s'expriment de façon littérale ou par un codage sur lequel les opérations arithmétiques telles que moyenne, somme,…, n'ont pas de sens. Il est représenté par autre chose que des chiffres.

Exemple :

Variable est " Rendement"

Modalités sont " Faible, Moyen, Elevé"

2.6.2. Caractère quantitatif

Un caractère statistique est quantitatif si ses valeurs sont des nombres sur lesquels des opérations arithmétiques telles que somme, moyenne,…, ont un sens. L'ensemble des valeurs est représenté par des chiffres.

Un caractère quantitatif est dit discret si les valeurs possibles sont des nombres isolés sur l'axe réel. Dans la pratique, il s'agit souvent de données de comptage.

Un caractère quantitatif est dit continu s'il peut prendre toutes les valeurs dans un intervalle réel.

Exemple :

– Degré d'acidité du vinaigre.

Modalités : 5° , 6°...

Type : Discret.

– La rigidité des ressorts.

Modalités : [10, 20] N/m

Type : continu.

2.7. Classe modale : C'est la classe ayant le plus grand effectif par unit´e d'amplitude. Dans le cas d'une classe modale unique, on parle de distribution continue unimodale.

Classe statistique : Intervalle correspondant à des valeurs observées pour un caractère quantitatif continu.

2.8. Classe de valeurs

On appelle classe de valeurs de X un intervalle de type [a, b[tel que X Є [a, b[si et seulement si a≤ X(w) < b, c'est à dire, que les valeurs du caractère sont dans la classe [a, b[.

2.9. Distribution statistique

Ensemble des modalités, valeurs, ou classes d'une variable, avec les effectifs observés correspondants. Une distribution d'effectifs univariée est la donnée de (x_1, n_1),… ,(x_k, n_k), où les x_i sont les valeurs distinctes du caractère statistique et n_i l'effectif associé x_i

2.10. Le mode Mo : C'est la valeur la plus fréquente dans la série. Le mode n'est pas forcément unique.
Si c'est le cas on parle de distribution unimodale. Sinon, on parle de distribution multimodale. Notons que le mode est calculable pour une variable qualitative.

2.11. La moyenne

C'est la somme des valeurs divisée par le nombre de valeurs. Pour une distribution d'effectifs (x_1, n_1),… ,(x_k, n_k), où x_i a pour effectif associé n_i, la moyenne notée $\overline{x}$ est donn´e par la formule :

$$\overline{x} = \frac{1}{n}(n_1 x_1 + \cdots + n_k x_k)$$

2.12. Ecart interquartile : C'est la différence I entre le 1er et le 3ème quartile : I = Q3 − Q1.

Ecart-type : pour une distribution d'effectifs (x_1, n_1),… ,(x_k, n_k), où x_i a pour effectif associé n_i, l'écart-type noté s_x est donné par la formule :

$$s_x = \sqrt{\left(\frac{1}{n}\right)(n_1(x_1 - \overline{x})^2 + \cdots + n_k(x_k - \overline{x})^2)}$$ où $\overline{x}$ est la moyenne de la série.

2.13. La médiane

On appelle médiane la valeur Me de la série statistique. X qui vérifie la relation suivante :

$$Fx(Me-) < 0.5\ Fx(Me+) = Fx(Me).$$

La médiane partage la série statistique en deux groupes de même effectif.

2.14. L'étendue

La différence entre la plus grande valeur et la plus petite valeur du caractère, donnée par la quantité

$$e = x_{max} - x_{min}$$

2.15. La Variance

La variance ϑ : La variance d'une série statistique vérifie :

$$Variance = \frac{Somme\ des\ carrés\ d'écart\ à\ la\ moyenne\ de\ la\ série}{Nombre\ de\ valeurs\ dans\ la\ série}$$

Pour une distribution d'effectifs (x_1, n_1),…, (x_k, n_k), où x_i a pour effectif associé n_i, la variance notée s_x^2 est donnée par la formule :

$$s_x^2 = \left(\frac{1}{n}\right)(n_1(x_1 - \overline{x})^2 + \cdots + n_k(x_k - \overline{x})^2)$$

La variance est le carré de l′ecart-type.

4.16. Fréquence (ou fréquence relative) :

C'est la proportion (ou le pourcentage) d'individus pour lesquels une variable statistique a pris une valeur donnée.

2.17. Fréquence cumulée :

Résultat de l'addition, de proche en proche, des fréquences d'une distribution observée, soit en commençant par le 1er :

$F_1 = f_1$, $f_2 = f_1 + f_2$,…, $f_i = f_1 + f_2 + \cdots + f_i$ (fréquences cumulées croissantes).

Soit en commençant par le dernier (en notant k le nombre total de valeurs distinctes) :

$F^*_k = f_k$, $F^*_{k-1} = f_k + f_{k-1}$, $\cdots$, $F^*_i = f_k + f_{k-1} + \cdots + f_i$ (fréquences cumulées décroissantes).

2.18. Quartiles :

Le premier quartile Q1 : c'est une valeur telle qu'il y a 25% de valeurs qui lui sont inférieures dans la série statistique et 75% qui lui sont supérieures.

Le troisième quartile Q3 : c'est une valeur telle qu'il y a 75% de valeurs qui lui sont inférieures dans la série statistique et 25% qui lui sont supérieures.

Notons que la médiane est le second quartile. Les 3 quartiles s'obtiennent en séparant la série réordonnée en quatre parties d′egale fréquence.

3. Représentation d'une série (Statistique à une dimension)

Les représentations recommandées et les plus fréquentes sont les tableaux et les diagrammes. Dans un document scientifique ou académique, il convient de les numéroter et de les légender. Cela facilite la lecture du document et permet de les référencer dans le texte.

Un tableau comprend 3 parties : le titre, le corps et la source d'information. Le titre permet de préciser le lieu, la période et les variables auxquels correspondent les données. La source d'information indique clairement s'il s'agit de données personnelles ou de données obtenues auprès d'un quelconque organisme ou média. Le corps du tableau dépend, lui, de la nature de la variable statistique étudiée.

31. Variable qualitative

A partir de l'observation d'une variable qualitative sur n individus statistiques, on peut construire un tableau dont le corps est :

***Table 1** : Corps de tableau pour une variable qualitative.*

Modalités	Effectifs	Fréquences
Modalités 1	n_1	f_1
Modalités 2	n_2	f_2
.	.	.
.	.	.
.	.	.
Modalités k	n_k	f_k
Totaux	N	1

où

$\mathbf{n_i}$ est l'effectif associé à la modalité i c'est-à-dire le nombre d'individus dans l'échantillon ayant cette modalité ;

n est la taille de l'échantillon (nombre total d'individus dans cet échantillon) ;

$\boldsymbol{f_i} = \mathbf{n_i/n}$ est la fréquence associée `a la modalité i c'est-à-dire la proportion d'individus dans l'échantillon ayant cette modalité ;

k est le nombre de modalités distinctes observées dans l'´echantillon.

Deux diagrammes permettent de représenter une variable qualitative : le diagramme à secteurs angulaires (dit camembert) et le diagramme en bandes (dit tuyaux d'orgue).

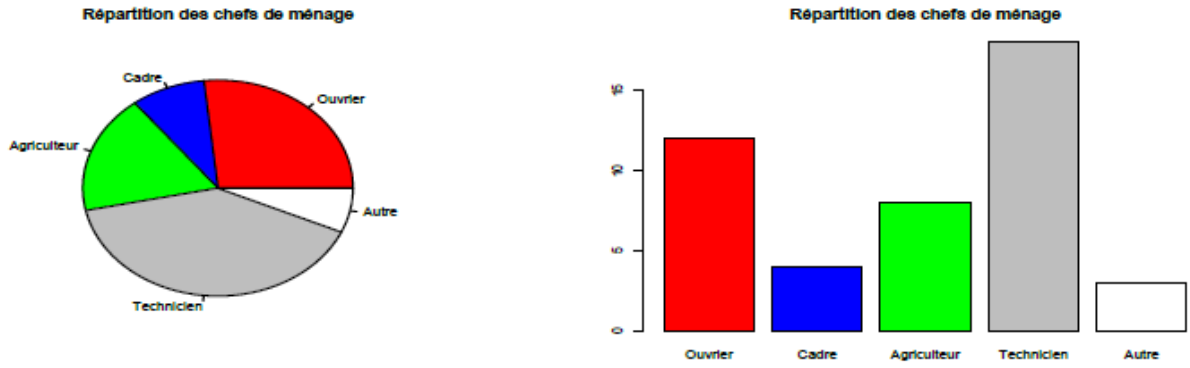


***Figure 1** : Représentations d'une variable qualitative.*

Le camembert est un disque partagé en secteurs, chaque secteur représentant une modalité et ayant une surface proportionnelle à la fréquence de cette modalité dans la série statistique.

Le diagramme en bandes (en colonnes) ou en bâtons est un ensemble de rectangles de même largeur, sépares par un espace, chaque rectangle représentant une modalité et ayant une hauteur proportionnelle à la fréquence de cette modalité dans la série statistique.

3.2. Variable quantitative discrète

Quand la variable est quantitative, on utilise les mêmes représentations à l'aide des fréquences absolues et relatives. La différence fondamentale entre les représentations pour des variables qualitatives et quantitatives tient au fait qu'il existe un ordre naturel sur les modalités (qui sont des nombres réels) pour les variables quantitatives. C'est pourquoi les diagrammes en bâtons sont toujours utilisés, mais pas les diagrammes sectoriels.

A partir de l'observation d'une variable quantitative discrète sur n individus statistiques, on peut construire un tableau dont le corps est donné par Table 2 :

***Table 2** : Corps de tableau pour une variable quantitative discrète*

Valeurs	Effectifs	Fréquences	Fréquences cumulées
x_1	n_1	f_1	F_1
x_2	n_2	f_2	F_2
.	.	.	.
.	.	.	.
.	.	.	.
x_k	n_k	f_k	F_k.
Totaux	n	1	-

Où

n_i est l'effectif associé à la valeur x_i c'est-à-dire le nombre d'individus ayant cette valeur dans l'échantillon ;

n est la taille de l'échantillon (nombre total d'individus dans cet échantillon) ;

$f_i = n_i/n$ est la fréquence associée à la valeur xi c'est-à-dire la proportion d'individus dans l'échantillon ayant cette valeur.

F_i est la fréquence cumulée en x_i c'est-à-dire la proportion d'individus dans l'échantillon ayant une valeur inférieure ou égalé à x_i. Le calcul des F_i peut se faire façon récurrente de la manière suivante :

$F_1 = f_1$ et $F_i = F_{i-1} + f_i$ pour $i \in \{2, \cdots k\}$

k est le nombre de valeurs distinctes observées dans l'échantillon.

Les valeurs distinctes sont par ordre croissant dans le tableau :

$$x_1 < x_2 < \ldots < x_k.$$

Deux diagrammes permettent de représenter une variable quantitative discrète : le diagramme en bâtons et le diagramme cumulatif.

Le diagramme en bâtons associe à chaque valeur de la variable un segment vertical de hauteur proportionnelle à la fréquence de cette valeur dans la série statistique.

Le diagramme cumulatif est une courbe en escalier représentant les fréquences cumulées relatives.

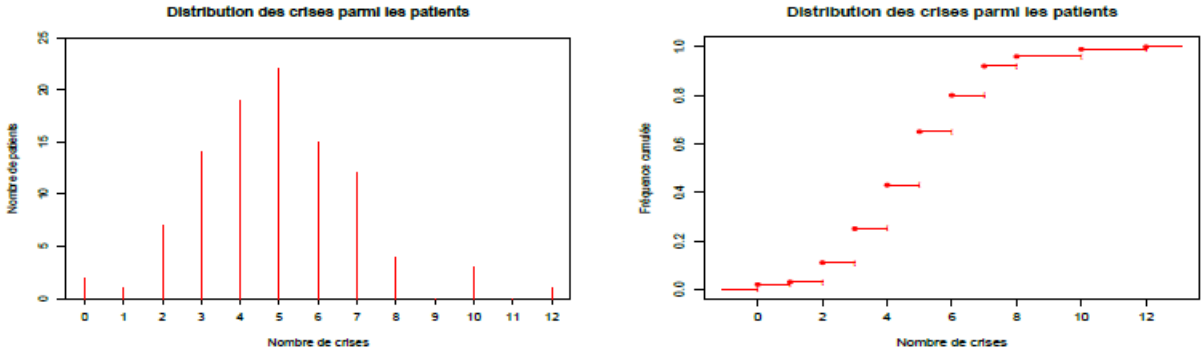


***Figure 2** : Représentations d'une variable quantitative discrète*

3.3. Variable quantitative continue

A partir de l'observation d'une variable quantitative continue sur n individus statistiques, on peut déterminer k classes statistiques et construire un tableau dont le corps est :

***Table 3**: Corps de tableau pour une variable quantitative continue.*

Classes statistiques	Effectifs	Fréquences	Fréquences cumulées
$]a_0, a_1]$	n_1	f_1	$F(a_1)$
$]a_1, a_2]$	n_2	f_2	$F(a_2)$
.	.	.	.
.	.	.	.
.	.	.	.
$]a_{k-1}, a_k]$	n_k	f_k	$F(a_k)$.
Totaux	n	1	-

Où

n_i est l'effectif associé à la classe $]a_{i-1}; a_i]$ c'est-à-dire le nombre d'individus ayant une valeur comprise entre a_{i-1} (exclus) et a_i dans l'échantillon ;

n est la taille de l'échantillon (nombre total d'individus dans cet échantillon) ;

$f_i = n_i/n$ est la fréquence associée à la classe $]a_{i-1}; a_i]$ c'est-à-dire la proportion d'individus ayant une valeur comprise entre a_{i-1} (exclus) et a_i dans l'échantillon ;

F(ai) est la fréquence cumulée en a_i c'est-a-dire la proportion d'individus dans l'échantillon ayant une valeur inferieure ou égale à a_i.

Le calcul des F(a_i) peut se faire façon récurrente de la manière suivante : F(a_1) = f_1 et F(a_i) = F(a_{i-1}) + f_i pour i $\in\{2,\cdots k\}$

k est le nombre de valeurs distinctes observées dans l'échantillon. Les bornes de classe vérifient bien évidemment : $a_0 < a_1 < a_2 < \ldots < a_k$.

Quand la variable étudiée est continue, les représentations du type diagramme en bâtons sont sans intérêt, car les données sont en général toutes distinctes, donc les fréquences absolues sont toutes égales à 1. On considérera ici deux types de représentations graphiques :

-l'histogramme et le polygone des fréquences qui lui est associé.

-la fonction de répartition empirique, qui permet notamment de construire des graphes de probabilités.

Ces deux types de représentations nécessitent d'ordonner les données. Si l'échantillon initial est noté $x_1,\ldots x_n$, l'échantillon ordonne sera noté $x_1^*,\ldots x_n^*$ (x_1^*= min($x_1,\ldots x_n$)).

-Histogramme et polygone des fréquences

L'histogramme est une juxtaposition de rectangles, chaque rectangle étant associé à une classe statistique et étant de surface (et non pas de hauteur) proportionnelle à la fréquence de cette classe.

L'histogramme est la figure constituée de rectangles dont les bases sont les classes et dont les aires sont égales aux fréquences de ces classes. Autrement dit, la hauteur du $i^{ème}$ rectangle est n_i/nh_i

L'histogramme n'est pas une approximation satisfaisante de la densité dans la mesure où c'est une fonction en escalier, alors que la densité est en général une fonction continue.

Une meilleure approximation est le polygone des fréquences, c'est à dire la ligne brisée reliant les milieux des sommets des rectangles, et prolongée de part et d'autre des bornes de l'histogramme de sorte que l'aire sous le polygone soit égale à 1 (comme une densité). Le polygone des fréquences est représenté dans la figure 3.

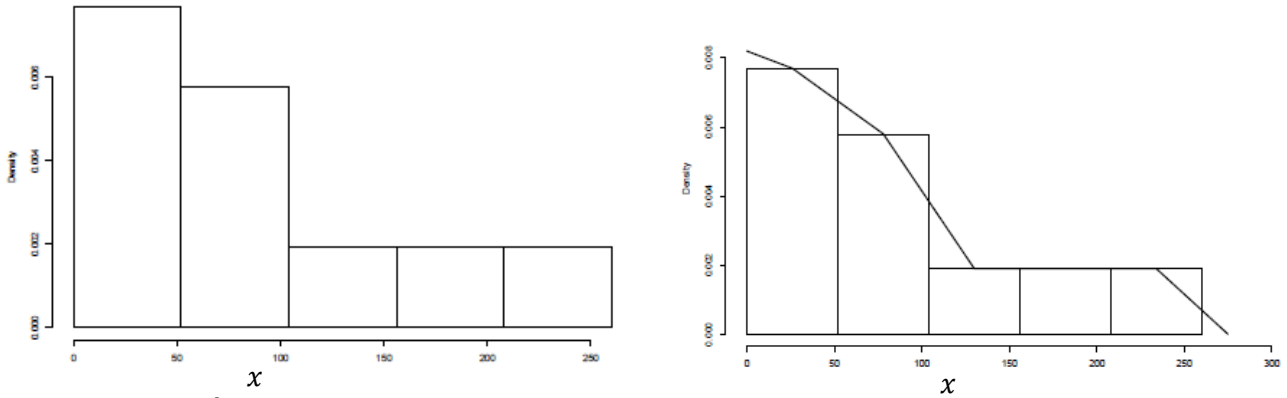


***Figure 3** : Représentations d'une variable quantitative continue. Droite ; Exemple d'histogramme Gauche ; Exemple d'histogramme à classes de même largeur et polygone des fréquences*

-Fonction de répartition empirique

La fonction de répartition empirique, qui permet notamment de construire des graphes de probabilités. La fonction de répartition empirique est la proportion d'éléments de l'échantillon qui sont inferieurs ou égaux à x. C'est une fonction de l'ensemble des valeurs prises par x dans [0, 1], la fonction de repartition empirique F_n associee à un échantillon $x_1,\ldots, x_n$ est la fonction définie par :

$$\forall x \in R\ ;\ F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{\{x_i \le x\}} = \begin{cases} 0 & si\ x < x_1^* \\ \frac{i}{n} & si\ x_i^* \le x < x_{i+1}^* \\ 1 & si\ x \ge x_n^* \end{cases}$$

$F(x)$ est la probabilité qu'une observation soit inferieure à x tandis que $F_n(x)$ est le pourcentage d'observations inferieures à x. On conçoit donc bien que $F_n(x)$soit une estimation de $F(x)$.

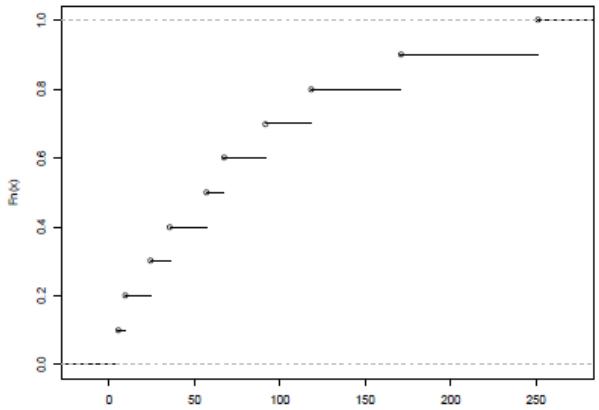


***Figure 3** : Représentations d'une variable quantitative continue. Exemple de fonction de répartition empirique pour une loi continue.*

4. Méthodes de statistique inférentielle

4.1. Inférence statistique : ensemble des méthodes permettant de formuler en termes probabilistes un jugement sur une population, à partir des résultats observés sur un échantillon extrait au hasard de cette population.

4.2. Les hypothèses de la statistique inférentielle:

-La population est considérée comme infinie (très grande).

-Les variables statistiques qui la décrivent peuvent être considérées comme des variables aléatoires.

4.3. Variable aléatoire

Une variable aléatoire est une grandeur dépendant du résultat d'une experience aléatoire, c'est-à-dire non prévisible a l'avance avec certitude.

4.3.1.Variables aléatoires discrètes

Une variable aléatoire X est dite discrète si et seulement si elle est à valeurs dans un ensemble E fini ou dénombrable. On peut noter E = {x_1, x_2, …}.

La loi de probabilité d'une **variables aléatoires discrètes** X est entièrement déterminée par les probabilités élémentaires $P(X = x_i)$, $\forall x_i \in E$.

La fonction de repartition de X est alors $F_X(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i)$.

4.3.2. Variables aléatoires continues

si et seulement si sa fonction de répartition F_X est continue et presque partout dérivable. Sa dérivée f_X est alors appelee densite de probabilite de X, ou plus simplement densite de X. Une **variable aléatoire continue** est forcément à valeurs dans un ensemble non dénombrable.

5. Quelques lois d'une variable aléatoire

51. La loi d'une variable aléatoire discrète

Une loi est à densité (de densité f) si les probabilités s'expriment comme des intégrales :

$$P(X \in [a, b]) = \int_a^b f(t)dt$$

Soit X une variable aléatoire à valeurs dans R et f (x) une densité de probabilité sur R. On dit que X est une variable aléatoire continue de densité f (x) si pour tout intervalle A de R on a :

$$P[X \in A] = \int_A f(x)dx$$

5.2. La loi d'une variable aléatoire continue est définie à partir d'une fonction f appelée densité qui vérifie les propriétés suivantes :

- f est positive ; pour tout $x \in R$, $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x)dx = 1$

La loi de la variable aléatoire X est la loi continue sur R, de densité f (x).

Pour déterminer la loi d'une variable aléatoire continue, il faut donc calculer sa densité.

Soit X une variable aléatoire continue, f sa densite

La probabilité que X appartienne à l'intervalle [a; b] : $P(a \leq X \leq b) = \int_a^b f(t)dt$

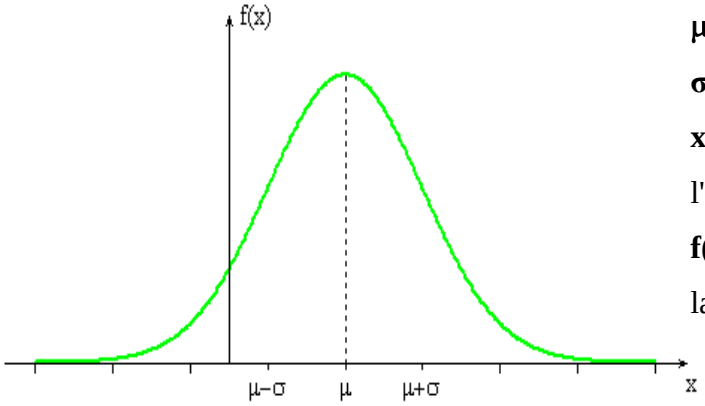
5.3. la loi normale ou gaussienne

La loi gaussienne (ou normale) est une des lois de probabilité les plus utilisées dans les sciences appliquées du fait de ses propriétés théoriques remarquables.

La loi gaussienne est une loi continue qui depend de deux paramètres $\mu \in R$ et $\sigma > 0$. Sa densiteé est :

$$p_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

Cette densité ne dépend que de deux paramètres μ et σ. L'allure de cette densité de probabilité est présentée sur la figure suivante.



μ est la moyenne

σ l'écart type

x le nombre total d'individus dans l'échantillon

f(x) le nombre d'individus pour lesquels la grandeur analysée a la valeur x.

***Figure 4**: (Densité de probabilité d'une variable aléatoire gaussienne)*

Pour chaque μ, σ, il existe une loi normale de moyenne μ et d''ecart-type σ. On la note N (μ, σ^2).

Exemple de représentation graphique de *f* -Loi normale centrée réduite :

Cas particulier μ = 0 et σ = 1 : loi normale centrée/réduite, suit la loi normale N (0, 1).Dont la densité de probabilité est donnée par la fonction: $P_{0,1}(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$

-Les autres lois normales

L'interprétation de σ comme l'écart-type de X explique son influence sur la forme de la représentation graphique de sa densité. Ci-dessous :

-en rouge, la densité de la loi normale N(0, 1/4)d'espérance 0 et d'écart-type 1/2

-en bleu, la densité de la loi normale N (0 , 1) d'espérance 0 et d'écart-type 1 ;

-en vert, la densité de la loi normale N (0 , 4) d'espérance 0 et d'écart-type 2.

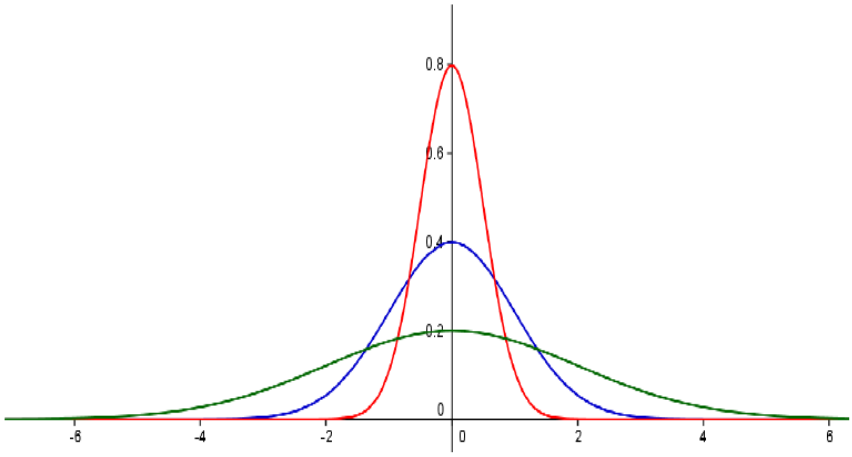


***Figure 5 :** La courbe représentative de la distribution d'une loi N(μ;σ^2) (N(0,1/4),N(0,1) et N(0,4)).*

5.4. Erreurs aléatoires ou « indéterminées »

En l'absence d'erreurs systématiques, on a affaire aux erreurs accidentelles « dues au hasard » qui ne peuvent être contrôlées car indéterminées. L'analyse mathématique de la courbe d'erreur conduit à la

conclusion que la moyenne arithmétique x des valeurs individuelles est la meilleure estimation de la moyenne vraie μ (voir figure 7). La symétrie de cette courbe et son aspect montrent que :

– il y a un nombre égal d'erreurs positives et négatives par rapport à la valeur centrale ;

– les petites erreurs sont plus nombreuses que les grandes erreurs ;

– la valeur la plus souvent rencontrée est la valeur centrale m (sans erreur).

La loi de distribution Normale (courbe de distribution de Gauss) est le modèle mathématique qui représente le mieux la répartition des erreurs dues au hasard:

$$f_{\mu,\sigma}(x)=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

On prend pour origine des x la moyenne vraie μ de la population et pour unité de mesure son écart-type σ (voir figure 7).

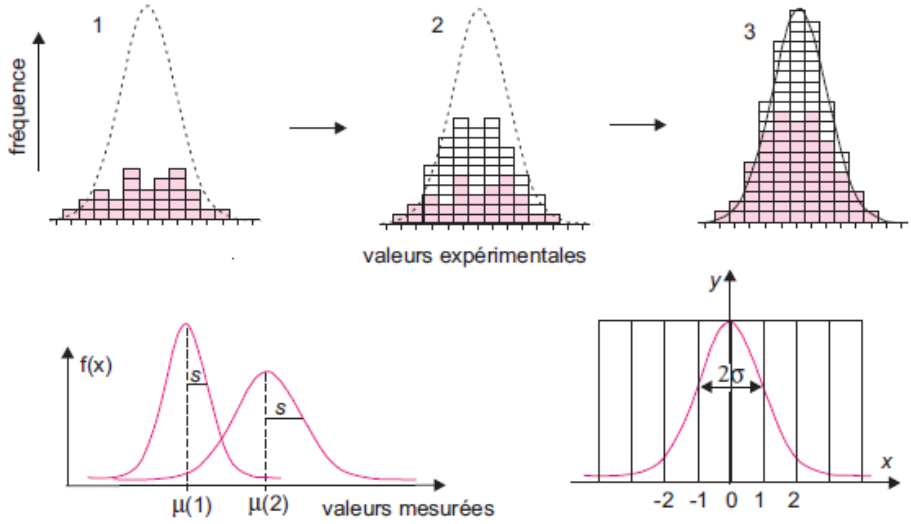


Figure 6: *Courbes de Gauss.*

Quand le nombre de mesures devient grand, et si l'intervalle de classe est étroit, le contour de cet empilement (mesure/fréquence) épouse la forme d'une courbe de Gauss (loi de distribution Normale).

La fonction de répartition est l'intégrale de la fonction f (X). Cette courbe est telle que 95,4% de l'aire est comprise dans un intervalle de ±2σ autour de la valeur centrale. On dit encore que les chances sont de 95,4% pour que l'erreur d'une mesure donnée soit comprise dans un intervalle de ±2σ. Une valeur élevée de l'écart-type σ signifie une courbe d'erreur évasée.

Chapitre 4

Etude de banques de données chimiques indexées par structure.

1. Présentation des bases de données

-Notion de fichier

La notion de fichier (File) a été introduite en informatique durant les années 50. Un fichier est un récipient d'information caractérisé par un nom, constituant une mémoire secondaire idéale, permettant d'écrire des programmes d'application indépendants des mémoires secondaires.

Il y avait deux types de fichiers :

Fichier des données : des séquences d'enregistrements dont l'accès est séquentiel ou indexé.

Fichier de traitement : un ensemble d'instructions permettant la manipulation des données des fichiers.

-L'intérêt d'un fichier

L'objectif des fichiers était de simplifier l'utilisation des mémoires secondaires des ordinateurs. Les fichiers fournissent des récipients de données plus manipulables aux programmes et sont gérées par un système de gestion de fichiers.

Suite à la sophistication des systèmes informatiques, les données stockées dans des fichiers sont devenues structurées. En effet, aujourd'hui, les fichiers sont à la base des systèmes d'information. De ce fait, le premier niveau d'un SGBD (Système de Gestion de Base de Données) est la gestion de fichiers. Le SGBD sera présenté dans les paragraphes suivants.

Les données gérées par l'entreprise et les programmes spécifiant les traitements sur les données sont stockés dans des fichiers gérés par le système informatique.

La gestion des fichiers permet de traiter et de stocker des quantités importantes de données, et de les partager entre plusieurs programmes. De plus, elle sert de base au niveau interne des Systèmes de Gestion de Bases de Données.

2. Définition de base de données (BD)

Une BD est un ensemble de données modélisant les objets d'une partie du monde réel et servant de support à une application informatique. Pour mériter le terme de BD, un ensemble de données

-doit être interrogeable par le contenu, c'est-à-dire que l'on doit pouvoir retrouver tous les objets qui satisfont à un certain critère.

Par exemple tous les éléments chimiques qui ont une masse atomique moins de 15,

- les données doivent être interrogeables selon n'importe quel critère. Il doit être possible aussi de retrouver leur structure.

Par exemple le fait qu'un élément chimique possède un symbole, un nom chimique, une masse atomique, etc.

3. L'intérêt d'une BD

L'intérêt d'une BD est de regrouper les données communes à une ou plusieurs applications dans le but:

-d'éviter les redondances et les incohérences que nous aurions si les données étaient réparties dans différents fichiers sans connexions entre eux,

-d'offrir des langages de haut niveau pour la définition et la manipulation des données,

-de partager les données entre plusieurs utilisateurs,

- de contrôler l'intégrité, la sécurité et la confidentialité des données,

-d'assurer l'indépendance entre les données et les traitements (applications).

4. Les caractéristiques d'une BD

Les BD ont les caractéristiques suivantes :

-Grandes : elles ont souvent une taille beaucoup plus grande que la mémoire principale d'un ordinateur,

-Persistantes : elles ont une longueur de vie indépendante de l'exécution des programmes qu'elles utilisent,

-Partagées : elles sont utilisées au même temps par plusieurs utilisateurs et par conséquent il faut disposer :

-de mécanisme d'autorisation d'accès,

-de mécanismes de contrôle de la concurrence.

5. Système de gestion de base de données

Un SGBD (en anglais DBMS pour Database Management System) est un logiciel système qui permet de manipuler (insertion, suppression, mise à jour, recherche efficace) de grandes quantités de données stockées dans une base de données. Ces données peuvent atteindre quelques milliards

d'octet partagé par de multiples utilisateurs simultanément. Les données stockées sont partagées en interrogation et en mise à jour d'une manière transparente. D'autres fonctions complexes peuvent être assurées par le SGBD telle que la protection des données partagées contre les incidents.

Contrairement aux systèmes de fichiers, les SGBD permettent de décrire les données de manière séparée de leur utilisation et de retrouver les caractéristiques d'un type de données à partir de son nom.

6. Architecture d'un SGBD

Un SGBD se compose de trois couches emboîtées de fonctions, depuis les mémoires secondaires vers les utilisateurs.

-Le gestionnaire de fichiers : (encore appelé système de gestion de fichiers SGF), gère les récipients de données sur les mémoires secondaires.

-SGBD interne : son rôle est la gestion des données stockées dans les fichiers, l'assemblage de ces données en objets, le placement de ces objets dans les fichiers, la gestion des liens entre objets et des structures permettant d'accélérer les accès aux objets.

-SGBD externe : La fonction essentielle de cette couche consiste dans la mise en forme et la présentation des données aux programmes d'applications et aux utilisateurs interactifs.

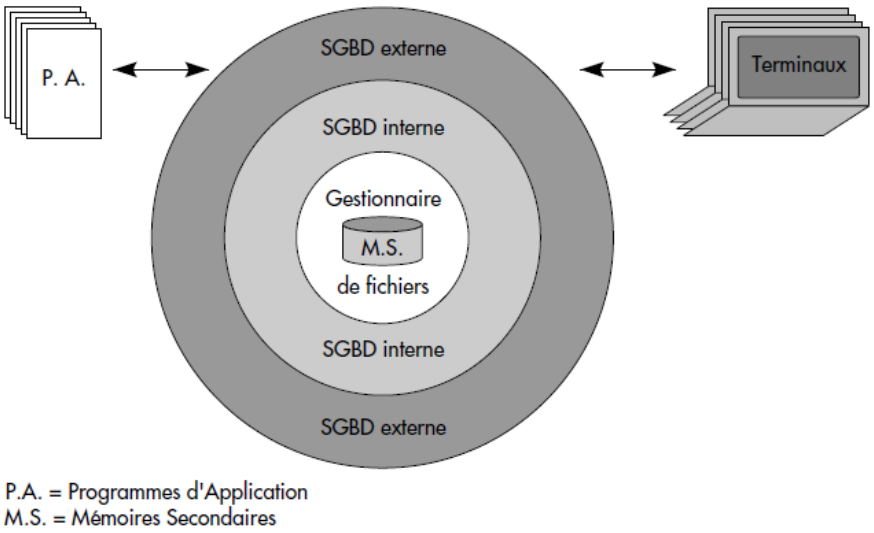


***Figure 1**. Architecture d'un SGBD.*

7. Objectifs des SGBD

-Indépendance physique : permettre le changement du schéma physique (modifier l'organisation physique des fichiers, d'ajouter ou supprimer des méthodes d'accès) sans changer le schéma conceptuel. Cela présente deux avantages : le fait de ne pas manipuler des entités complexes rend les

programmes d'application plus simples à écrire, la modification des applications n'est pas obligatoire dans le cas de modification des caractéristiques du niveau physique.

-Indépendance logique : possibilité de modification du niveau conceptuel sans changement du schéma externe. Cela a deux avantages : pour les programmes d'application du niveau vue, il n'est pas nécessaire d'avoir une vue globale de l'entreprise. En outre, en cas de modification du schéma du niveau logique, les applications du niveau vue sont réécrites seulement si cette modification entraine celle de la vue.

-Manipulation des données : permettre à tous types d'utilisateurs d'accéder à la base selon leurs besoins et connaissances. Par conséquent, un ou plusieurs :

- Administrateurs de la base doivent avoir la possibilité de décrire les données aux niveaux interne et logique,
- Développeurs d'applications écrivent des programmes d'application pour les utilisateurs finaux ou pour eux-mêmes, cela est à partir du niveau conceptuel ou externe,
- Utilisateurs peuvent manipuler les données via un langage simple dont ils ont besoin.

8. Types de modèles de données

8.1. Modèle sémantique

Le modèle sémantique est parmi les modèles de bases de données les moins courants. Il comprend des informations sur la façon dont les données stockées sont rattachées au monde réel.

8.2. Modèle Entité-Association

Le modèle Entité-Association (EA) en français, ER en anglais (Entit Relationship) permet de décrire l'aspect conceptuel des données à l'aide d'entités et d'associations. Le modèle entité/association (E/A) est basé sur une perception du monde réel qui consiste à distinguer des agrégations de données élémentaires appelées entités et des liaisons entre entités appelées associations.

-Entité

Une entité correspond à un objet du monde réel défini en général par un nom. Ces entités sont identifiables de manière unique, interagissent et font ou subissent des actions. Intuitivement, une entité correspond à un objet du monde réel généralement défini par un nom, par exemple un élément chimique, un enseignant, une voiture, une commande, etc. Une entité est une agrégation de données élémentaires.

-Association

Une association correspond à un lien logique entre deux entités ou plus. Elle est souvent définie par un verbe du langage naturel.

Un type d'association peut avoir des propriétés particulières définies par des attributs spécifiques.

-Schéma conceptuel de données

Le langage utilisé pour représenter le modèle E/A est un langage graphique visuel. On représente l'entité par un rectangle contenant au-dessus le nom de l'entité séparé aux propriétés par une ligne. L'association est représentée par un ellipse contenant le nom de l'association séparé aux propriétés par une ligne (figure 2).

Dans la figure on a deux entités : Element_Chim et Molécule. Ces deux entités sont reliées par une association appelée rassemble. Le nom de l'association peut être aussi rassemblé-en et cela tout dépend du sens de la lecture du schéma soit :

-une ou plusieurs molécules rassemblent un ou plusieurs éléments chimiques.

-ou un ou plusieurs éléments chimiques sont rassemblés en une ou plusieurs molécules.

Chacune des deux entités est caractérisée par des propriétés. Pour l'entité Element_Chim, le nombre de propriétés est de huit (Z, symb, nom etc.) alors que pour Molécule, le nombre est de deux. Quant à l'association, elle est caractérisée par une seule propriété (nbre-atom).

Dans la conception des BD, on désigne par schéma conceptuel ou Modèle Conceptuel de Données (MCD) le résultat de la modélisation d'un système d'information d'une entreprise ou d'une application.

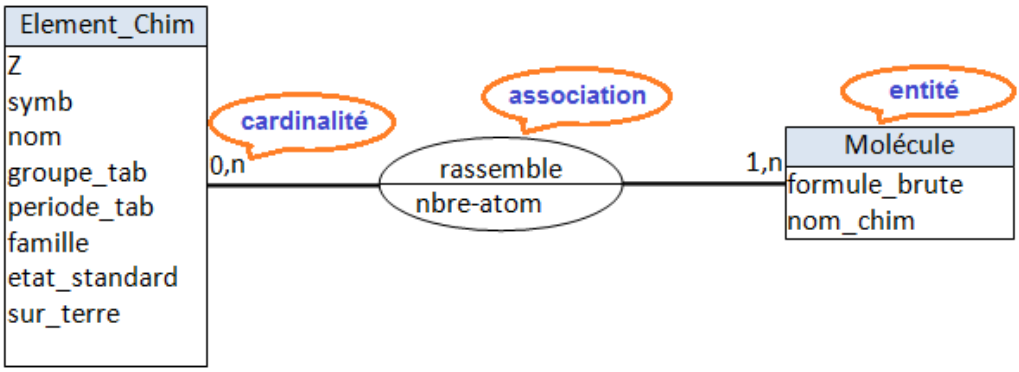


Figure 2 : *Exemple de MCD pour la chimie*

8.3. Modèle hiérarchique

Le modèle hiérarchique organise les données dans une structure arborescente, où chaque enregistrement n'a qu'un seul parent (racine). Les enregistrements frères et sœurs sont triés dans un ordre particulier. Ce modèle convient à la description de plusieurs relations du monde réel.

8.4. Modèle réseau

Modèle réseau ou graphe est un modèle hiérarchique étendu qui autorise relations transverses (i.e. relations plusieurs-à-plusieurs entre des enregistrements liés).

Un enregistrement peut être un membre ou un enfant dans plusieurs ensembles. Cela permet de traduire des relations complexes.

8.5. Modèle relationnel

Le modèle relationnel a été introduit par E. F. Codd en 1970. La première volonté du modèle relationnel fut d'être un modèle ensembliste simple, supportant des ensembles d'enregistrements aussi bien au niveau de la description que de la manipulation.

Le schéma relationnel est l'ensemble des RELATIONS qui modélisent le monde réel ; tel que les relations représentent les entités du monde réel (par exemple : des personnes, des objets, etc.) ou les associations entre ces entités..

9. Le langage SQL

Plusieurs langages permettant de manipuler des bases de données relationnelles ont été proposés, en particulier QUEL [Zook77], QBE [Zloof77] et SQL [IBM82, IBM87].

SQL est le langage informatique standard pour la communication avec les SGBD relationnelles. Il peut être logiquement vu comme étant composé de 3 sous langages: un Langage de Manipulation de Données (LMD), un Langage de Description de Données (LDD) et un Langage de Contrôle de Données (LCD).

Aujourd'hui, le langage SQL est normalisé [ISO89, ISO92] et constitue le standard d'accès aux bases de données relationnelles. Ils comportent quatre opérations de base :

1. la recherche (mot clé SELECT) permet de retrouver des tuples ou parties de tuples vérifiant la qualification citée en arguments ;
2. l'insertion (mot clé INSERT) permet d'ajouter des tuples dans une relation ; les tuples peuvent être fournis par l'utilisateur ou construits à partir de données existant déjà dans la base ;
3. la suppression (mot clé DELETE) permet de supprimer d'une relation les tuples vérifiant la qualification citée en argument ;
4. a modification (mot clé UPDATE) permet de mettre à jour les tuples vérifiant la qualification citée en argument à l'aide de nouvelles valeurs d'attributs ou de résultats d'opérations arithmétiques appliquées aux anciennes valeurs.

10. Base de données chimiques

Une base de données chimique est une base de données spécifiquement dédiée à l'information chimique. La plupart des bases de données chimiques stockent des informations sur des molécules stables. Les grandes bases de données chimiques devraient être capables d'assurer le stockage et la recherche d'informations sur des millions de molécules sur des téraoctets du mémoire physique.

11. Exemple de bases de données chimiques

Il en existe plusieurs milliers, elles regroupent des données plus homogènes établies autour d'une thématique ou d'une méthode spécifique de production des données.

Elles possèdent une grande valeur ajoutée, par la qualité et la quantité des données croisées disponibles.

- Bases de données cristallographiques

Les versions modernes des bases de données cristallographiques sont basées sur le modèle de base de données relationnelle. La communication avec la base de données se fait généralement via un dialecte du SQL (Structured Query Language). Les bases de données Web traitent généralement l'algorithme de recherche sur le serveur en interprétant les éléments de script pris en charge, tandis que les bases de données de bureau exécutent des moteurs de recherche installés localement et généralement précompilés.

- Base de données de la littérature

Les bases de données de la littérature chimique mettent en corrélation des structures ou d'autres informations chimiques avec des références pertinentes telles que des articles universitaires ou des brevets. Ce type de base de données comprend STN , Scifinder et Reaxys . Des liens vers la littérature sont également inclus dans de nombreuses bases de données axées sur la caractérisation chimique.

- Base de données des spectres RMN

Les bases de données de spectres RMN mettent en corrélation la structure chimique avec les données RMN. Ces bases de données incluent souvent d'autres données de caractérisation telles que le FTIR et la spectrométrie de masse.

- Base de données des réactions

La plupart des bases de données chimiques stockent des informations sur des molécules stables , mais dans des bases de données pour les réactions, des intermédiaires et des molécules instables créées temporairement sont également stockés. Les bases de données de réaction contiennent des informations sur les produits et les mécanismes de réaction.

Méthodologie de la recherche d'informations en chimie

Chapitre 5

1. Information chimique

Beaucoup de connaissances chimiques ont été dérivées des données. La chimie doit offrir une gamme riche des données sur les propriétés physiques, chimiques, et biologiques, par exemple, données binaires pour la classification, vraies données pour la modélisation, et données spectrales ayant une densité élevée de l'information. Ces données doivent être introduites dans une forme favorable à l'échange d'information facile et analyse de données.

2. Méthodologie : est un mot qui est composé par trois mots grecs : **metà** (« après, qui suit »), **odòs** (« chemin, voie, moyen ») et **logos** (« étude »). La notion se rapporte aux méthodes de recherche permettant d'arriver à certains objectifs au sein d'une science.

Le terme méthodologie représente l'ensemble des méthodes et techniques mises en place dans un domaine particulier.

«La notion de méthodologie, en tant qu'ensemble de règles et de démarches adoptées pour conduire une recherche, si importante dans l'histoire de la structuration des disciplines scientifiques, est cruciale". (de Mourat et al., 2015) »

3. Méthodologie de la recherche

La méthodologie de la recherche comme objet d'enseignement, est récente et son origine montre en même temps sa nature : elle est une codification des pratiques considérées comme valides par les chercheurs séniors d'un domaine de recherche

4. La méthode scientifique : est l'ensemble de règles et de démarches à suivre pour atteindre des objectifs et pour conduire une recherche scientifique.

La méthode est définit *« Comme l'ensemble des opérations intellectuelles permettant d'analyser, de comprendre et d'expliquer la réalité étudiée. (Jean Louis LAUBET Del Bayle, 2010) ».*

La recherche scientifique est un effort de l'esprit qui vise la connaissance du monde et de l'univers qui nous entoure.

La nécessité est donc forte de renoncer à une définition triviale de la connaissance. *« La connaissance n'est pas l'œil qui regarde et enregistre la chose réduite à son essence simple, son noyau irréductible qui constituerait la substance. La connaissance est une machinerie qui informe le monde – les choses, les plantes, les animaux, les hommes – c'est-à-dire le soumet à son expérimentation, en vue d'un travail de remise en forme » (Dolle, 1984).*

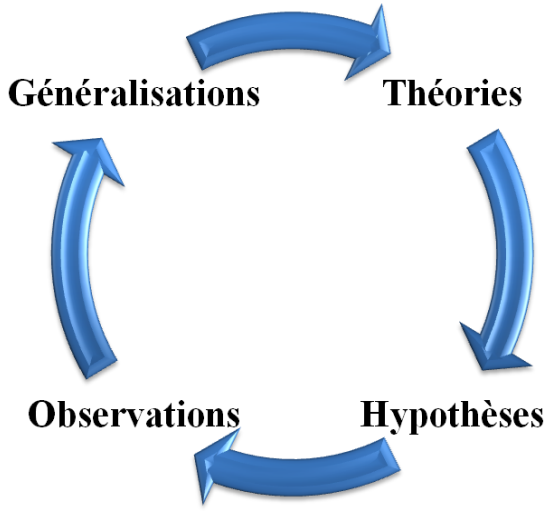


Figure 1 : *La roue de la connaissance scientifique*

La méthodologie scientifique permet la mise en œuvre des exigences théoriques et opératoires de l'observation ; ainsi elle confère aux résultats un fondement légitime. Ce sont donc les façons de procéder, les modes opératoires directs mis en jeu dans le travail de recherche :

-méthode déductive (raisonnement qui va du général au particulier, du principe à la conséquence),

-méthode inductive (raisonnement qui va du particulier au général, des faits aux lois)

-méthode analytique (décomposition de l'objet d'étude en allant du plus complexe au plus simple),

-méthode expérimentale (expériences en laboratoire ou sur le terrain permettant de dégager des lois).

4.1. Les critères d'une méthode scientifique:

-l'utilisation d'un cadre de référence : le cadre de référence cerne les limites de l'étude : on part d'une question, de la définition d'un problème, d'outils, de techniques. Ce critère s'applique surtout aux études descriptives ;

-la compatibilité des données dans un système théorique : ceci réfère aux relations entre les différentes variables retenues au sein d'un cadre théorique donné ; les relations entre les variables devraient mener à une explication d'un phénomène et donc à une simplification de la réalité ; pour ce faire, données quantitatives et données qualitatives pourront être successivement utilisées ;

-le principe de vérification : c'est un principe important en recherche scientifique que les résultats d'une recherche doivent être vérifiables compte tenu d'un protocole de recherche donné ;

-une vision critique et la recherche de l'objectivité ; la dimension critique est essentielle car elle permet de remettre en question les connaissances acquises. Il est important de soumettre les faits et les théories à un examen critique.

5. La recherche

5.1. Définition de la recherche :

Selon l'Oxford English Dictionary (2002), la recherche est définie comme « l'étude systématique des matériaux et des sources afin d'établir des faits et parvenir à des conclusions nouvelles »

McMillan et Schumacher (1997) définissent la recherche comme «un processus systématique de collecte et d'analyse des informations (données) dans un but défini ».

Kerlinger (1986) définit la recherche scientifique comme « un examen systématique, contrôlé, empirique et critique de phénomènes naturels, guidé par la théorie et des hypothèses au sujet de relations présumées entre ces phénomènes ».

5.2. Le travail de recherche :

Le travail de recherche scientifique consiste en une investigation sur un thème spécifique que l'auteur (étudiant ou chercheur) doit développer à partir de son point de vue, en tenant compte des sources d'information nécessaires, pour la réalisation la réalisation d'un «objectif scientifique».

La construction d'un «objet scientifique». Il permet de:

- Découvrir un phénomène
- Résoudre un problème
- Questionner des résultats fournis dans des travaux
- Étudier un nouveau procédé, une nouvelle solution, une nouvelle théorie
- Appliquer une pratique à un phénomène
- De décrire un phénomène
- Expliquer un phénomène

5.3. La recherche scientifique est un développement dynamique ou une démarche rationnelle qui permet d'examiner des phénomènes, des problèmes à résoudre, et d'obtenir des réponses précises à partir d'études. Les fonctions de la recherche sont de décrire, d'expliquer, de comprendre, de contrôler, des phénomènes.

6. La recherche bibliographique

Pour réaliser un travail scientifique dans n'importe quel domaine, il faut établir sa bibliographie, c'est-à-dire dresser la liste des documents utiles à la recherche sur un sujet donné.

6.1. Définition de la recherche bibliographique :

La recherche bibliographique est une démarche méthodologique constituée par des étapes permettant de chercher, identifier, récupérer et traiter des documents et des informations sur un sujet donné.

Se documenter, c'est savoir où et comment trouver l'information, savoir poser les bonnes questions, savoir de quelle information on a besoin, savoir la lire, la comprendre, la critiquer, évaluer si elle répond à ses besoins et savoir la diriger.

6.2 Etapes de la recherche documentaire

6.2.1. Préparation de la recherche

Ce travail préliminaire se déroule en deux phases :

-Analyse du sujet

Ce besoin d'informations diffère selon les individus, en fonction de leurs connaissances antérieures du sujet. Pour cela, il est conseillé d'utiliser la méthode 3QPOC : il s'agit d'une méthode empirique de questionnement, permettant d'affiner au maximum l'objet d'une recherche.

Elle tente de répondre aux questions QUOI, QUI, QUAND, POURQUOI, OÙ et COMMENT.

Il faut donc organiser un questionnement structuré pour le définir au mieux :

- Quoi ? Quel est le thème de votre travail ? Peut-il être décliné en sous-thèmes ?

Quelles en sont les caractéristiques ? Quels en sont les risques ?

- Qui ? Quels auteurs ont travaillé sur votre sujet, votre thème ? Quels acteurs sont concernés par le résultat, par la mise en œuvre ?
- Où ? Où le problème apparaît-il ?
- Quand ? Quelle période est couverte par votre sujet ? De quel délai disposez-vous pour réaliser ce travail ? Quel est le temps disponible pour celui-ci ?
- Comment ? Repère-t-on des modes de fonctionnements spécifiques ? Comment se déroule le processus étudié ?
- Combien ? Quels moyens disponibles ?
- Pourquoi ? Quel est l'objectif de votre travail ? Quel est l'objectif de chaque étape méthodologique (recherche d'informations, études qualitatives, etc) ?

Cette méthode QQOQCCP (Quoi? Qui? Où? Quand? Comment? Combien ? Pourquoi?) appelée « l'Hexamètre mnémotechnique de Quintilien », a été identifiée par les Romains pour résoudre les enquêtes criminelles.

-Elaboration d'une stratégie de recherche

Il s'agit ici d'identifier et de hiérarchiser les ressources susceptibles d'apporter les informations recherchées.

6.2.2. Exécution de la recherche

Il faut alors procéder comme suit :

-Formulation du sujet de recherche

Parce que le langage parlé n'est pas toujours adapté aux outils de recherche documentaire, il est nécessaire de traduire le sujet de la recherche par des **mots-clés**.

- Ecriture des équations de recherche

Il s'agit de combiner les mots-clés définis précédemment afin d'écrire une requête.

D'un point de vue général, poser une requête revient à combiner les mots-clés grâce aux **opérateurs de recherche :**

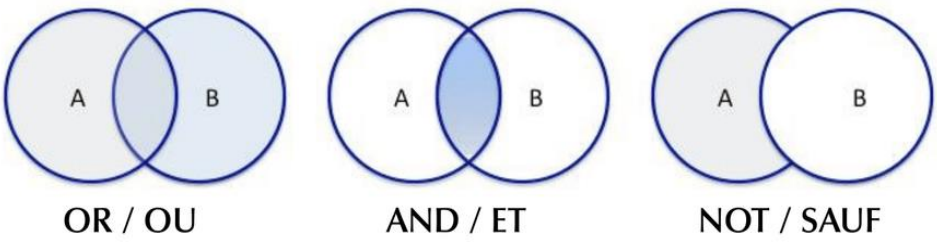
*** Les opérateurs booléens : « ET », « OU », « SAUF »**

Une équation de recherche simple, correctement formulée avec des opérateurs logiques, permet de repêcher plus rapidement des documents pertinents sur le Web.

AND : Recherche très précise, tous les termes reliés par AND doivent être présents dans le document

OR : Recherche très large, au moins un des termes reliés par OR doit être présent dans le document

NOT : Recherche orientée, le terme relié par NOT doit être exclu du document



*** Les troncatures :**

Certains symboles placés en fin de mot permettent d'étendre la recherche à des mots issus de la même famille :

"+" Inclusion : le mot précédé du signe + doit être présent à l'identique dans le document

"-" Exclusion : le mot précédé du signe - est exclu de la recherche

"*" / "?" Proximité : L'astérisque remplace n'importe quelle chaîne de caractère et le point d'interrogation remplace un caractère au début, au milieu ou à la fin d'un terme, ce qui permet d'étendre la recherche

Exemple : pédia* : pédiatrie, pédiatre / wom?n : women, woman

*** La recherche par expression :**

L'utilisation des guillemets "" permet de lancer une recherche sur une "chaîne de caractères".

Elle est particulièrement utile lorsqu'une recherche entraîne un trop grand nombre de résultats ou pour rechercher précisément une expression.

Exemple : " caractérisation de polymères π- conjugués " : recherchera les références contenant cette expression dans l'ordre où sont saisis les termes.

6.2.3. Evaluation des résultats

Cette dernière étape est essentielle puisqu'elle valide la qualité et la pertinence des informations collectées. Seules les informations répondant positivement à ces deux critères sont à exploiter.

-Evaluation de la qualité des sources

-Evaluation de la pertinence des sources

7. Les outils de recherche d'information

L'énorme quantité de données en chimie a mené au développement des outils de données pour stocker et disséminer les données en forme électronique. Il existe plusieurs types de documents : (voir figure 3)

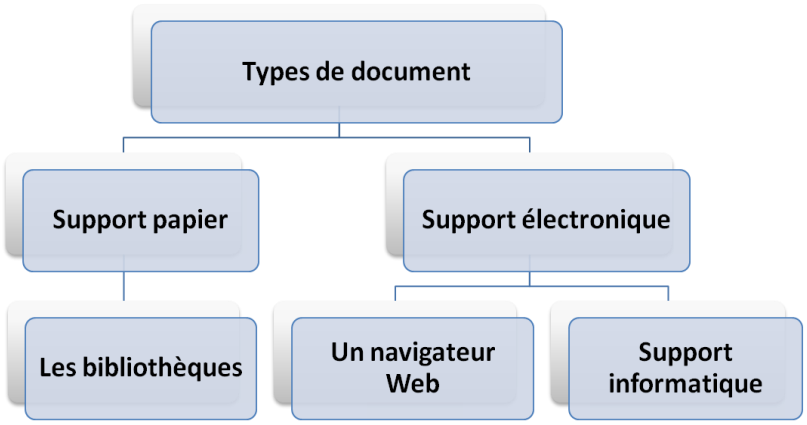


***Figure 3** : Types de document*

Parmi ces types, on peut distinguer :

7.1. Recherche sur les catalogues informatisés

-Catalogue de bibliothèque :

Au sens général, un catalogue est une liste (du grec katalogos : liste).

Concernant une bibliothèque, c'est la liste de tous les documents possédés par cette bibliothèque, quel que soit leur type : livre, thèse, revue ...

Un catalogue permet donc d'identifier un document puis de localiser.

La consultation des catalogues vous permet de repérer les documents possédés par une bibliothèque, même s'ils sont en prêt, en traitement, en commande, quelle que soit leur localisation.

Les catalogues de bibliothèques : ils sont incontournables pour trouver de la documentation papier : ***le catalogue de la bibliothèque universitaire de Biskra***

L'intérêt

-A savoir si la bibliothèque possède les documents dont l'enseignant vous a donné les références.

-A rechercher les ouvrages ou les thèses traitant d'un sujet particulier.

A savoir si la bibliothèque de votre université possède la revue où l'article dont vous avez la référence.

-Portail documentaire de bibliothèque :

Un portail documentaire de bibliothèque est un site internet qui propose un ensemble de services :

-accès au catalogue de la bibliothèque

-interrogation simultanée de diverses ressources (catalogue de la bibliothèque, sites internet, autres catalogues, moteurs de recherche...)

-accès à des documents en ligne en texte intégral (revues en ligne, bases de données, thèses...)

-accès à des services et à des ressources personnalisés en fonction du profil de l'utilisateur (informations ciblées)

-possibilité d'utiliser des applications en ligne pour gérer ses documents

-possibilité de faire des réservations, de suggérer de nouvelles acquisitions, etc.

7.2. Recherche sur les banques de données

Elles sont constituées d'un ensemble structuré de références bibliographiques sur un sujet, un domaine, un type de document, etc. Elles peuvent contenir une analyse, un résumé et de plus en plus souvent l'accès au texte intégral du document lui-même.

- Les banques de données vous permettent de repérer en particulier des articles.

Exemple : sndl

www.sndl.cerist.dz :

Système National de Documentation en Ligne Centre de Recherche de l'Information Scientifique et Technique.

7.3. Recherche sur INTERNET

Internet représente un espace numérique d'information, dans lequel circulent un ensemble considérable de ressources documentaires (informations générales, sommaires, résumés d'articles, journaux électroniques, livres, bases de données, informations sur l'actualité, etc.).

-L'information y est pléthorique, fluctuante, interactive, hétérogène, d'où la nécessité de bien évaluer l'information.

-Les ressources sont innombrables mais leur qualité est extrêmement variable et l'information y est volatile.

Quelques sites recommandés pour la recherche d'informations scientifiques et académiques classés par catégorie :

-Des moteurs de recherche spécialisés

Les moteurs de recherche représentent la mise en pratique de la recherche d'information. Ils sont présents depuis longtemps

-Google Scholar (http://scholar.google.fr/)

-Google Books (http://books.google.fr/)

-Theses.fr (http://www.theses.fr/)

-Exemple : Google Scholar

Google Scholar est un moteur de recherche spécialisé sur le Web. Au lieu de simplement chercher dans Google en espérant trouver des sources fiables et pertinentes, il est préférable d'utiliser Google Scholar afin de trouver des résultats probants.

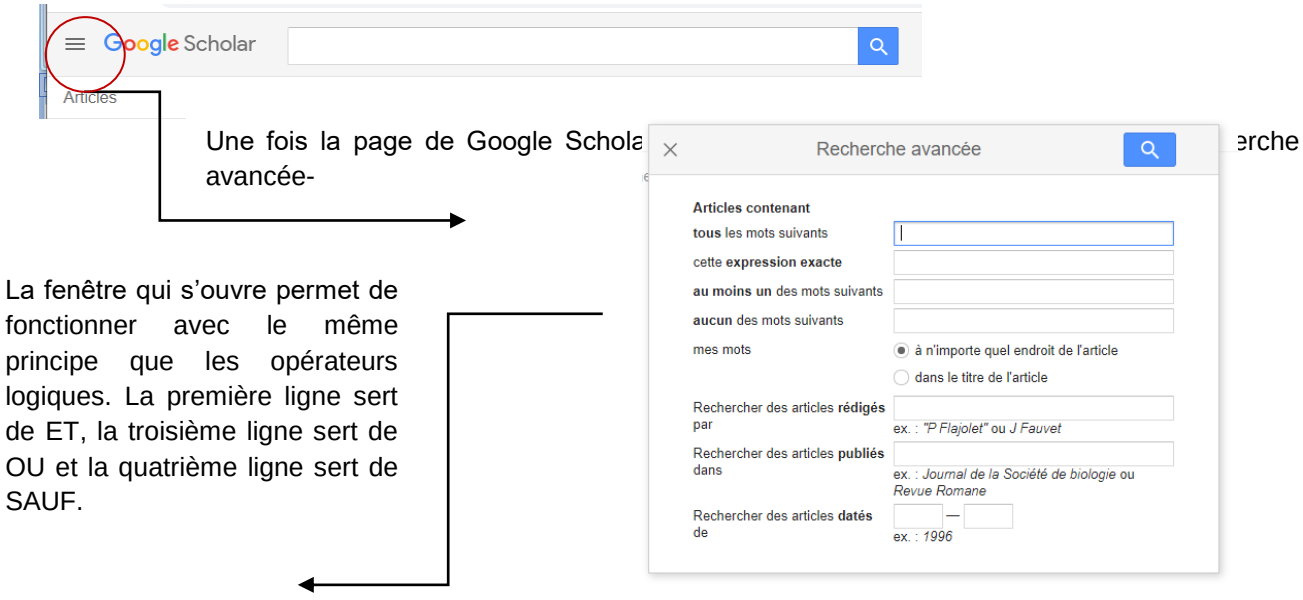


Figure 4 : *Les opérateurs logiques*

-Des portails scientifiques ou thématiques

Les portails d'accès à la littérature scientifique intègrent plusieurs sources différentes dans une même base de données. Ce sont essentiellement des ressources en libre accès mais aussi des ressources que les portails vont directement rechercher sur les sites des éditeurs ou sur des sites officiels (universités, sites gouvernementaux, institutions internationales...).

-Université en ligne (http://uel.unisciel.fr)

-Centre international de recherche scientifique (http://www.cirs.fr)

-Bouquets de revues :

Les bouquets de revues : ensemble de revues électroniques mises en ligne, qu'elles existent déjà sous forme imprimées ou non. Elles peuvent être privées ou publiques, d'accès gratuit ou payant.

Exemple: Science Direct

Service en ligne de l'éditeur de revues scientifiques, techniques et médicales, Elsevier-Masson.

Propose plus de 3200 titres de revues francophones et anglophones.

-Bases de données :

C'est un « Ensemble de données relatif à un domaine défini de connaissance et organisé pour être offert aux utilisateurs.

L'énorme quantité de données en chimie a mené au développement des bases de données pour stocker et disséminer les données en forme électronique. Par exemple, des bases de données ont été développées pour la littérature chimique, composés chimiques, structures 3D, réactions, et spectres. L'Internet est de plus en plus employé pour distribuer des données et l'information en chimie.

Exemple:

PubMed est le principal moteur de recherche de données bibliographiques de l'ensemble des domaines de spécialisation de la biologie et de la médecine.

ChemSpider (Structure, nomenclature, propriétés)

CIR (Chemical Identifier Resolver)

NIST Chemistry WebBook

PubChem

The Cambridge Structural Database (CSD) (voir TP)

Méthodes de recherche

Les capacités de recherche des bases de données diffèrent considérablement. La fonctionnalité de base comprend la recherche par mots-clés, propriétés physiques et éléments chimiques.

La recherche par nom de composé et par paramètres de réseau revêt une importance particulière. Les options de recherche qui permettent l'utilisation de caractères génériques et de connecteurs logiques

dans les chaînes de recherche sont très utiles. Si elle est appuyée, la portée de la recherche peut être limitée par l'exclusion de certains éléments chimiques. Des algorithmes plus sophistiqués dépendent du type de matériau traité. Les composés organiques peuvent être recherchés sur la base de certains fragments moléculaires. Les composés inorganiques, par contre, pourraient présenter un intérêt pour un certain type de géométrie de coordination. Des algorithmes plus avancés traitent de l'analyse de la conformation (organique), de la chimie supramoléculaire (organique), de la connectivité interpolyédrique («non organique») et des structures moléculaires d'ordre supérieur (macromolécules biologiques). Les algorithmes de recherche utilisés pour une analyse plus complexe des propriétés physiques, par exemple les transitions de phase ou les relations structure-propriété, peuvent appliquer des concepts théoriques de groupe.

Certaines méthodes de recherche de base de données sont couramment disponibles:

Nom composé

Peut inclure les noms officiels IUPAC et les noms communs.

Formule moléculaire

Soit une formule exacte, soit une plage.

Structure moléculaire

Cette méthode nécessite une interface d'éditeur moléculaire.

Numéro d'enregistrement

Généralement le numéro de registre CAS, mais la plupart des bases de données ont également leur propre système de numérotation.

Plage de pics ou autres caractéristiques spectrales

L'utilisateur entre numériquement des données liées à un spectre d'un composé inconnu. Ces données sont utilisées pour les composés qui partagent les décalages dans des contraintes spécifiées. Cela permet aux utilisateurs de localiser exactement le composé ou les molécules avec des groupes fonctionnels similaires.

Recherche de spectres

Le logiciel est utilisé pour rechercher dans une base de données des spectres qui ressemblent aux spectres soumis.

Applications locales ; Représentation de la structure 3D

Chapitre 6

1. Représentation moléculaire

Une structure moléculaire est un enchaînement atomique caractérisé par la nature de ces atomes constitutifs de leur ordre, de leur mode de liaison, de leur géométrie interne (distance et angles de liaison) et de leur conformation tridimensionnelle.

La chimie travaille sur des objets qui ne sont pas visible. Pourtant ces objets existent! Il faut les représenter. Plusieurs niveaux de représentation : 2D et 3D.

Exemple : L'eau : trois atomes, deux éléments, deux liaisons, une molécule. Un atome d'oxygène (ici en rouge), se lie à deux atomes d'hydrogène (ici en blanc).

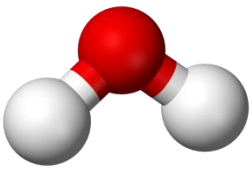


Figure 1 : *Schéma de la molécule d'eau.*

1.1. Représentation en deux dimensions

Une représentation plane donne l'enchaînement des atomes les uns aux autres, sans indiquer leur position dans l'espace :

Formule développée: toutes les liaisons sont représentées.

Formule semi-développée: les liaisons C-H n'apparaissent pas.

Formule de Lewis: tient compte de la valence et de la règle de l'octet.

Formule simplifiée ou topologique: les C et H ne sont pas représentés, seuls les hétéroatomes sont représentés.

1.2. Représentations tridimensionnelles

Les représentations tridimensionnelles font appel à des conventions pour représenter sur une feuille ou un écran une structure en 3D.

-Représentation de Cram

La représentation de CRAM permet de visualiser dans l'espace les liaisons autour du carbone.

-Représentation de Newmann

En représentation de Newman l'axe de la liaison C – H n'apparaît plus. La projection de Newman d'un composé organique est sa représentation sur papier permettant d'étudier ses différentes conformations (on passe d'un conformère à un autre par rotation autour d'une liaison simple carbone-carbone) d'un composé organique. Cette projection est généralement utilisée uniquement avec des atomes de carbone tétravalents (liés à quatre autres atomes).

-Représentation de Fischer

Ce n'est pas une représentation plane. La convention de Fischer est plus utilisée par les biochimistes que par les chimistes. Elle permet une simplification très significative de la nomenclature des monosaccharides.

2. La visualisation

La visualisation tridimensionnelle des molécules sur un écran d'ordinateur ou infographie moléculaire permet de représenter l'image dans l'espace d'un modèle moléculaire. Pour cela, chaque atome constituant la molécule est identifié par ses coordonnées spatiales. Ces dernières sont obtenues par cristallographie, RMN, modélisation moléculaire.

La visualisation des molécules peut être faite dans une grande variété de modes graphiques permettant dans chaque cas de mettre en valeur les informations demandées (boules et bâtons, bâtons, compact,…).

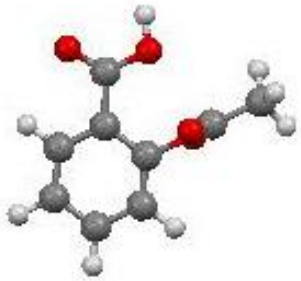
Display : permet de choisir un mode de représentation de la molécule. Il existe plusieurs sortes de modèles moléculaires :

-Wireframe : les atomes sont représentés par des joints par des lignes colorés qui simulent les liaisons. Ce mode de représentation est très simple. Du fait de sa simplicité, elle économise la mémoire et les ressources vidéo de l'ordinateur et accélère l'affichage sur l'écran. Toute fois sa lisibilité diminue avec la taille des molécules.

-Stick (ou bâtons) : C'est une version plus consistante du mode Wireframe ou les liaisons sont indiqués par des bâtons colorés ou des cylindres au lieu des lignes. Ce mode introduit par Dreiding, est surtout utile quand on veut mettre en évidence la géométrie d'un réseau de liaisons

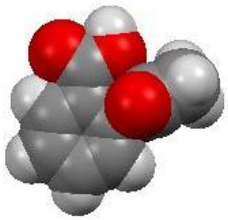
-Ball and stick (boules et bâtons, appelé aussi modèle éclaté) : Dans ce mode, chaque atome est figuré par un sphère de diamètre arbitraire. Les couleurs obéissent au code CPK. Les liaisons sont

représentées par des bâtons de mêmes couleurs que les atomes correspondants. Cette manière de représenter les molécules est surtout utile pour montrer leur composition atomique globale car elle met en évidence le codage des couleurs.



Exemple de modèle éclaté – Aspirine-

-Compact : Ces modèles introduits par Corey, Pauling et Koltun, sont également connus sous le nom CPK. Cette représentation est particulièrement utile pour montrer les volumes et les formes. Ils utilisent simultanément les rayons de Van Der Waals Rvdw (représentant l'encombrement spatial du nuage électronique de l'atome et les rayons de covalence R ' représentant la distance internucléaire entre deux atomes liés.



Exemple de modèle compact – Aspirine-

-La manipulation des molécules sur l'écran :

1- Déplacer l'ensemble de la molécule.

2- Déplacer un atome ou un groupe d'atomes.

3- Couper une molécule.

4- Associer plusieurs molécules.

3. Logiciels utilisés pour la représentation et la visualisation moléculaire

Une gamme entière des méthodes pour la représentation sur l'ordinateur des composés et des structures de produit chimique a été développée comprenant des codes linéaires, des tables de raccordement, et des matrices

Des méthodes spéciales ont dû être conçues pour représenter uniquement une structure chimique, pour percevoir des dispositifs tels que l'aromaticité et pour traiter la stéréochimie, les structures 3D, ou les surfaces moléculaires. Différentes techniques expérimentales permettent de découvrir la structure 3D des structures: RMN, cristallographie aux rayons X, ...

- Quelques logiciels

ACD/ChemSketch Freeware

C'est un logiciel gratuit (Freeware) en anglais qui permet de dessiner des molécules et de les visualiser en 3d.

On peut le télécharger à l'adresse http://www.acdlabs.com/download/chemsk.html (37 Mo il faut remplir un formulaire et fournir une adresse électronique valide).

Guide animé accessible à cette adresse :

http://sciences-physiques.tice.ac-orleans-tours.fr/moodle/file.php/61/ressources/sante/pages-html/tutoriel-chemsketch/index.html

Avogadro est un éditeur de structures chimiques permettant de dessiner une composition moléculaire tri-dimensionnelle, avec plusieurs angles et perspectives.

PyMOL

PyMOL est un logiciel de visualisation moléculaire, édité par la société DeLano Scientific. Il s'agit d'un logiciel libre et gratuit.

Guide animé accessible à cette adresse :

http://www.alchem.org/IMG/PymolIntroduction.pdf

ChemDraw (voir TP)

Initiation à la modélisation moléculaire

Chapitre 7

1. Quelques définitions

1.1. Modèle

Le principe d'un modèle est de remplacer un système complexe en un objet ou opérateur simple reproduisant les aspects ou comportements principaux de l'original. Donc le modèle ensemble de paramètres et de fonctions mathématiques qui permettent une représentation simplifiée de la réalité.

1.2. Modélisation

Transformation d'un ensemble le plus grand possible d'observations expérimentales en un ensemble le plus petit possible de paramètres.

2. Modélisation moléculaire

L'utilisation de méthodes théoriques pour l'obtention de modèles qui puissent prédire et comprendre les structures, les propriétés et les interactions moléculaires est connue sous le nom de «Modélisation Moléculaire ». Celle-ci permet de fournir des informations qui ne sont pas disponibles par l'expérience et joue donc un rôle complémentaire à celui de la chimie expérimentale. Ainsi, la modélisation moléculaire peut par exemple permettre de se faire une idée précise de la structure de l'état de transition pour une réaction donnée, ce qui est difficile, voire impossible, pour la chimie expérimentale.

3. Logiciel de modélisation moléculaire

Un logiciel de modélisation moléculaire comprend de manière générale les modules suivants :

- Construction, visualisation et manipulations des molécules.

-Calculs.

- Sauvegarde des structures et gestion des fichiers.

- Etude des propriétés moléculaire.

Les logiciels de calculs très utilisés:

1- Logiciel ChemOffice ultra :

ChemOffice ultra réunit chem3D ultra, ChemDraw Ultra, chemFinder ultra,….

Chem 3D ultra est un pro-logiciel (produit-Logiciel) de modélisation et de visualisation moléculaire doté d'une nouvelle interface graphique. Il permet de faire la mécanique moléculaire avec MM2 pour l'optimisation de la géométrie et de la dynamique moléculaire, calculs semiempiriques (AM1, PM3, MNDO,….) avec interfaces aux programmes : GAMESS, Gaussian, Tagmar et MOPAC.

Il optimise la géométrie des états de transition et évalue certaines propriétés physiques (dipôles, charges, densités,…).

2- Logiciel Hyperchem

Hyperchem est un logiciel de modélisation moléculaire développé par Autodesk, INC, et distribué par Hypercube INC (Ontario, Canada).

C'est un logiciel de modélisation moléculaire sophistiqué qui est connu pour sa qualité, sa flexibilité, et sa faculté d'usage.

Unissant 3D visualisation et animation, hyperChem peut faire des calculs de la mécanique moléculaire et de la dynamique moléculaire, MM ou a les champs de force (AMBER, MM+, OPLS, BIO+….)

Il offre aussi la possibilité de faire des calculs semi-empiriques (AM1, PM3, CNDO, MINDO3,…) et mêmes des calculs quantiques simples ab-initio (base minimale STO-3G, Small 3-21G, medium 6-21G*, large 6-21 G**, ….DFT.

3- Logiciel Gaussian

Gaussian est un logiciel de chimie numérique crée à l'origine par John Pople et sorti en 1970 (Gaussian 70). Le nom vient de l'utilisation des fonctions gaussiennes pour représenter les orbitales atomiques (OA). Ceci à faciliter le développement de la chimie numérique en particulier les méthodes ab-initio comme Hartree-Fock pour calculer les orbitales moléculaires (OM) à partie des orbitales atomiques (OA).

Gaussian est rapidement devenu un programme sélectif électronique très populaire et largement utilisé. Gaussian peut faire:

-Les calculs de la mécanique moléculaire type AMBER ou Champs de force UHF).

-Les calculs semi-empiriques : AM1, PM3, CNDO

-Théorie de la fonctionnelle densité (DFT) : B3lYP, HF (RHF, UHF).

-La méthode ONIOM (QM/MM).

4. Méthodes de la Modélisation moléculaire

La modélisation moléculaire a pour but de prévoir la structure et la réactivité des molécules ou des systèmes de molécules. Les méthodes de la modélisation moléculaire comprennent : les méthodes quantiques, la mécanique moléculaire et la dynamique moléculaire.

4.1. Mécanique quantique:

La mécanique quantique est le prolongement de la théorie de quanta, elle explique la quantification de certaines grandeurs (énergie, moment cinétique) et fait émerger le principe d'exclusion de Pauli.

Le développement de la mécanique quantique a commencé au début du vingtième siècle avec la découverte de la quantification du rayonnement du corps noir par le physicien allemand Max Planck (prix Nobel de physique en 1918) et par l'explication de l'effet photo-électrique par Albert Einstein (prix Nobel de physique en 1921).

En 1900, Planck détermine la loi de répartition spectrale du rayonnement thermique du corps noir sans en maîtriser l'interprétation physique : l'énergie émise par les atomes entre les états excités est quantifiée alors que la mécanique classique prédit, a contrario, un continuum d'états. En 1905, Einstein expose ses théories sur la nature corpusculaire de la lumière suite à ses études sur l'effet photoélectrique. Il reprend les travaux de Planck et démontre que la lumière se comporte simultanément comme une onde et un flux de particules. L'effet photoélectrique corrobore ainsi l'hypothèse des quantas énergétiques avancée par Planck quelques années auparavant.

De cette dernière et de ses conséquences dont la vision duale de la nature de la lumière, vision qui s'avèrera ultérieurement étendue à toutes les composantes de la matière quantique

Cette dualité onde-corpuscule de la lumière est ensuite généralisée en 1924 par de Broglie à l'ensemble des particules matérielles qui doivent être associées à une onde réelle elle-même reliée à la quantité de mouvement.

La nouvelle conception des particules qui découle de la dualité onde –corpuscule, explicitée dans les travaux de De Broglie (1923) conduit à la mécanique ondulatoire.L'objectif de la mécanique quantique est principalement de déterminer l'énergie et la distribution électronique. Cette approche est ensuite généralisée en 1925 par Schrödinger qui introduit alors son équation éponyme.

La chimie quantique définit la structure moléculaire comme un noyau autour duquel gravitent des électrons, qui sont décrits par leur probabilité de présence en un point et représentés par des orbitales. Les équations de la chimie quantique sont basées sur la résolution de l'équation de Schrödinger qui s'écrit pour les états stationnaires.

Résolution de l'équation de Shrödinger

Pour décrire un système à l'échelle atomique et subatomique, ainsi que répondre à des problématiques où la mécanique newtonienne échouait, la mécanique quantique s'est développée au début du XX^{ème} siècle. Dans les théories quantiques, les états fondamental et excités d'un système, à

un instant t et à un point $\vec{r}$ de l'espace, peuvent être décrits par des fonctions $\Psi_k(\vec{r}, t)$, appelées fonctions d'ondes et où k décrit les états énergétiques d'un système. L'évolution temporelle et spatiale du système est décrite par l'équation dépendante du temps proposée par Schrödinger en 1926:

$$\hat{H}\Psi_k(\vec{r}, t) = i\hbar \frac{\partial}{\partial t}\Psi_k(\vec{r}, t) \quad \textbf{1}$$

où $\hat{H}$ désigne l'opérateur hamiltonien. Si nous souhaitons connaître à un instant t donné, l'évolution spatiale du système, l'équation se simplifie par :

$$\hat{H}\Psi_k(\vec{r}, t) = E_k\, \Psi_k(\vec{r}, t) \quad \textbf{2}$$

C'est l'équation de Schrödinger indépendante du temps. Les termes E_k sont les valeurs propres de $\hat{H}$associées aux vecteurs propres $\Psi_k(\vec{r})$, et correspondent aux énergies des différents états.

La plus petite des valeurs propres, $|E_0|$, détermine l'énergie de l'état fondamental du système décrit par $\Psi_0(\vec{r})$.

L'étude quantique d'un système a pour but de déterminer les $\Psi_k(\vec{r})$, en résolvant exactement l'équation **2**. Cependant, la résolution exacte de cette équation est impossible pour les systèmes à trois particules et plus, ce qui nous oblige à travailler avec des solutions approchées. L'expression exacte de l'hamiltonien $\hat{H}$ d'un système isolé à N électrons aux positions $\vec{r}_i$(i =1..N) et M noyaux aux positions $\vec{R}_\alpha$ (α = 1..M), s'écrit :

$$\hat{H}\Psi_k(\vec{r}, t)=\frac{-1}{2}\sum_i^N \nabla_i^2 - \sum_\alpha^M \nabla_i^2 - \frac{1}{2}\sum_i^N \sum_\alpha^M \frac{Z_\alpha}{|\vec{r}_i-\vec{R}_\alpha|} + \frac{1}{2}\sum_i^N \sum_{i\neq j}^N \frac{1}{|\vec{r}_i-\vec{R}_{j\neq i}|} + \frac{1}{2}\sum_\alpha^M \sum_{\beta\neq\alpha}^M \frac{Z_\alpha Z_\beta}{|\vec{r}_\alpha-\vec{R}_\beta|} \quad 3$$

ou plus succinctement, en faisant correspondre termes à termes :

$$\hat{H} = \hat{T}_e + \hat{T}_n + \hat{V}_{e-n} + \hat{V}_{e-e} + \hat{V}_{n-n} \quad 4$$

où $\hat{T}_e$décrit l'énergie cinétique des électrons et $\hat{T}_n$ celle des noyaux. $\hat{V}_{e-n}, \hat{V}_{e-e}, \hat{V}_{n-n}$ Représentent respectivement l'interaction coulombienne entre les électrons et les noyaux, celle entre les électrons, et celle entre les noyaux.

- Approximation de Born-Oppenheimer

Une première approximation pour la résolution de l'équation de Schrödinger a été proposée par M. Born et R. Oppenheimer en 1927. La masse d'un noyau étant environ 1836 fois plus élevée que celle d'un électron, les noyaux peuvent être considérés comme quasi-immobiles visà-vis des électrons, et leurs mouvements décorrélés de ceux des électrons. Dans ce cas, l'énergie cinétique des noyaux est négligeable et $\hat{V}_{n-n}$ est une interaction constante, indépendante des électrons. Nous obtenons alors un nouvel hamiltonien électronique,

$\hat{H}_{elec}$ qui est la somme de $\hat{T}_e, \hat{V}_{e-n}, \hat{V}_{e-e}$ et qui nous permet de décrire les états électroniques du système $\Psi_0^{\text{elec}}(\vec{r})$ tels que :

$$\hat{H}_{elec}\Psi_{\text{k}}^{\text{elec}}(\vec{r}) = E_k^{elec}\Psi_{\text{k}}^{\text{elec}}(\vec{r}) \quad \textbf{5}$$

Ainsi l'énergie totale du système est la somme de l'énergie électronique et celle des noyaux E_n, telle que :

$$E_n = \frac{1}{2}\sum_{\alpha}^{M}\sum_{\beta\neq\alpha}^{M}\frac{Z_\alpha Z_\beta}{|\vec{r}_\alpha - \vec{R}_\beta|} \quad \textbf{6}$$

$$E_k = E_k^{elec} + E_n \quad \textbf{7}$$

Par la suite, puisque toutes les résolutions et raisonnements se concentrent sur la partie électronique, $\hat{H}$ désignera l'hamiltonien électronique, $\Psi_k(\overrightarrow{r_1}; :::; \overrightarrow{r_N})$ la fonction d'onde électronique du k^{ime} état énergétique et E_k son énergie.

-Approximation orbitalaire

L'hamiltonien électronique n'est encore solvable que pour les systèmes hydrogénoïdes, c'est-à-dire ne possédant qu'un électron. Pour résoudre des systèmes possédant n électrons, l'équation de Schrödinger électronique est transformée en un système d'équations monoélectroniques en considérant un modèle de particules indépendantes.

L'approximation orbitalaire, introduite par Hartree en 1928 consiste à développer la fonction d'onde d'un système multi-électronique en un produit de spin-orbitales monoélectroniques supposées normalisées, ceux sont les orbitales atomiques s, p, d, f…des atomes hydrogénoïdes.

La fonction d'onde Ψ devient le produit de Hartree:

$$\Psi(1,2,\dots) = \varphi_1(1)\varphi_2(2)\dots\varphi_n \quad \textbf{8}$$

Sachant que chaque spin-orbitale est le produit d'une fonction de position de l'électron ϕi et d'une fonction de spin η(si).

$$\varphi(n_i) = \phi_{\text{i}}(\text{r}_{\text{i}})\,\eta(\text{si}) \quad \textbf{9}$$

On associe à la fonction de spin $\eta(s_i)$ deux formes : α pour le spin +1/2 et β pour le spin -1/2.

Où chaque ϕ_i ne dépend explicitement que des coordonnées d'un seul électron. Les $\{\phi_i\}$ Solutions d'un hamiltonien monoélectronique, forment une base orthonormée.

L'utilisation de cette approximation n'implique pas qu'on néglige l'énergie de répulsion électron-électron, ce qui serait inacceptable si on prétend à des résultats quantitatifs. Mais on se contentera d'évaluer leur énergie moyenne de répulsion en calculant l'énergie de l'état décrit par $\varphi_1\varphi_2$ avec l'hamiltonien complet H.

En 1930, Fock démontre que la méthode de Hartree ne respecte pas le principe d'antisymétrie de la fonction d'onde. En effet, d'après le principe d'exclusion de Pauli, deux électrons ne peuvent pas être simultanément dans le même état quantique.

Ainsi Slater propose d'écrire la fonction d'onde comme un déterminant des n spinorbitales.

L'approximation orbitalaire telle qu'est exprimée par l'équation **8**, ne respecte pas le principe de Pauli puisqu'elle ne satisfait pas à l'exigence d'antisymétrie. Si les coordonnées 1 et 2 sont permutées, φ(1) φ(2) est remplacé par φ(2) φ(1) qui ne présentent en général aucune relation entre eux. C'est pourquoi Fock a proposé d'écrire la fonction d'onde totale Ψ sous forme d'un déterminant, appelé le déterminant de Slater:

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1(1) & \varphi_2(2) \dots & \varphi_N(1) \\ \varphi_1(2) & \varphi_2(2) \dots & \varphi_N(2) \\ \dots & \dots & \dots \\ \varphi_1(N) & \varphi_2(N) \dots & \varphi_N(N) \end{vmatrix} \qquad \textbf{10}$$

Si on permute les coordonnées 1 et 2, on permute 2 lignes du déterminant et on change donc son signe. Par ailleurs, si deux électrons sont dans le même état quantique (même spin-orbitale), par exemple φ1 = φ2, le déterminant présente deux colonnes identiques et s'annule.

On retrouve ainsi le principe d'exclusion de Pauli.

Le développement de ce déterminant est une somme de N!

- Approximation CLOA :

La partie d'espace Φ des spin-orbitales peut être prise a priori sous la forme d'une combinaison linéaire d'orbitales atomiques c dont l'ensemble constitue une base (normée, mais pas orthogonale). C'est le cas dans la quasi-totalité des programmes de chimie quantique.

Chaque orbitale moléculaire s'exprime comme une combinaison linéaire de fonctions de base χ_p

$$\Phi_i = \sum_{p=1}^{M} C_{pi}\chi_p \qquad \textbf{11}$$

Ce sont les coefficients c_{pi} qui sont les paramètres variationnels du calcul HF

La méthode CLOA exprime les orbitales moléculaires sous la forme d'une combinaison linéaires d'orbitales centrées sur chaque noyau, appelées par commodité orbitales atomiques (OA).

La méthode est connue sous le sigle CLOA, en anglais LCAO. En réalité, ces orbitales de base peuvent être assez différentes des seules orbitales atomiques connues exactement, à savoir les OA de l'hydrogène et des hydrogénoïdes. Le choix de cette base est un des éléments essentiels de la qualité du résultat des calculs de chimie quantique.

4.1.1 Méthodes Méthode quantique : Ab initio

Les travaux effectués au début du vingtième siècle par Planck, Einstein, Bohr, De Broglie, Schrödinger et Heisenberg ont abouti à l'élaboration de la mécanique des microsystèmes.

En 1925, grâce aux efforts de W. Heisenberg et E. Schrödinger et de P. Dirac, J. Von Neumann, N. Bohr, M. Born et d'autres, une nouvelle mécanique a été crée : la Mécanique quantique, qui a permis d'expliquer de nombreuses propriétés physiques, telles que les propriétés chimiques des éléments et la formation des liaisons chimiques.

La résolution exacte de l'équation Schrödinger n'est possible que pour l'atome d'hydrogène et les systèmes hydrogénoïdes. Pour les systèmes polyélectroniques, on fait appel aux méthodes d'approximation. Les principales variantes sont la méthode de Huckel et les méthodes de champ auto cohérent (Self consistent Field, SCF).

Les méthodes ab initio sont caractérisées par l'introduction d'une base arbitraire pour étendre les orbitales moléculaires et alors le calcul explicite toutes les intégrales exigées qui impliquent cette base. Les calculs ab initio peuvent être exécutés au niveau d'approximation de Hartree-Fock, qui est équivalent à un calcul du champ auto – cohérant SCF (Self Consistent Field). L'option et les niveaux d'Hartree-Fock incluent les effets de corrélation qui n'est pas incluse au niveau d'approximation d'Hartree-Fock d'une solution non- relativiste pour l'équation de Schrödinger.

-Approximation de Hartree-Fock

Encore appelée approximation du champ self consistant SCF, Hartree (1928), Fock (1930) elle fut proposée en 1928 par Hartree. Le problème électronique est un problème multi corps et du fait de la présence des termes d'interaction entre les électrons, il est impossible de traiter séparément les différents électrons.

La méthode HF consiste à approcher l'énergie de l'état fondamental en restreignant la minimisation du principe variationnel à des fonctions d'onde Φ correspondant exclusivement à des déterminants de Slater:

$$E_0 = \min \Psi \left\langle \Psi \middle| \hat{H} \middle| \Psi \right\rangle \rightarrow E_{HF} = \min \left\langle \Phi \middle| \hat{H} \middle| \Phi \right\rangle \qquad \textbf{12}$$

Notons que $E_0 - E_{HF} = E_c < 0$ (énergie de corrélation électronique)

La méthode de Hartree-Fock permet une résolution approchée de l'équation de Schrödinger d'un système quantique à n électrons et N noyaux dans laquelle la fonction d'onde poly-électronique ψ_{HF} est écrite sous la forme d'un déterminant de Slater composé de spinorbitales mono-électroniques qui respecte l'antisymétrie de la fonction d'onde:

$$\Psi = \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1(1) & \varphi_2(2) \dots & \varphi_N(1) \\ \varphi_1(2) & \varphi_2(2) \dots & \varphi_N(2) \\ \dots & \dots & \dots \\ \varphi_1(N) & \varphi_2(N) \dots & \varphi_N(N) \end{vmatrix}$$

Cette méthode itérative est connue sous le nom de méthode du champ autocohérent(SCF).Toutefois, la méthode de Hartree-Fock souffre d'un inconvénient majeur: dès lors que la répulsion électronique est moyennée, une partie de la corrélation électronique est négligée.

- Méthode post-Hartree-Fock

Les méthodes les plus importantes, dans l'utilisation courante, pour introduire la corrélation électronique sont d'une part les méthodes appelées post-HF avec l'interaction de configuration (CI), les méthodes de « coupled cluster »(CC) et les « many-body perturbation theory » (MP2, MP4, …) La résolution des équations d'HF donne une fonction d'onde de référence sous la forme d'un déterminant de Slater. La fonction d'onde de la méthode CI est une combinaison linéaire de déterminants de Slater représentant l'état fondamental et des configurations excitées.

Ces dernières correspondent à l'excitation d'un, de deux ou plus électrons d'une orbitale occupée à une virtuelle. Les coefficients de cette combinaison linéaire sont déterminés selon le principe variationnel; l'énergie qui en découle est donc une limite supérieure à l'énergie exacte.

L'IC totale ("Full CI") est la limite que l'on peut atteindre dans une base donnée, c'est-à-dire qu'elle comprend toutes les excitations possibles des n électrons. Cependant l'IC totale accroît énormément les calculs avec le nombre d'électrons et la dimension des bases utilisées. Pour ces raisons de tels calculs servent habituellement de référence pour des systèmes comprenant un petit nombre d'électrons ($n \leq 20$). La méthode CI limitée à un nombre d'excitation inférieure à la totalité des possibilités ("truncated CI) n'est pas "size consistent" (c'est à dire, par exemple, que l'énergie d'un dimère placé à distance infinie n'est pas égale à deux fois l'énergie du monomère). Cette propriété importante est garantie par des méthodes non variationnelles comme "many-body perturbation theory" ou "coupled cluster methods".

Dans le formalisme de Moller-Plesset, l'hamiltonien est représenté comme la somme de l'hamiltonien HF de l'état fondamental et d'une différence, entre l'hamiltonien exact et HF, traitée comme une perturbation. Dans la théorie de perturbation, la fonction d'onde et l'énergie, pour un état donné, donnent l'état appelé "zéro" ou fondamental auquel est ajouté des corrections successives résultant des différents ordres de perturbation pour le traitement du système. L'approximation MP2 comprend des substitutions simples et doubles, la théorie des perturbations d'ordre 4 (MP4) ajoute des substitutions triples et quadruples. Les méthodes "coupled cluster" (CC) sont actuellement les plus puissantes des méthodes ab initio. La théorie CC commence par un postulat sur la fonction d'onde à n électrons à savoir que cette fonction d'onde est multipliée par une exponentielle naturelle d'une somme d'opérateurs d'excitation d'électrons (T); T2 implique une double substitution. CCSDT comprend jusqu'à un opérateur de triple excitation.

4.1.2. Méthodes semi empiriques

La mécanique quantique est une technique mathématique rigoureuse basée sur l'équation de Schrödinger. La solution de cette équation permet d'obtenir des informations précises sur les propriétés géométriques et électroniques de la molécule. Les calculs peuvent être de type ab initio ou semi-empirique (ex : CNDO, PM3). En ab initio, on tient compte de tous les électrons de la molécule et on vise une solution rigoureuse de l'Hamiltonien.

Les calculs semi-empiriques traitent seulement les électrons de valence et utilisent un Hamiltonien plus simple ayant des facteurs de correction basés sur des données expérimentales. L'équation de Schrödinger d'un système moléculaire peut être résolue sans approximation (ab initio) ou en introduisant des approximations (semi-empirique). En mécanique quantique, on se préoccupe de la distribution des électrons (orbitales) dans l'espace. Les meilleurs programmes comportent des processus d'optimisation de la géométrie.

Les modèles semi-empiriques de valence ne portent que sur les électrons de valence, Ils supposent que ces électrons sont soumis à un écran connu et constant des électrons de cœur. Ils se fondent aussi sur un espace actif minimal : l'espace de valence. Il n'y a pas d'espace de cœur n_i d'espace externe. Le développement des orbitales se limite alors à une base minimale de valence comportant autant de fonctions atomiques que d'orbitales atomiques de valence des atomes constituant la molécule. Les différentes méthodes semi-empiriques vont se différencier suivant le type d'approximation utilisée. Il existe cependant des points communs en toutes ces méthodes.

Pour les systèmes polyélectroniques, l'équation de Schrödinger n'est pas résolue car on traite ici d'un problème à N-corps. Il faut donc aussi faire des approximations orbitalaires en tenant compte de chaque électron de façon indépendante. On parle alors d'un développement linéaire de combinaisons d'orbitales atomiques « LCAO» pour chaque électron.

Les orbitales de la molécule sont développées en combinaisons linéaires des orbitales atomiques (OA) de valence.

$$\Phi_i = \sum_{p=1}^{M} C_{pi} \aleph_p$$

Où les Φ_i sont les orbitales de valence, et les χ_p les OA de valence.

Deux variantes sont d'un très bon rapport qualité/prix et sont largement utilisées dans le calcul des molécules organiques, les méthodes AM1 et PM3, PM6, PM7 disponibles dans le programme GAUSSIAN

4.1.3 Théorie de la fonctionnelle de la densité (DFT)

La théorie de la fonctionnelle de la densité (DFT d'après les initiales en anglais) a pour objet de décrire un système en considérant la densité $\rho(\vec{r})$ comme variable de base. Ainsi, le problème à *n*

électrons est étudié dans l'espace de $\rho(\vec{r})$ qui est de dimension 3 au lieu de l'espace de dimension 3*n* de la fonction d'onde $|\Psi\rangle$.

Historiquement, les premiers à avoir exprimé l'énergie en fonction de la densité furent L. H. Thomas et E. Fermi en 1927. Dans leur modèle, les interactions électroniques sont traitées classiquement et l'énergie cinétique est calculée en supposant la densité électronique homogène. Ce modèle, même amélioré par P. A. Dirac avec un terme d'échange, ne permet pas de rendre compte de la stabilité des molécules vis à vis des dissociations. Un peu plus tard, J. C. Slater proposa un modèle basé sur l'étude d'un gaz uniforme améliorée avec un potentiel local. Cette méthode, appelée Hartree-Fock-Slater ou X_α, fut essentiellement utilisée en physique du solide. Mais la DFT a véritablement débuté avec les théorèmes fondamentaux de Hohenberg et Kohn en 1964 qui établissent une relation fonctionnelle entre l'énergie de l'état fondamental et sa densité $E[\rho(\vec{r})]$.

Théorie de Hohenberg-Kohn-Sham

La densité électronique d'un système à *n* électrons associée a une fonction d'onde $\Psi(\vec{r}, \vec{r}_2, ..., \vec{r}_n)$ est donnée par l'expression suivante : $\rho(\vec{r}) = n\int |\Psi(\vec{r},\vec{r}2,\ldots,\vec{r}n)|^2 d\vec{r},\vec{r}2,\ldots,\vec{r}n$,

elle est normée au nombre d'électrons : $\int \rho(\vec{r})d\vec{r} = n$.

La DFT repose sur deux théorèmes de Hohenberg et Kohn, initialement d'démontres pour un état fondamental non dégénéré. Ces théorèmes démontrent que toutes les propriétés d'un système dans son état fondamental sont complètement déterminées par sa densité´ $\rho(\vec{r})$, elle-même étant obtenue par un principe variationnel appliqué à la fonctionnelle énergie du système $E[\rho]$.

Les différents types de fonctionnelles d'échange-corrélation

Differents types d'approximation de la fonctionnelle d'échange-corrélation E_{xc} ont été développées. On peut les regrouper par générations.

-La première génération est celle de l'approximation de la densité locale (LDA, d'après son acronyme anglais « (*Local Density Approximation*) »). Elle consiste a supposer que la densité $\rho(\vec{r})$ est localement uniforme, si bien que E_{xc} s'écrit :

$$E_{xc} = \int \rho(\vec{r})\, \varepsilon_{xc}[\rho(\vec{r})]d\vec{r} \qquad \textbf{13}$$

ou` $\varepsilon_{xc}[\rho]$ est la densité d'énergie d'échange-corrélation par électron. La plupart des calculs tiennent compte de la polarisation de spin, et l'approximation locale s'appelle dans ce cas LSDA pour Local Spin Density Approximation.

-Les fonctionnelles de la deuxième génération dépendent à la fois de la densité $\rho(\vec{r})$ et de ses dérivées. L'idée est de faire un developpement en gradient de la densité (appelé GEA pour Gradient Expansion Approximation) : $E_{xc}^{GEA}[\rho] = \int \rho(\vec{r})\varepsilon xc(\rho(\vec{r}))d\vec{r} + \int B_{xc}(\rho(\vec{r}))|\nabla\rho(\vec{r})|^2 d\vec{r} + \ldots$ **14**

Les premiers résultats obtenus avec ces méthodes se sont avères nettement moins bons que ceux de LSDA. Les raisons sont notamment reliées au fait qu'une fonctionnelle quelconque de type GEA ne vérifie pas les règles de somme, contrairement a la fonctionnelle LSDA.

-La troisième génération des fonctionnelles est celle des fonctionnelles hybrides basées sur la méthode de la connexion adiabatique. Ces fonctionnelles prennent mieux en compte l'énergie d'échange. Les ´ fonctionnelles hybrides contiennent a la fois un terme d'échange calculé en DFT. Ainsi, Becke a développé des fonctionnelles de la forme:

$$E_{xc} = aE_x^{exact} + (1-\alpha)E_x^{LDA} + bE_x^{GGA} + cE_c^{GGA} \qquad \textbf{15}$$

Ou les paramètres a, b, c sont optimises sur un jeu de valeurs connues.

La partie d'échange est en général la fonctionnelle de Becke (B), la partie de corrélation celle de Lee, Yang et Parr (LYP) ou celle de Perdew-Wang (PW) avec les variantes 86 et 91, d'où finalement les mots-clés BLYP, BPW86 et BPW91.

L'approximation GGA a fait ses preuves dans de très nombreux cas et est connue pour donner de meilleurs résultats que la LDA, notamment pour les systèmes magnétiques. Les systèmes avec des fortes variations de densité électronique sont ainsi décrits plus correctement.

Enfin, il s'est avéré que dans les méthodes LDA, il y avait du bon à prendre, que d'autre part, comme on l'a vu, la méthode HF traitait correctement l'énergie d'échange, d'où des méthodes hybrides basées sur une combinaison empirique de ces énergies avec l'énergie GGA. La plus répandue est la méthode de « Becke à trois paramètres » (B3) ; ainsi, la fonctionnelle B3LYPutilise la fonctionnelle LYP pour la partie GGA. Les paramètres ont été ajustés pour reproduire les valeurs des énergies d'atomisation. La partie GGA peut être également les fonctionnelles PW91 et PW86.

4.2. Mécanique moléculaire

L'expression ''Mécanique moléculaire'' désigne actuellement une méthode de calcul qui permet, d'obtenir des résultats de géométrie d'énergie moléculaires en se basant sur la mécanique classique. La mécanique moléculaire est apparue en 1930, mais s'est développée à partir des années soixante, quand les ordinateurs furent plus accessibles et plus performant. Les méthodes de la mécanique moléculaire permettent le calcul de propriétés structurales et thermodynamiques de systèmes moléculaires comportant jusqu'à plusieurs milliers d'atomes. Les électrons n'y sont pas traités explicitement comme dans un calcul de mécanique quantique, mais les atomes y sont représentés par des masses ponctuelles chargées reliées les unes aux autres par des ressorts. Contrairement à la mécanique quantique, l'énergie des systèmes moléculaires ne provient pas de la résolution de l'équation de Schrödinger, mais est décrite par les fonctions empiriques auxquelles sont associés des paramètres dérivant de l'expérience ou de calculs précis quanto-chimiques. Le « champ de force »

établi par cette méthode représente aussi bien que possible les variations de l'énergie potentielle avec la géométrie moléculaire. L'énergie de la molécule est exprimée sous la forme d'une somme de contributions associées aux écarts de la structure par rapport à des paramètres structuraux de référence :

$E = E_{liaison} + E_{angle} + E_{dièdre} + E_{van\ der\ Waals} + E_{électrostatiq}$

Les variables du calcul sont les coordonnées internes du système : longueurs de liaisons, angles de valence, angles dièdres, ainsi que les distances entre atomes non liés dont les interactions sont représentées par un potentiel de Van der Waals et un potentiel électrostatique le plus souvent de type Coulombien. Habituellement, on distingue dans l'équation de l'énergie du champ de force les termes intramoléculaires concernant les atomes liés chimiquement (liaisons, angle, dièdres, etc.) des termes intermoléculaires concernant les interactions entre les atomes non-liés chimiquement (électrostatiques, Van der Waals, etc.).

4.3. Dynamique moléculaire

La dynamique moléculaire (DM) est une méthode de simulation du mouvement des atomes et des molécules en calculant leurs déplacements. Cette technique est largement utilisée pour simuler les propriétés des solides, des liquides, et des gaz. Elle Permet de calculer les propriétés physico-chimiques d'un système sachant que le potentiel d'interaction est connu.

La dynamique moléculaire utilise la seconde loi de Newton pour décrire le mouvement d'une molécule en fonction du temps.

$$\vec{F}_i = \boldsymbol{m_i}\vec{a}_i$$

$\vec{F}_i$: est le vecteur force agissant sur l'atome i.

m_i :est la masse de l'atome i.

$\vec{a}_i$:est le vecteur d'accélération de l'atome i.

Cette équation montre que la vitesse et le sens du mouvement atomique dépendent des forces qui s'exercent entre les atomes.

La force Fi qui s'exerce sur un atome i se trouvant en position $r_i(t)$ est déterminée par : $\vec{F}_i = \frac{d\vec{E}(r_i)}{dr_i}$

L'énergie potentielle totale du système, se calcule en utilisant les techniques de la mécanique moléculaire.

5. Bases atomiques :

Une base en chimie quantique est un ensemble de fonctions utilisées afin de créer des orbitales moléculaires, qui sont développées comme combinaison linéaire de telles fonctions avec des poids à déterminer. Ces fonctions sont habituellement des orbitales atomiques.

5.1. Bases minimales

Sont constitués de nombre minimum de fonctions de base requis pour représenter tous les électrons de chaque atome. Une base minimale est celle dans laquelle est utilisée pour chaque atome du système, une seule fonction de base est utilisée pour chaque orbitale.

Les bases minimales les plus courantes sont STO-nG ou n est un nombre entier.

$NY_{lm}(\theta,\varphi)r^{n-1e^{-\xi}}$

L'idée la plus simple est de partir des seules orbitales connues sans approximation, les OA de l'hydrogène et des hydrogénoïdes, en se limitant aux OA occupées et aux orbitaux vides de la couche de valence : 1s pour H, 1s, 2s, 2p pour C, N, etc. On a ainsi une base minimale. Par analogie avec les OA de l'hydrogène $\langle n,l,m\rangle = NY_{lm}(\theta,\varphi)r^{n-1p(r)^{n-1}e^{(-\frac{2r}{na_0})}}$

Cependant, dans ce type de fonction, l'exponentielle pose de grandes difficultés dans le calcul des intégrales lorsque plus de deux atomes sont présents. On la remplace donc généralement par une gaussienne exp(-αr2) avec α> 0. L'exponentielle : elle décroît plus vite quand on s'éloigne du noyau, mais surtout elle n'a pas le même comportement pour r = 0 (par exemple, sa dérivée est nulle). Elle est donc remplacée par une combinaison linéaire, en général de trois gaussiennes (voir figure 1). On parle de base STO- 3G, orbitales de Slater « approximée » par 3 gaussiennes.

Cette base est l'option par défaut dans le programme Gaussian. Elle est d'une qualité médiocre pour des résultats quantitatifs, mais peut être utilisée pour obtenir rapidement les représentations des OM. Nous l'utiliserons aussi dans la suite pour des calculs modèles d'une interprétation aisée.

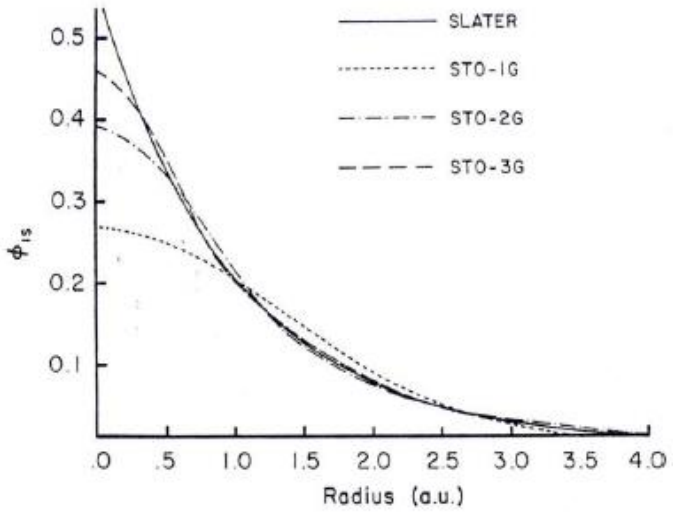


Figure1 : *Approximation d'une fonction de Slater par une deux et trois gaussiennes*

5.2. Bases étendues

Dans les bases les plus utilisées, la partie radiale de chaque OA est représentée une combinaison linéaire de n gaussiennes :

$$\sum_{p=1}^{M} d_i \, e^{-ar^2}$$

Les OA sont adaptées aux atomes, de symétrie sphérique. Il n'est pas étonnant qu'elles le soient moins à des systèmes de symétrie quelconque ou sans symétrie, dès qu'on s'éloigne du noyau. Pour comprendre les stratégies d'amélioration des bases, on peut découper l'espace en trois zones (voir figure 2)

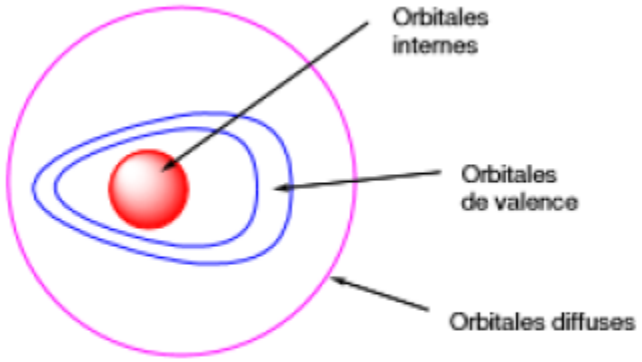


***Figure 2** : Les zones à traiter la conception d'une base*

-Les orbitales internes

Les électrons y sont proches d'un seul noyau : le potentiel nucléaire est pratiquement à symétrie sphérique. Les orbitales atomiques sont donc bien adaptées, mais l'énergie étant très sensible à la position de l'électron au proche voisinage du noyau, il sera préférable de prendre un nombre élevé de gaussiennes.

-La zone de valence

C'est la région « délicate » de la molécule, où la densité électronique est délocalisée entre plusieurs atomes, loin de la symétrie sphérique. On utilisera donc pour la décrire au mieux :

- la démultiplication de la couche de valence, ou multiple zêta de valence (ou, en anglais split valence). Par exemple, pour le carbone, une base « double zêta » utilisera deux orbitale s de valence, 2s et 2s' et six orbitales p, 2p et 2p'. Les bases usuelles de bonne qualité sont double zêta (DZ) ou triple zêta (TZ).

- l'ajout d'orbitales de polarisation. Il faut donner à la densité électronique un maximum de plasticité. Ceci se fait en ajoutant à la couche de valence des fonctions de l supérieur : orbitales p, d ...pour H, d, f, g ....pour les atomes de la deuxième période etc. En effet, au voisinage d'un atome

d'hydrogène ne possédant qu'une orbitale 1s, aucune direction de l'espace ne peut être privilégiée. Avec les orbitales p, on peut particulariser une direction, et avec un mélange sp, une direction et un sens, et ainsi de suite avec les hybrides sd, spd etc.

-La zone diffuse

Au-delà de la couche de valence, loin des noyaux, l'écart à la symétrie sphérique s'estompe à nouveau. On peut ajouter des orbitales diffuses, c'est-à-dire d'exposant a faible, qui diminuent lentement quand on s'éloigne du système. Ces OA ne sont pas indispensables dans les systèmes usuels, mais le deviennent quand on s'intéresse à des interactions à longue distance (complexes de Van der Waals), ou quand on a un anion. Dans ce cas l'électron supplémentaire tend à s'éloigner sensiblement du noyau et il faut fournir les fonctions permettant d'optimiser cette situation. La polarisation est moins importante pour ces OA, et un ensemble s et p est en général suffisant.

5.3. Nomenclature de bases usuelles

Outre la base minimale STO-3G, un jeu de bases très utilisé est symbolisé par n-n'n''... (++)G(**) n désigne le nombre de gaussiennes de la couche interne.n'n''... indiquent le nombre de gaussiennes utilisée dans chaque couche de valence. ++ (facultatif) désigne un (+) ou deux (++) ensembles de diffuses** (facultatif) désigne pour la première * des fonctions d sur les atomes de la deuxième période et des fonction p sur H. Une notation équivalente est (...)G(d,p).

Par exemple, la base très utilisée 6-31G** comporte, pour le carbone, 6 gaussiennes pour l'orbitale 1s, un double ensemble de valence, 2s 2p décrit par 3 gaussiennes et 2s' 2p' décrit par 1 gaussienne, avec des orbitales de polarisation d (p sur les hydrogènes). Ce code est reconnu par le programme GAUSSIAN.

Travaux Pratiques

Ces TPs viseront à vous faire découvrir et maîtriser certains outils informatiques couramment utilisés par le chimiste moléculaire.

TP1 : Analyse de données statistique et de graphiques avec Excel.

TP2 : Introduction à l'utilisation du logiciel d'analyse de données : « Sigma-Plot».

TP3 : Tracer un graphe à partir de données numérique avec OriginPro.

TP4 : Etude de banques de données chimiques indexées par structure: (Cambridge Structural Database).

TP5 : Outils de dessin des molécules : ChemDraw .

TP6 : Initiation à la modélisation moléculaire:

Analyse de données statistiques et de graphiques avec Excel

TP1

But

Excel offre d'innombrables possibilités de recueillir des données statistiques, de les classer, de les analyser et de les représenter graphiquement.

L'objectif de ce TP est de montrer :

1- Application et utilisation de fonctions simples et statistiques sur Excel (**Excel 2007)**.

2- Elaboration automatique des diagrammes.

1. Calcul simple et intégration de formule

1.1. Exécuter des opérations simples dans une feuille de calcul

(addition, soustraction, multiplication, division)

-Dans un tableur Excel, ouvrir une nouvelle feuille de calcul.

Entrer dans la case A1 la valeur 10, dans la case B1 la valeur 5, puis dans les cases C1 à G1 les formules **=A1+B1, =A1-B1, =A1*B1, =A1/B1**.

1. Interpréter les valeurs qui apparaissent dans les cases C1 à G1.

2. Remplacer dans la case A1 la valeur 10 par la valeur 2. Que peut-on remarquer ?

1. 2. Les fonctions statistiques :

L'Excel permet de calculer une moyenne l'écart-type, etc.

Dans cet exemple on va calculer la moyenne, l'écart-type et la médiane.

On utilise la série statistique suivante : 19.9, 19.4, 20.6, 19.8, 19.6, 19.5.

Pour calculer la moyenne et l'écart-type, il faut d'abord ordonner les données

Données ordonnées : 19.4, 19.5, 19.6, 19.8, 19.9, 20.6.

Syntaxe:

-la moyenne : **= MOYENNE(liste)**

-l'écart-type **: =ECARTYPE(liste)**

1. trouver la valeur moyenne et l'écart type de cette série

2. Essayez de trouver la médiane.

2.Diagramme à secteurs et histogramme

On a mesuré la température d'ébullition de 100 solvants. Nous obtenons le tableau statistique suivant:

Températures d'ébullitions(°C)	Nombre de solvants
[10-20[	15
[20-30[	30
[30-40[	25
[40-50[	20
[50-60[	8
[60-70[	2

-Présentez les données sous forme d'un diagramme à secteur et histogramme

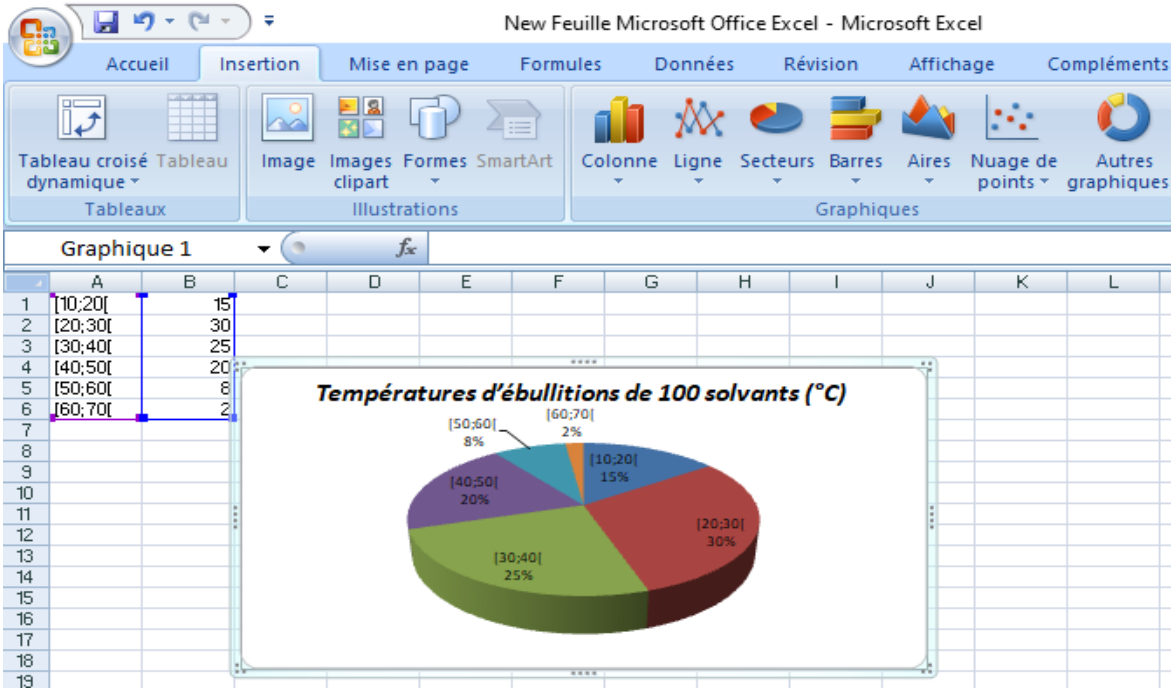
3.1. Diagramme à secteurs

1. Recopier le tableau dans Excel

2. Sélectionnez le tableau (Sélectionner la zone des cellules), A partir de l'onglet Insertion, cliquez sur cliquez sur «Graphiques ».

3. Maintenez le bouton (voir image) pour avoir un premier aperçu du sous-type de graphique sélectionné.

4. Choisir le type de graphique « **Secteurs** ». Le diagramme à secteur s'affiche dans la fenêtre suivante:



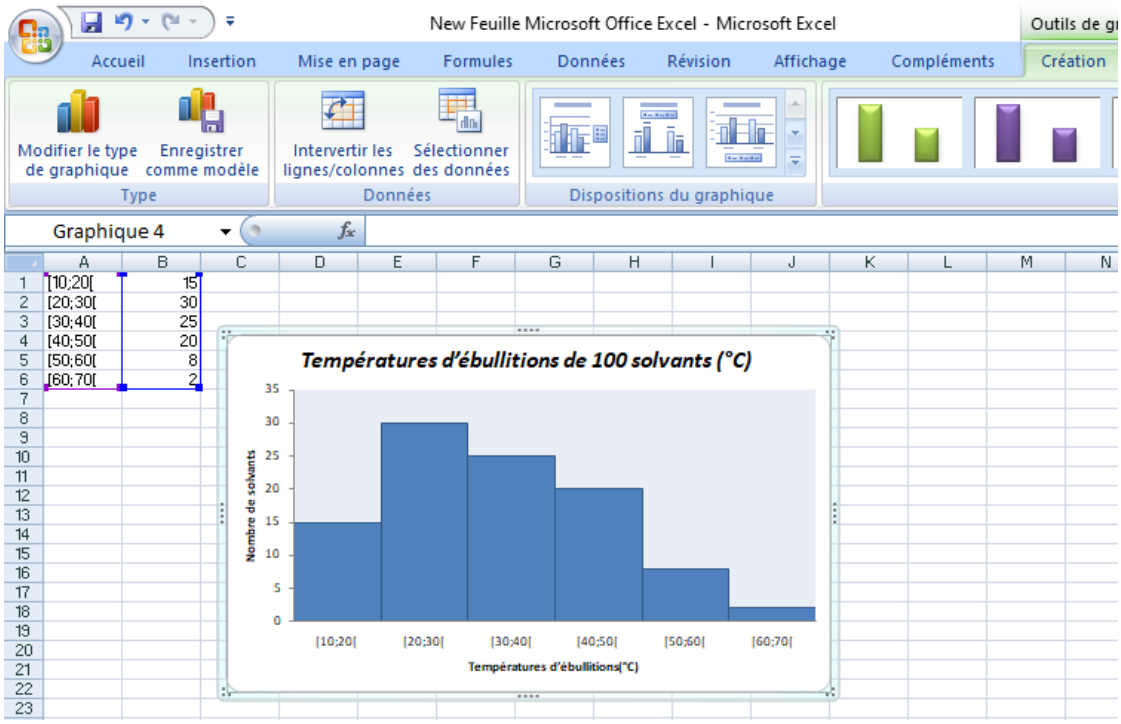
3. 2. Construire l'histogramme des effectifs

1. Recopier le tableau dans Excel

2. Sélectionnez le tableau (Sélectionner la zone des cellules), A partir de l'onglet Insertion, cliquez sur cliquez sur «Graphiques ».

3. Maintenez le bouton (voir image) pour avoir un premier aperçu du sous-type de graphique sélectionné.

4. Choisir le type de graphique « **Colonne**». L' histogramme s'affiche dans la fenêtre suivante :



Annexe

EXCEL est un tableur qui va vous permettre de faire des tableaux avec des calculs automatisés, des graphiques qui les illustrent et du texte qui les commente. Un tableur se présente sous la forme de classeurs en deux dimensions : colonnes et lignes.

-Un classeur est un ensemble de feuilles de calcul, c'est un ensemble de feuilles de calcul stockées dans un même fichier. Chaque feuille est repérable par un onglet à son nom.

-Une feuille de calcul est un ensemble de cellules organisées en tableau.

-Une cellule est l'intersection d'une ligne et d'une colonne. Une cellule active est une cellule qui apparaît en surbrillance à l'écran.

-Le ruban est constitué des composants suivants :

-Des onglets pour chaque catégorie des tâches d'Excel rassemblant les commandes les plus utilisées.

-Des groupes rassemblant des boutons de commande.

-Des boutons de commande dans chaque groupe que vous pouvez sélectionner pour accomplir une action.

-Des lanceurs de boîte de dialogue dans el coin inférieur droit de certains groupes vous permettant d'ouvrir une boîte de dialogue contenant un certain nombre d'options supplémentaires.

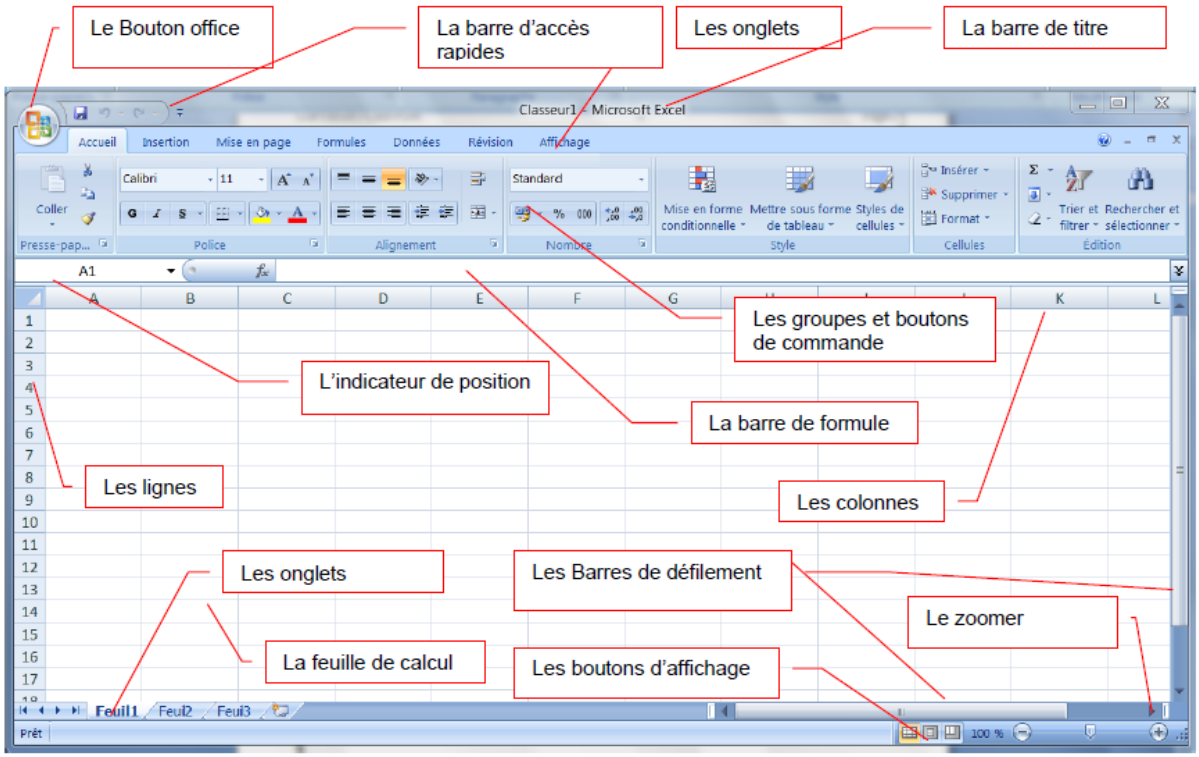


Figure 1 : *Page d'accueil et interface générale.*

1. Les fonctions simples

Le tableur étant un logiciel basé sur des tableaux de chiffres, il permet de réaliser des calculs. Une formule est un ensemble de données saisies dans une cellule. Elle sert à effectuer un calcul ou une analyse des données dans la feuille de calcul.

Remarque : une formule de calcul dans Excel commence toujours par le signe = (égal).

Vous pouvez ensuite effectuer toutes les opérations courantes en combinant les noms de cellules (A1, B3, C4,...)

+	addition
-	soustraction
*	multiplication
/	division

-Pour créer un calcul, il vous faut utiliser les opérateurs suivants : +,-,*,/

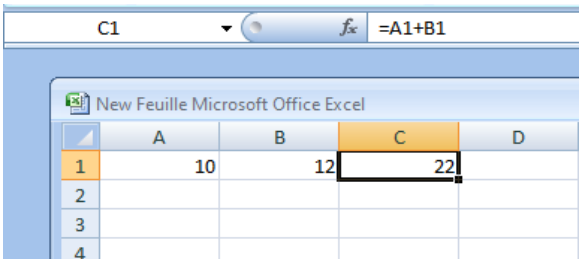
Excel est capable d'utiliser des références dans les calcules. Une référence est le nom de la cellule, à savoir la lettre de la colonne plus le chiffre de la ligne et se trouve à gauche de la barre de formule.

La référence est la cellule dans laquelle doit s'afficher le résultat, dans notre exemple la C1.

Exemple :

Pour l'addition, tapez la formule suivante :

= SOMME(A1+B1)



-Validez par la touche Entrée ou par le bouton Valider.

2. Les fonctions statistiques

Par Exemple:

Pour obtenir la moyenne, utiliser la fonction MOYENNE.

Ecrire dans la cellule où doit apparaître le résultat : **=MOYENNE(liste)**

-Sélectionner ensuite à la souris la plage de cellules désirée

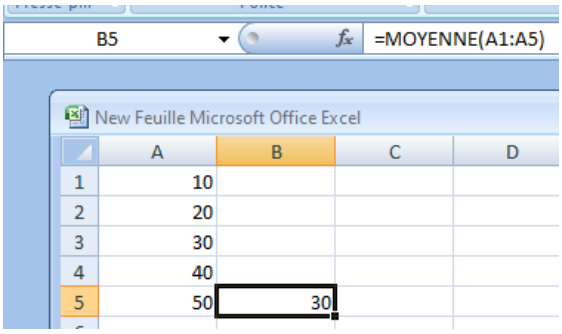
-Valider avec la touche Entrée.

-La fonction renvoie la moyenne des sommes sélectionnées.

Exemple : Dans la colonne A on a écrit des nombres.

On voudrait obtenir la moyenne des nombres de la colonne A.

En B5 on écrit la formule suivante : =MOYENNE(A1:A5)

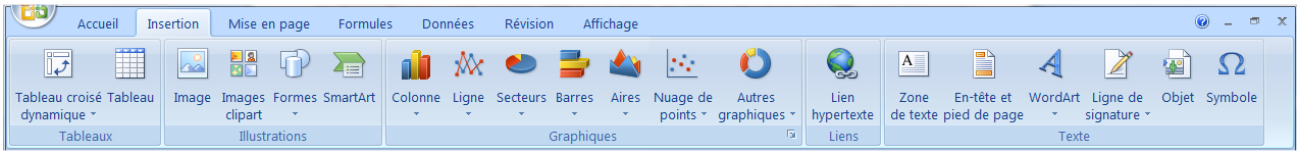


-Validez par la touche Entrée ou par le bouton Valider.

3. Les graphiques

Les graphiques sont utilisés pour afficher des séries de données numériques sous forme graphique afin d'appréhender plus facilement d'importantes quantités de données et les relations entre différentes séries de données.

Pour créer un graphique dans Excel, il faut commencer par entrer les données dans une feuille de calcul et ensuite tracer ces données dans un graphique : plusieurs formes de graphiques sont disponibles dans le groupe Graphiques de l'onglet Insertion.



1. Créez votre base de données.

2. Sélectionnez les cellules contenant les données que vous voulez utiliser pour votre graphique.

3. Cliquez sur l'onglet Insertion et choisissez votre graphique

-Pour notre exemple : choisissez le graphique → diagramme à secteur et histogramme.

Introduction à l'utilisation du logiciel d'analyse de données : « Sigma-Plot»

TP2

But

SigmaPlot est un logiciel d'analyse de données pratique et le plus largement utilisé qui vous permettra de créer une feuille de calcul simple ainsi qu'un graphique scientifique de haute qualité.

L'objectif de ce TP est de montrer :

1. Application et utilisation de fonctions simples sur SigmaPlot **(SigmaPlot 14.0)**

1. Utilisation du logiciel SigmaPlot

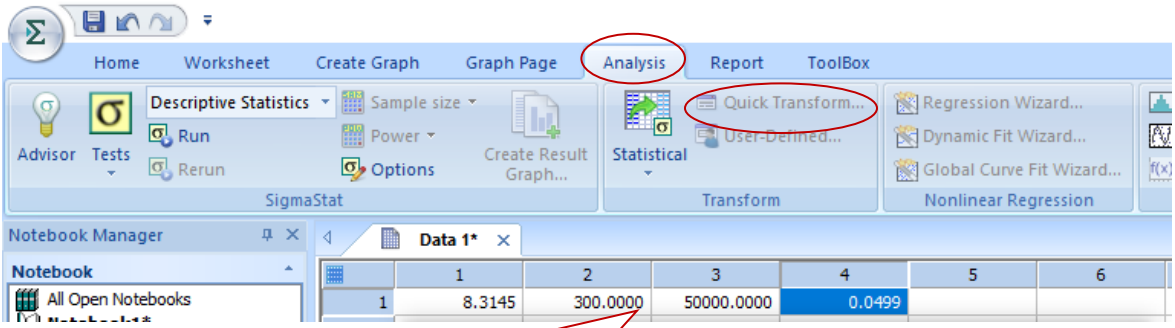
1.1. Utilisation de fonctions simples sur SigmaPlot

-Pour une mole de gaz l'équation des gaz s'écrit : **PV = RT** (R=8.31451 $Pa.m^3.mol^{-1}.K^{-1}$)

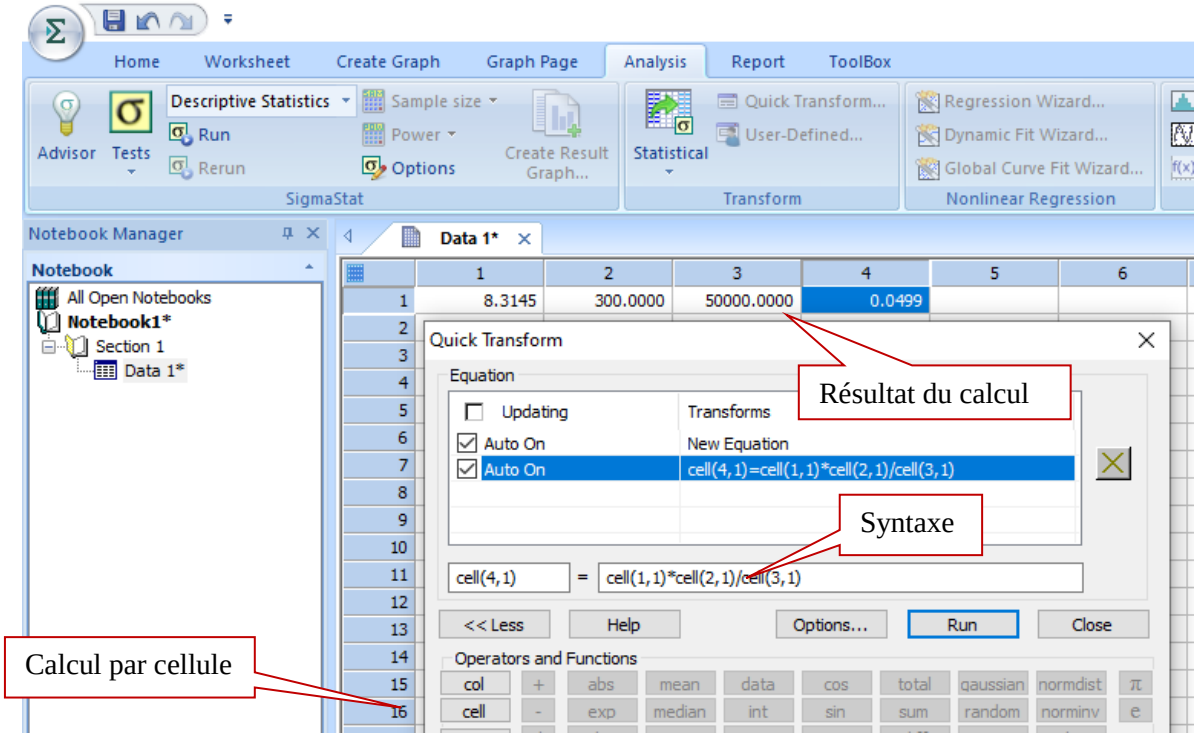
Nous allons utiliser **SigmaPlot** pour calculer le volume de gaz à Pression P= **50000** Pa pour une température donnée **300** K.

-Calcul par cellule de Sigma Plot

1. Après avoir double cliqué sur l'icône **Sigma Plot**, la fenêtre d'application s'affiche « **Worksheet** »
2. Placez le curseur à la cellule 1.1 de Sigma Plot, entrez 8.31451
3. Déplacez le curseur en cellule 2.1, Entrez 300
4. Déplacez le curseur en cellule 2.1, Entrez 50000
5. On cliqué **Analysis**, on choisit en suite **Quick transform**
6. On indique les cellules de données. On désigne ensuite la formule « **cell(1,1)*cell(2,1)/cell(3,1) »** qui permet de calculer le volume
7. On cliqué **Run**
8. Le résultat du calcul s'affiche dans la cellule 4,1 **(cell (4,1))** (voir figure ci-dessous).

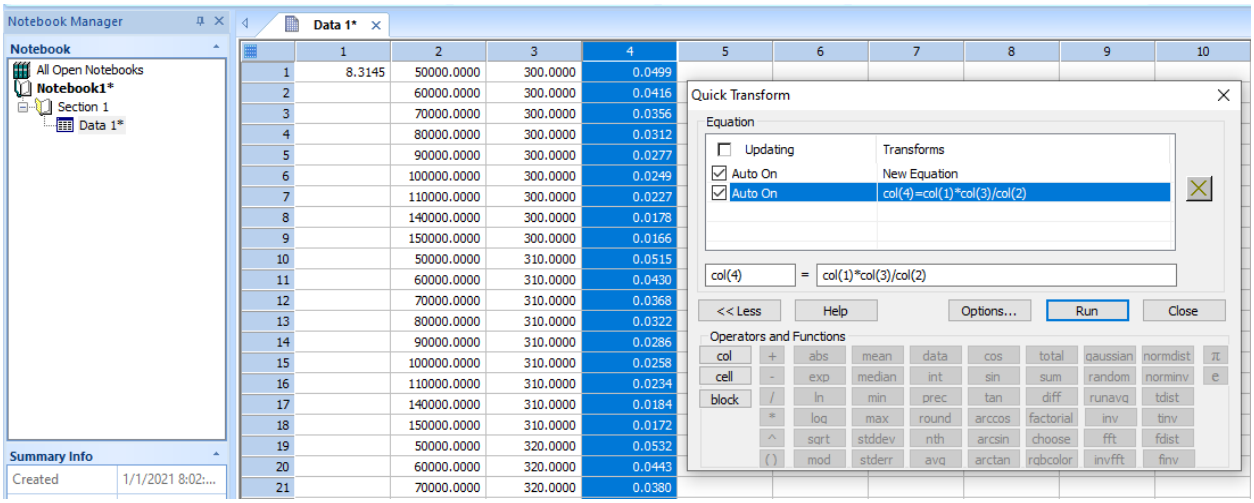


Les cellules 1.1, 2.1, 3.1 et 4.1

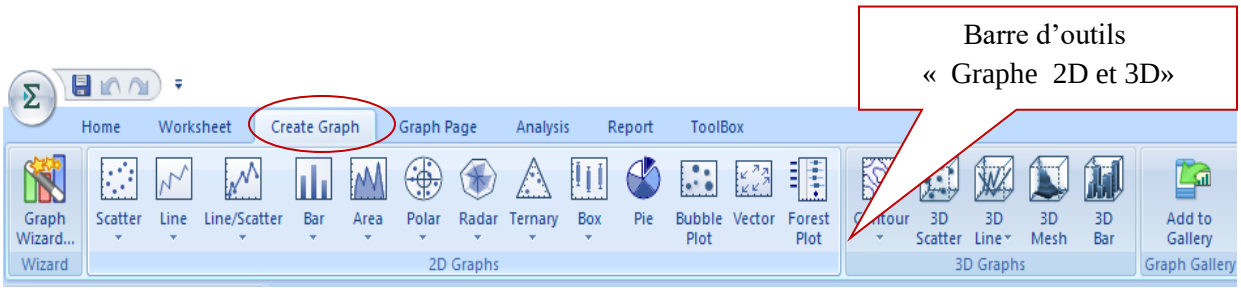


1.2. Graphique 3D en utilisant Sigma Plot

1. Préparer un tableau des valeurs de volumes dans la colonne 4, V=RT/P (**Calcul par colonne de Sigma Plot)**.



2. on clique **Create Graph** dans le menu **Graph** qui ouvre la boite de dialogue, on sélectionne le type de courbe, dans ce cas on cliqué **3D Mesh Plot**; (voir figure ci-dessous).



Après **Next,** On indique le format de données parmi **XYZ Triplet**→Next

X : Column 2 (Température T)

Y : Column 1 (Pression P)

Z : Column 3 (Volume V)

3. On clique finalement sur la case **Finish** et la page graphique s'affiche dans la fenêtre Sigma Plot

4.Cliquer sur la souris à droite→**Graph Properites→Graph→Rotation.** Jouer avec horizontal et vertical view

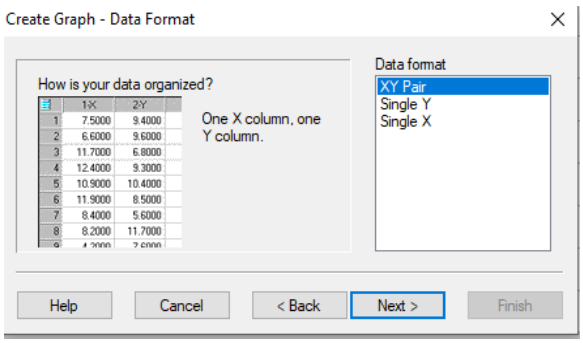
5. Ajoutez les titrez aux colonnes

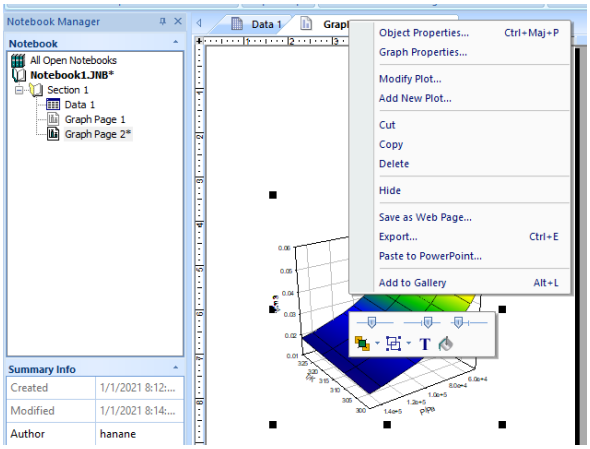
Cliquer Y data, changer le titre en T/K.

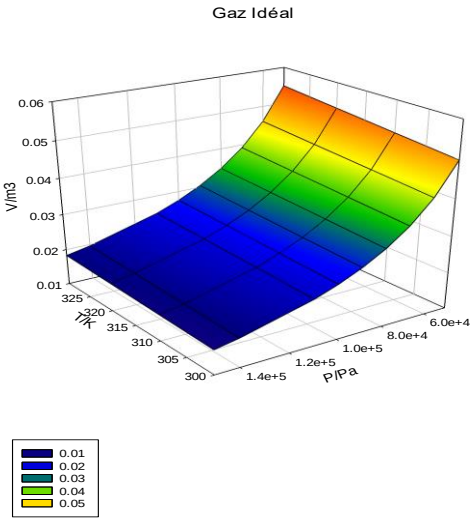
Cliquer X data, changer le titre en P/Pa.

Cliquer Z data, changer le titre en V/m3.

Cliquer 3D Graph, changer le titre en Gaz Idéal.







Graphe crée sous Sigma Plot et importé dans Word

Annexe

SigmaPlot est un logiciel graphique scientifique désigné pour fonctionner sur la plate-forme Windows. Il est spécialement conçu pour la présentation graphique de données expérimentales. La création d'un graphe est assez simple et on peut utiliser la feuille de style pour retracer le même type de graphe. C'est un logiciel exhaustif qui offre:

-Un large choix de modèles graphiques.

-Analysez vos données en toute simplicité.

-Ajustez facilement vos données.

-Publiez vos graphiques où vous voulez.

-Utilisez Sigmaplot depuis Excel.

Pour obtenir plus d'information, veuillez consulter le site Web:

https://support.alfasoft.com/hc/en-us/articles/360001311138-SigmaPlot-Tech-Tips-and-Tricks

Présentation du logiciel

La fenêtre de SigmaPlot est illustrée dans la Figure 1, Comme la plupart des applications Windows, en retrouve la barre de menus accessible à l'aide de la souris ou le raccourci de touches et la barre d'outils. Les différentes commandes de la barre d'outils graphiques sont résumées sur la Figure 2.

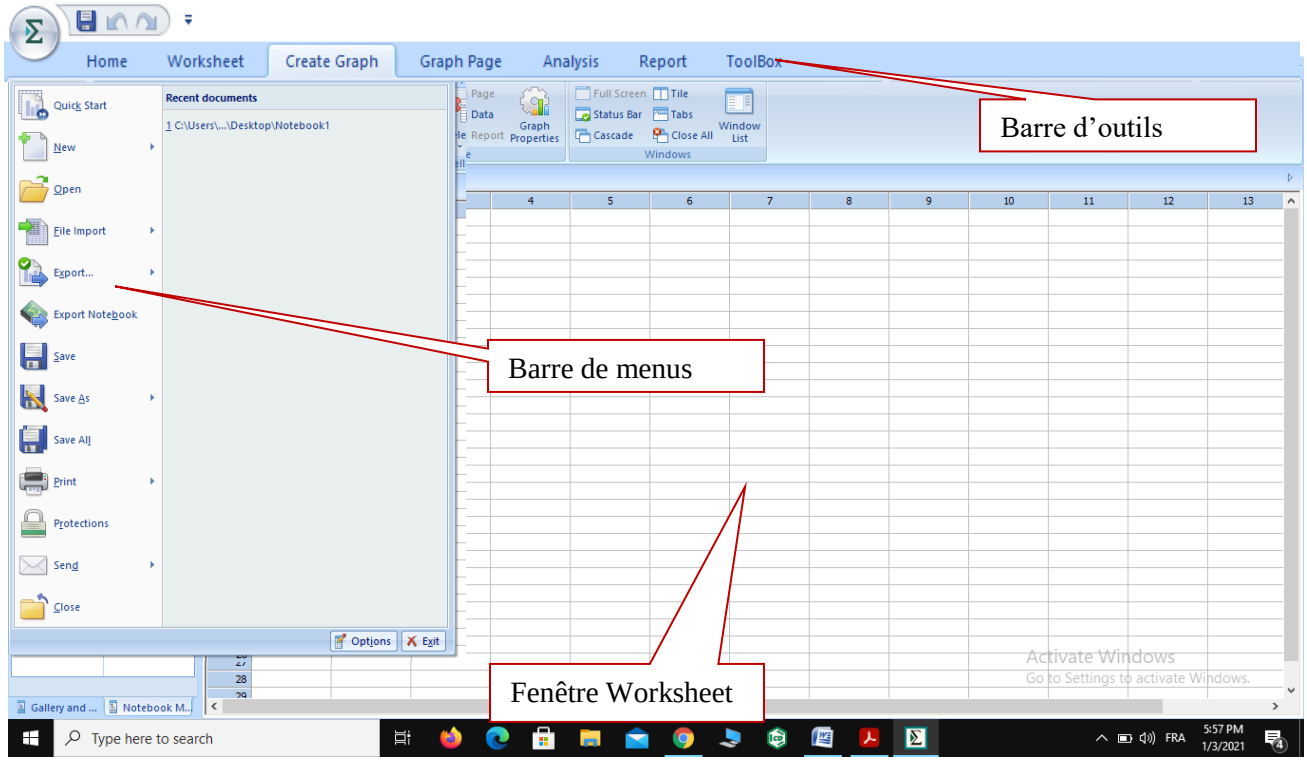


Figure 1 : *Page d'accueil et interface générale.*

Tracer un graphe à partir de données numérique avec OriginPro

TP3

But

Origin est un logiciel de traitement et d'analyse de données scientifiques pour environnement Microsoft Windows développé par OriginLab. Il permet notamment de tracer des courbes, des graphes 2D et 3D et possède des fonctions d'analyse et d'interpolation. Il est aussi capable d'importer des fichiers de divers formats tels que Excel, ASCII, Mathematica ou SigmaPlot, et d'exporter les graphes sous format JPEG, GIF, Tiff etc.

L'objectif de ce TP est de montrer :

-Application et utilisation du logiciel d'analyse de données en chimie « OriginPro 8.5».

-Importer des fichiers de formats Excel et SigmaPlot.

Tracer un graphe à partir de données numérique

1. Excel (Excel 2007)

Exploiter des résultats d'expérience à l'aide d'un tableur (Excel).

1. Créer une nouvelle feuille de calcul.

Vb (mL)	0	4	8	10	12	14	14,5	15	15,5	16	18	20	24	26	30
pH	2.9	3.6	3.9	4.1	4.5	5	5.8	7.6	10	11	11.3	11.4	11.5	11.6	11.7

2. Recopier ce tableau sur une feuille d'un classeur Excel.

3. Tracer du graphe pH = f (Vb)

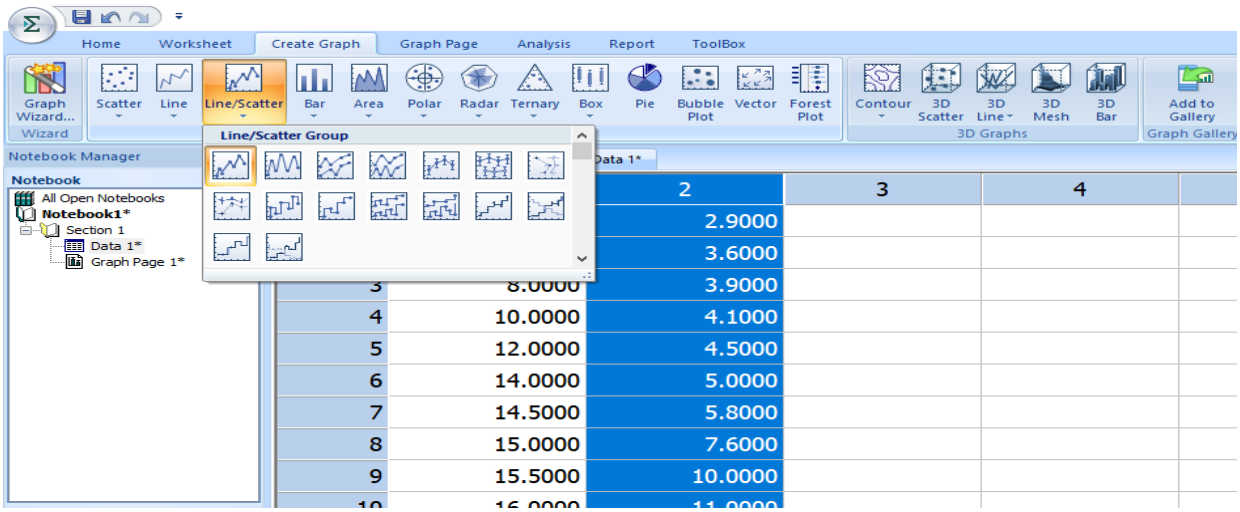
2. Sigma Plot

Exploiter des résultats d'expérience précédente à l'aide « **Worksheet** » de **Sigma Plot**

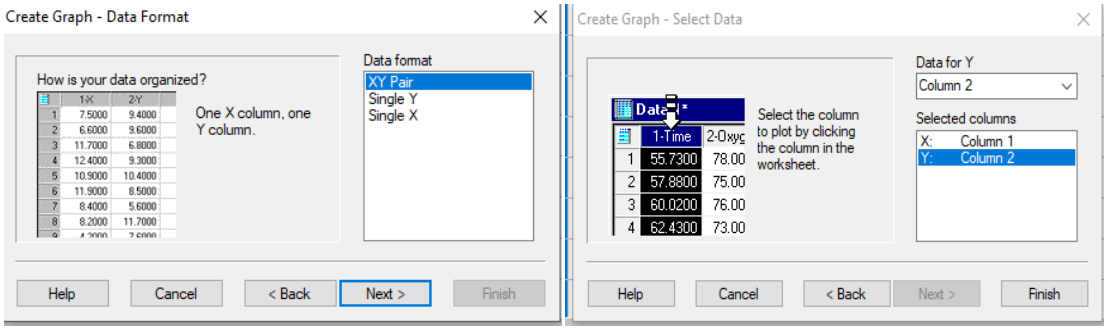
1. Après avoir double cliqué sur l'icône **Sigma Plot**, la fenêtre d'application s'affiche comme montrer dans la figure ci-dessous.

2. Recopier le tableau dans la cellule de Sigma Plot.

3. on clique **Create Graph** dans le menu **Graph** qui ouvre la boite de dialogue, on sélectionne le type de courbe, dans ce cas on cliqué **Line/Scatter**; on choisit en suite le style de courbe **Simple Straiht Lines & scatter** (voir figure ci-dessous).



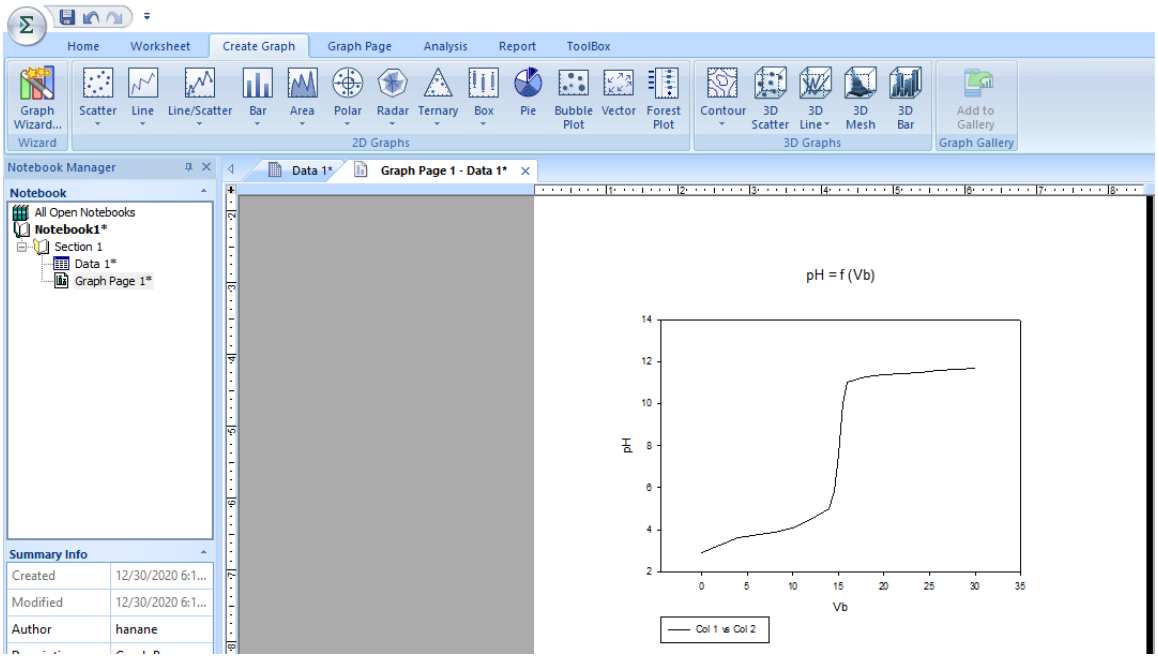
4. On indique le format de données parmi **XY Pairs**. On désigne ensuite la colonne 1 comme X et la colonne 2 comme Y (voir figure ci-dessous).

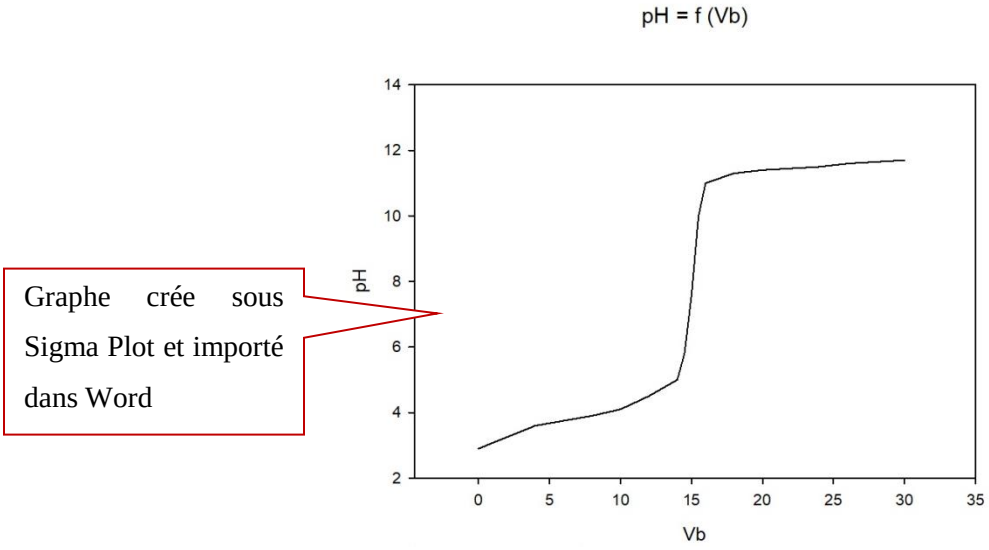


Format de données. **Affectation de colonnes aux axes X, Y.**

5. On clique finalement sur la case **Finish** et la page graphique s'affiche dans la fenêtre Sigma Plot

6. Insérer le graphique dans votre rapport.





Graphe crée sous Sigma Plot et importé dans Word

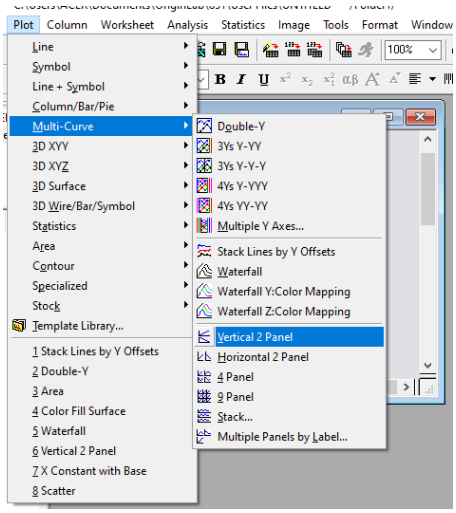
2. Origin pro 8.5

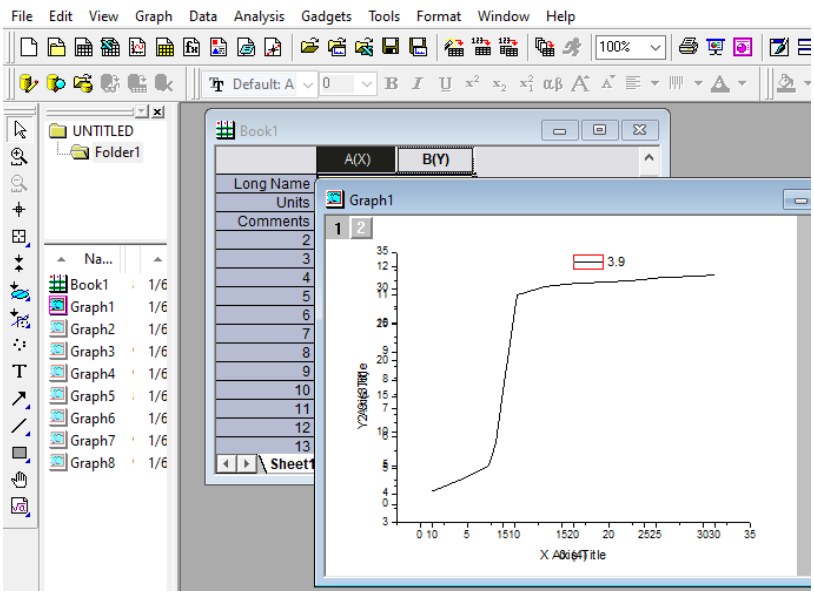
1. Après introduction des valeurs (Origin considère par défaut la première colonne comme axe des abscisses et la seconde colonne comme axe des ordonnées.

- Pour le plot (tracer la courbe)

1.Sélectionner les colonnes B et C, et sélectionner **Plot**→Multi-Curve → Vertical 2 Panel pour dessiner la courbe.

2. Pour réarranger ces deux couches, utiliser: **Layer Management** la page graphique s'affiche dans la fenêtre (voir figure ci-dessous).

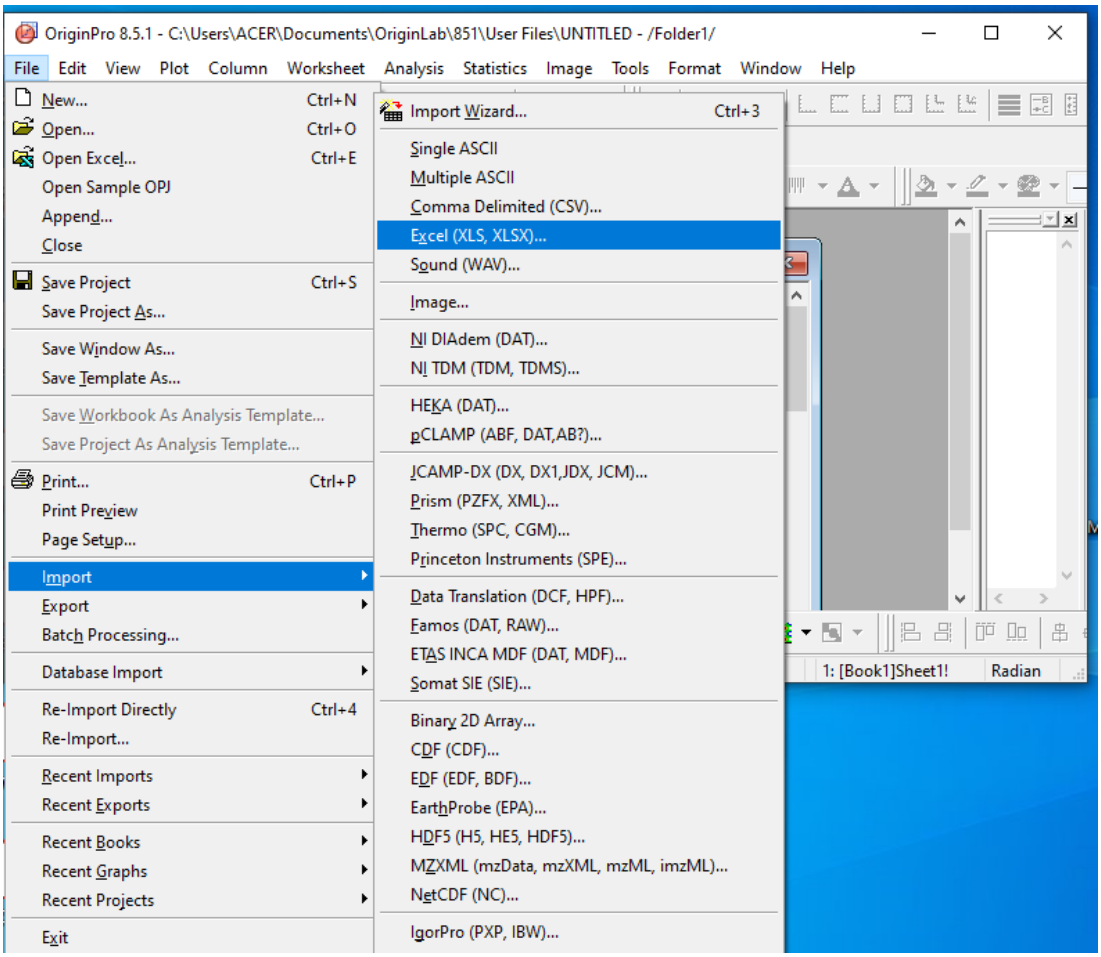




2. Importer des données dans Origin pro 8.5 (Importer des fichiers de formats Excel et SigmaPlot)

L'importation d'un fichier se fait de la façon suivante:

1. Accédez à **File**, cliquez sur « Importer » puis Excel

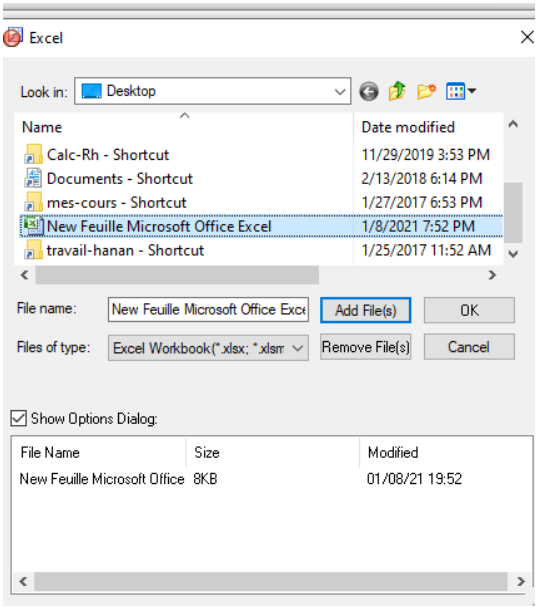


2. Recherchez et double-cliquez sur le fichier Excel que vous voulez ouvrir.

Enfin, cliquez sur **« OK »** pour terminer l'opération d'importation.

Origin ouvre automatiquement le fichier et affiche les données dans un nouveau tableau dans lequel on introduit les valeurs

.



Annexe

1. Comment réaliser un graphique avec Excel (EXCEL 2007)

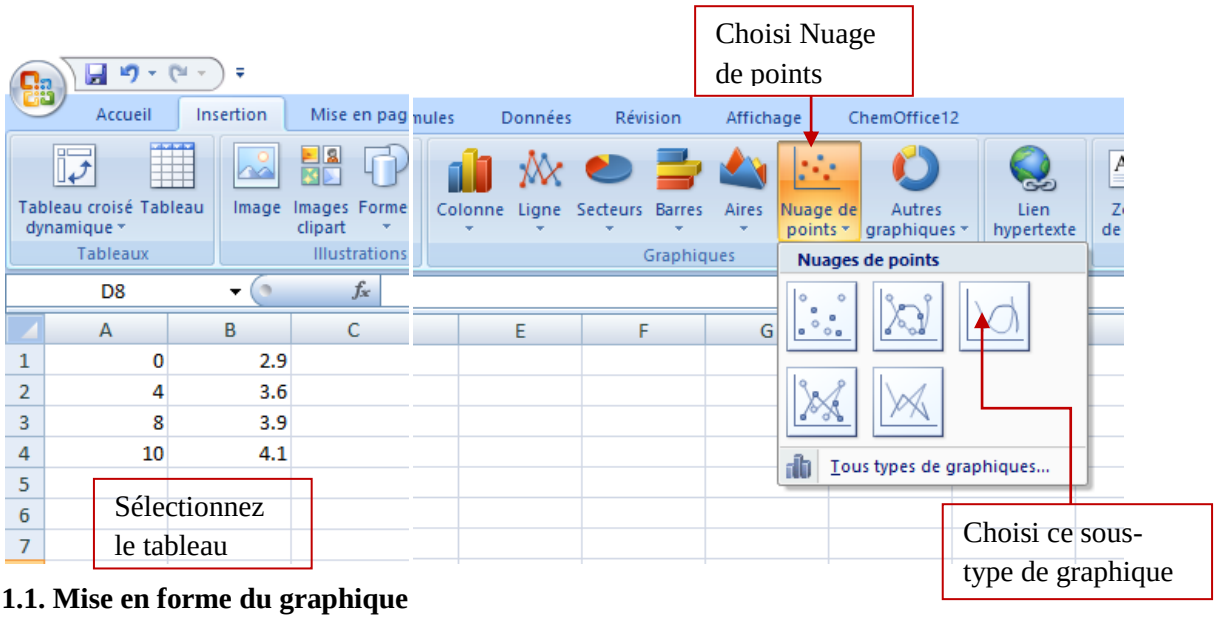
Excel est un logiciel dit « tableur » (fichier .xls ou « classeur »)

A chaque démarrage d'Excel, un classeur vierge s'ouvre avec 3 feuilles. Vous pouvez passez d'une feuille à l'autre en cliquant simplement sur le nom de la feuille.

1. Créez le tableau

2. Sélectionnez le tableau (Sélectionner la zone des cellules), A partir de l'onglet Insertion, cliquez sur «Graphiques ».

3. Maintenez le bouton (voir image) pour avoir un premier aperçu du sous-type de graphique sélectionné.



1.1. Mise en forme du graphique

-Titrer le graphique

1. Sélectionner l'onglet Disposition ;
2. Dans le sous-menu Étiquettes, sélectionner Titre de graphique et Au-dessus du graphique ;
3. Écrire le titre en Times New Roman, 12 points ;
4. Si le titre n'est pas centré, sélectionner Disposition, Options, Alignement, Centré ;
5. Supprimer la fenêtre affichant série 1 et série 2 ;

-Titrer les axes

1. Sélectionner l'onglet Disposition ;
2. Dans le sous-menu Étiquettes, sélectionner Titres des axes et Titre de l'axe horizontal principal puis Titre en dessous de l'axe ;
3. Écrire le titre de votre axe.
4. Faites la même procédure pour l'axe y. Sélectionner Titre de l'axe vertical principal puis Titre→pivoté.

1.2. Insérer le graphique dans votre rapport

1. Enregistrer le travail.

2. Copier et coller le graphique dans Word.

3. Si le graphique est en format paysage, il doit être inséré sur une page aussi en orientation paysage.

Dans Mise en page sélectionner: Orientation paysage.

2. L'Origin

c'est un outil complet d'analyse de données et de mise en forme graphique fournissant tout un éventail de fonctionnalités (analyse de pics, ajustement de courbes, statistiques...) permettant de satisfaire aux exigences de qualité et aux besoins spécifiques de la communauté scientifique (chercheurs,étudiants,ingénieurs, techniciens...).

Pour obtenir plus d'information, veuillez consulter le site Web: https://www.originlab.com/

Les bonnes raisons d'utiliser ce logiciel

-Import les données brutes de vos expérimentations avec les nombreux formats disponibles : .csv, .dat, .xls...

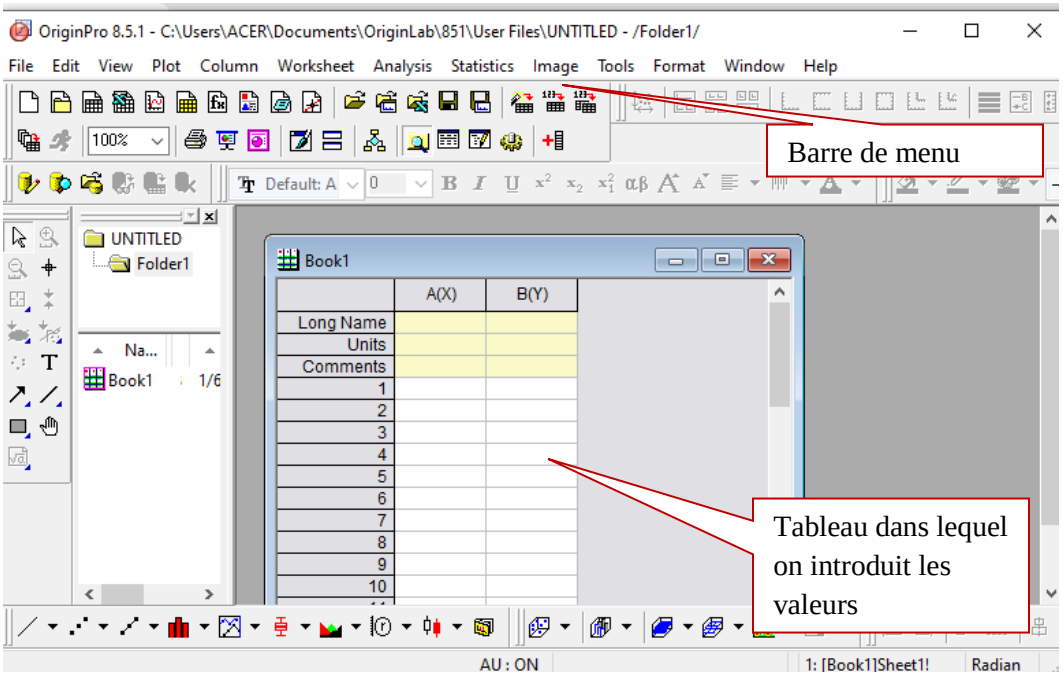
-Souplesse pour l'analyse de données (ajustement de courbes et de surface, analyse de pics, interpolation, traitement du signal, statistiques, ANOVA, tests paramétriques et non-paramétriques, analyse multivariée, calculs de puissance et d'effectifs d'échantillons, analyse de survie et courbes ROC. Chaqueanalyse est accompagnée d'un rapport quirésume les paramètres et les résultats).

-Large choix de graphiques (2D et3D).

-Automatisation des tâches répétitives.

-Connectivité avec Excel, Matlab, LabView, Minitab, Python, ...

2.2. L'interface utilisateur graphique



Sur le tableau on fait introduire les valeurs des mesures. Le nom du tableau est nommé par défaut Data1.

- On peut ajouter un autre tableau en cliquant juste sur l'icône qui se trouve dans la barre des icônes.

Etude de banques de données chimiques indexées par structure (Cambridge Structural Database)

TP4

But

-Recherche de molécules dans une banque de données

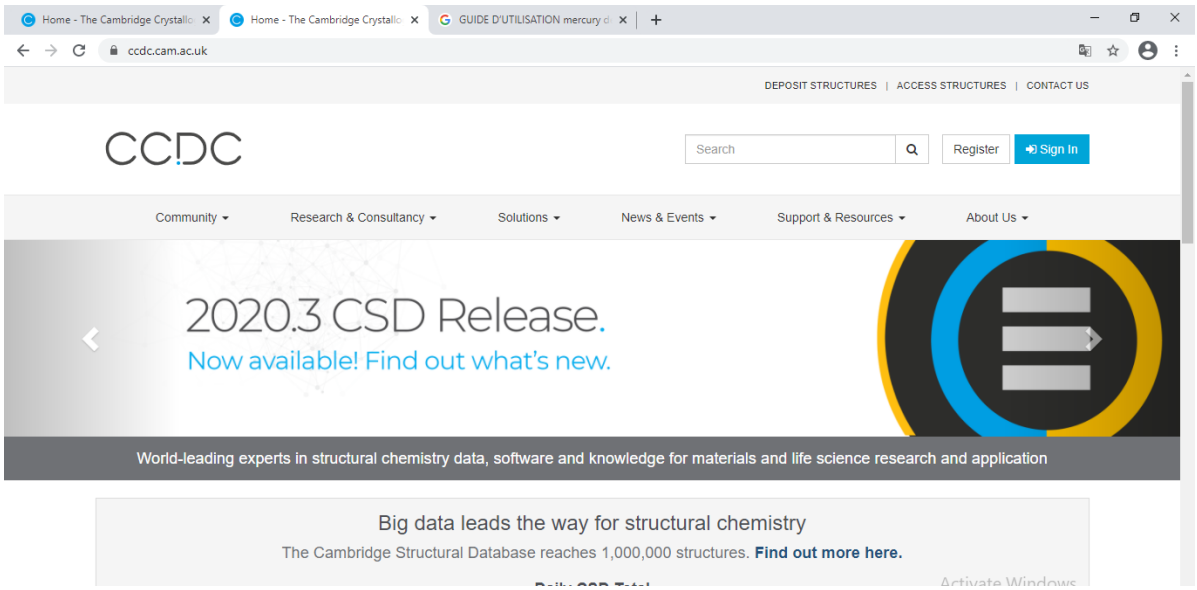
-Recherche structurale en utilisant la base de Cambridge

1. Accès à la Cambridge Structural Database

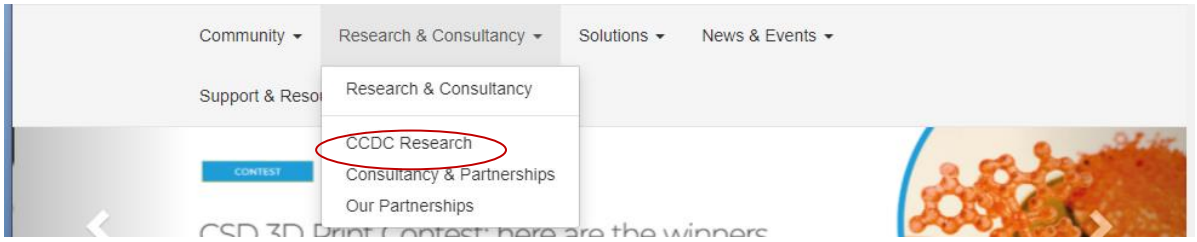
Aller sur la banque « The Cambridge Crystallographic Data Centre (CCDC) » https://www.ccdc.cam.ac.uk/

Étapes de conversion d'une structure cristallographique en format cif

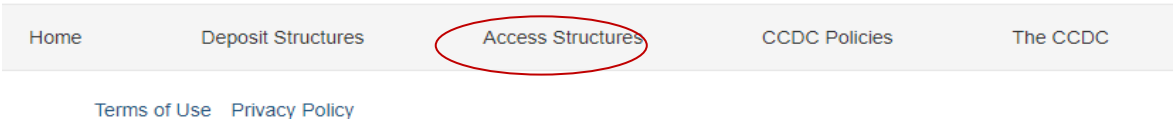
1. La page d'accueil du CSD se présente comme suit :



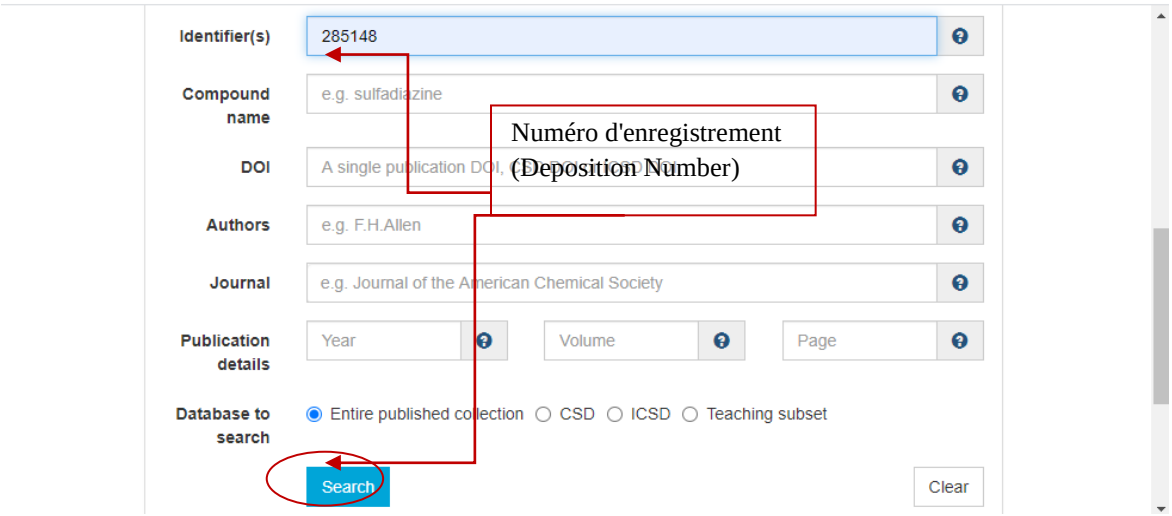
2.Veuillez cliquer sur l'hyperlien "**CCDC Research** "



3. Cliquez sur "**Access Structures** " dans la page suivante

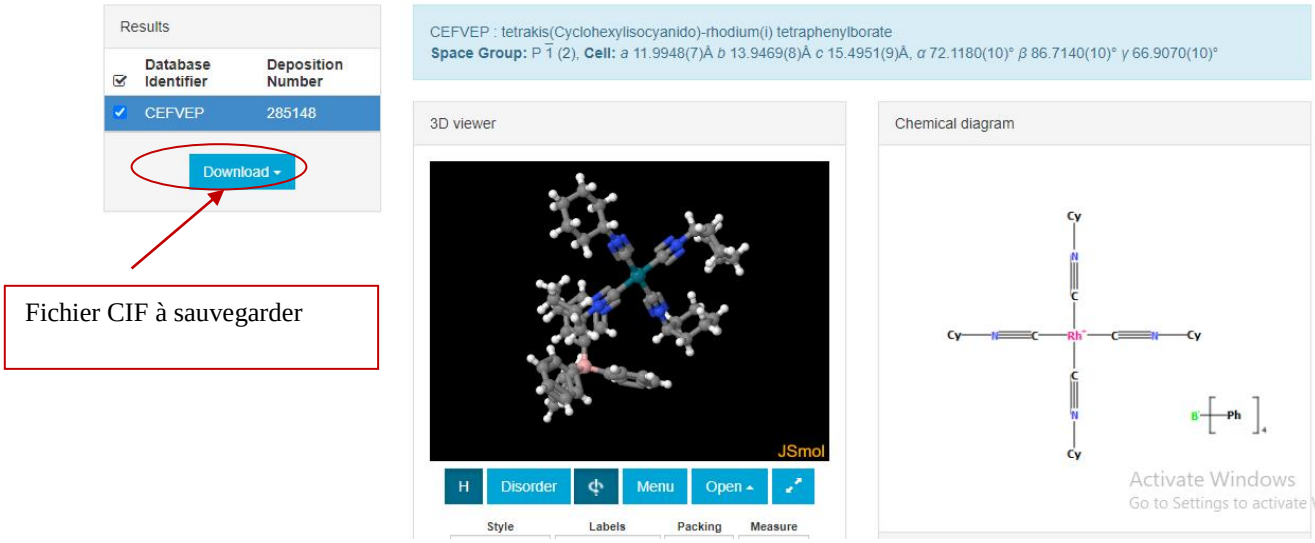


4. Vous arriverez alors sur la page suivante :



Additional details	
Deposition Number	285148
Data Citation	N.T.Tran, J.R.Stork, D.Pham, M.M.Olmstead, J.C.Fettinger, A.L.Balch CCDC 285148: Experimental Crystal Structure Determination, 2006, DOI: 10.5517/cc9kqbw
Deposited on	28/09/2005

5. Cliquez ensuite sur le bouton « **Search**» ; Vous arriverez alors sur la page suivante :

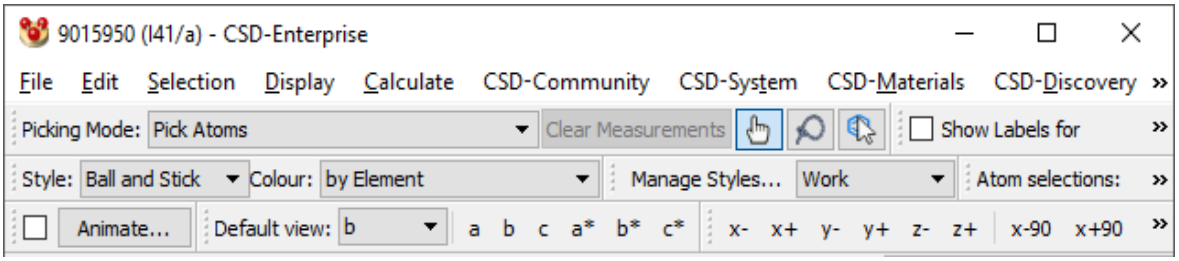


6. Nous avons téléchargé le Cif qui porte le numéro : 285148 (fichier cif publier par Cambridge Crystallographic Data Centre obtenu par la diffraction RX lors de la synthèse de la molécule).

2. visualisation avec Mercury

Mercury (Hg) peut lire directement des fichiers CIF (Crystallographic Information Files) ou vous permettre de construire votre propre structure cristalline en entrant à la main, les paramètres cristallins, le groupe d'espace et les coordonnées réduites des atomes dans la maille cristalline.

Ce logiciel gratuit a été à l'origine développé par CCDC.

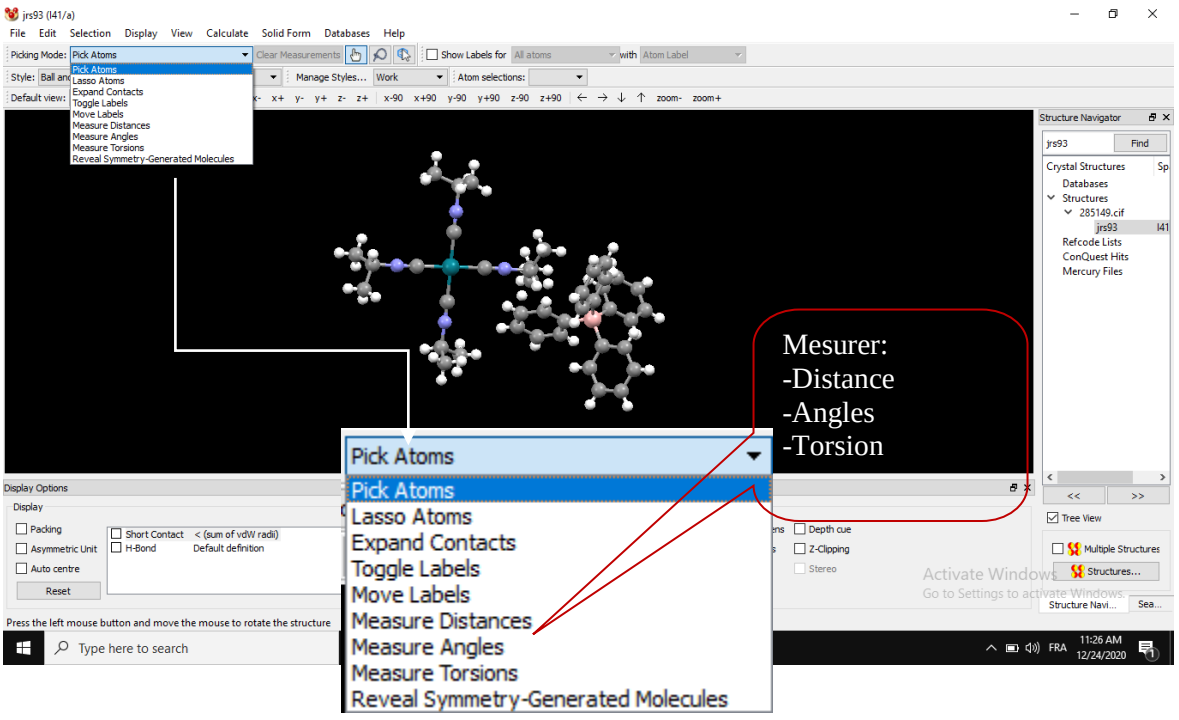


1. Recopier le fichier de données .cif dans un répertoire de votre choix

2. Démarrer le programme **Mercury (Hg)**.

3. Cliquez sur le bouton File → Cliquez sur « Open … ». Choisissez le fichier de données, ici,«.cif ».

4. Vous arriverez alors sur la page suivante :



Application : Recherche Simple

Rendez vous sur la banque de données « The Cambridge Crystallographic Data Centre (CCDC) » https://www.ccdc.cam.ac.uk/, et récupérez les structures cristallographiques possédant les numéros suivant: **285100** et **285101**.

1.Retrouvez sur cette banque :

-Les noms de ces composés

- les informations cristallographies

2. Quelles sont les différences principales entre les deux structures?

3.Télécharger ces structures moléculaires.

-Quel sera le contenu des fichiers CIF ?

-Comment est –il possible de visualiser ces molécules sur le Pc ? Citer les étapes

3. Effectuer la recherche bibliographique des applications de méthodes spectroscopiques (RMN, IR,UV, …) sur ces complexes.

Annexe

1. Base de données structurales

Les bases de données structurales contiennent les informations sur les structures de composés organiques, organométalliques et inorganiques mais aussi sur les métaux et les alliages.

Leur rôle est multiple:

-vérifier que la structure cristalline en cours n'a pas déjà été étudiée

-récupérer une ou plusieurs structures pour des études de modélisation

-rechercher des structures possédant des fragments structuraux définis

-explorer les interactions moléculaires d'une série de composés

-visualiser les structures en 3D

2. Principales Bases de Données Structurales

-CSD (Cambridge Structural Database): C'est la seule base regroupant les données structurales complètes de composés organiques et organométalliques.

-ICSD (Inorganic Crystal Structure DataFile): C'est l'équivalent de la CSD pour l'ensemble des structures inorganiques.

-CRYSTMET (MDF – Metals Data File): Elle contient les données structurales pour les métaux et les alliages.

-CDIF (Crystal Data Identification File): Elle fournit les symétries, les paramètres de maille, la composition et les références bibliographiques pour plus de 237000 composés.

-ICDD (International Centre for Diffraction Data): Elle contient l'ensemble des données structurales sur poudre.

3. Cambridge Structural Database (CSD)

Dans ce TP nous utiliserons la base de données structurelle de Cambridge (CSD) :

La base de données structurelle de Cambridge (CSD) est à la fois un dépôt et une validation et organisée pour trouver les données de structure en trois dimensions de molécules contenant généralement au moins du carbone et de l' hydrogène , comprenant une large gamme d' organique, organométallique et organométalliques molécules.

Les entrées spécifiques sont complémentaires aux autres bases de données cristallographiques telles que la banque de données sur les protéines (PDB), la base de données sur les structures cristallines inorganiques et le Centre international de données de diffraction.

Les données, généralement obtenues par cristallographie aux rayons X et moins fréquemment par diffraction électronique ou par diffraction neutronique, et soumises par des cristallographes et des chimistes du monde entier, sont librement accessibles (telles que déposées par les auteurs) sur Internet via le site Web de l'organisation mère du CSD (CCDC, référentiel).

Le CSD est supervisé par la société à but non lucratif constituée en personne morale appelée Cambridge Crystallographic Data Center , CCDC. L'intérieur du siège du CCDC Cambridge, Royaume-Uni Le CSD est un référentiel largement utilisé pour les structures cristallines organiques et métallo-organiques à petites molécules pour les scientifiques. Les structures déposées auprès du Cambridge Crystallographic Data Center (CCDC) sont accessibles au public pour téléchargement au moment de la publication ou avec le consentement du déposant. Ils sont également enrichis scientifiquement et inclus dans la base de données utilisée par les logiciels proposés par le centre. Des sous-ensembles ciblés de la CDD sont également disponibles gratuitement pour soutenir l'enseignement et d'autres activités.

Cambridge Crystallographic Data Centre CCDC: crée en 1965 par le Dr Olga Kennard du Department of Organic, Inorganic and Theoretical Chemistry of the University of Cambridge.

CCDC web site : http://www.ccdc.cam.ac.uk

Elle contient l'ensemble des structures de petites molécules organiques et Organométalliques. Toutes ces structures ayant été analysées soit par diffraction des rayons X, soit par diffraction des neutrons. La base contient plus de 322.000 composés.

Des logiciels pour la recherche, la récupération, l'examen et l'analyse des informations fournis par la CSD sont à la disposition des utilisateurs. Ces logiciels sont utilisés dans le monde entier, aussi bien par le monde académique que par la recherche industrielle et sont développés en permanence par le personnel du Cambridge Crystallographic Data Centre(CCDC) à Cambridge.

- Formats de fichiers

De nombreux formats de fichiers sont utilisés en CSD on peut citer : Format CIF (Crystallographic Information File).

Crystallographic Information File (CIF) est un format de fichier texte standard pour échanger des informations sur la structure des cristaux On y retrouve donc principalement les informations suivantes :

-Formule chimique (brute et/ou structurale).

-Paramètres de la maille les longueurs et les angles de liaison entre les atomes.

4. Visualiser les données

Chaque ensemble de données dans CSD peut être ouvertement visualisé et récupéré à l'aide du service gratuit **Access Structure**. Grâce à ce service basé sur un navigateur Web, les utilisateurs peuvent visualiser

l'ensemble de données en 2D et 3D, obtenir des informations de base sur la structure et télécharger l'ensemble de données déposé. Des fonctions de recherche plus avancées et des informations organisées sont disponibles via le système CSD basé sur l'abonnement.

Outre l'utilisation du système CSD , les fichiers de structure peuvent être visualisés à l'aide de l'un des nombreux programmes informatiques open source tels que Jmol . Certains autres programmes gratuits, mais non open source, incluent MDL Chime , Pymol , UCSF Chimera , Rasmol , WINGX, le CCDC fournit une version gratuite de son programme de visualisation Mercury .

À partir de 2015, Mercury de CCDC fournit également la fonctionnalité pour générer un fichier prêt pour l'impression 3D à partir de structures dans CSD.

- **Mercury** est un programme permettant de visualiser les structures cristallines en trois dimensions. Il est capable de lire les structures cristallines dans différents formats et de faire pivoter et traduire l'affichage tridimensionnel de la structure cristalline.

-Il est également capable de mesurer et d'afficher les distances, les angles et les angles de torsion impliquant des atomes, des milieux et des plans.

Possibilité d'afficher les concentrateurs de cellules, le contenu de n'importe quel nombre de cellules dans n'importe quelle direction ou la section d'un cristal dans n'importe quelle direction.

-Localisez et affichez les liaisons hydrogène entre molécules et / ou molécules, des contacts courts illimités et des types de communication définis par l'utilisateur

Version actuelle Mercury peut lire les types de fichiers " cif ", " .mol ", ".mol2", " .pdb ", ".res", ".sd" et ".xyz". Mercury a son propre format de fichier avec l'extension de nom de fichier ".mryx".

Outils de dessin des molécules: Logiciel ChemDraw

TP5

But

Le but de TP consiste à connaître suffisamment bien le logiciel pour pouvoir :

-Dessiner des structures moléculaires en 2D ou 3D

- Visualiser ces molécules en 3D et les copier sous cette forme

-Calculer certaines propriétés des molécules à l'aide des outils de ChemDraw

1. Des molécules simples :

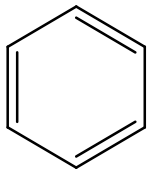
1.1. Création de la molécule :Phénol

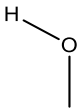
Ouvrir le logiciel **Chemdraw Ultra (Professional**) présent sur le bureau.

L'interface du logiciel apparaît

Il suffit d'un double clic dans la barre d'outils (voir la figure). On est maintenant prêt à dessiner une molécule:

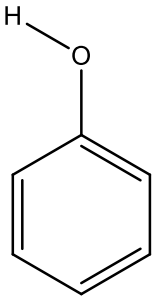
1. Commencer à créer la molécule en introduisant du cycle de carbone situé en bas à droite de la barre d'outils 



2. puis Introduire comme suit les deux liaisons 

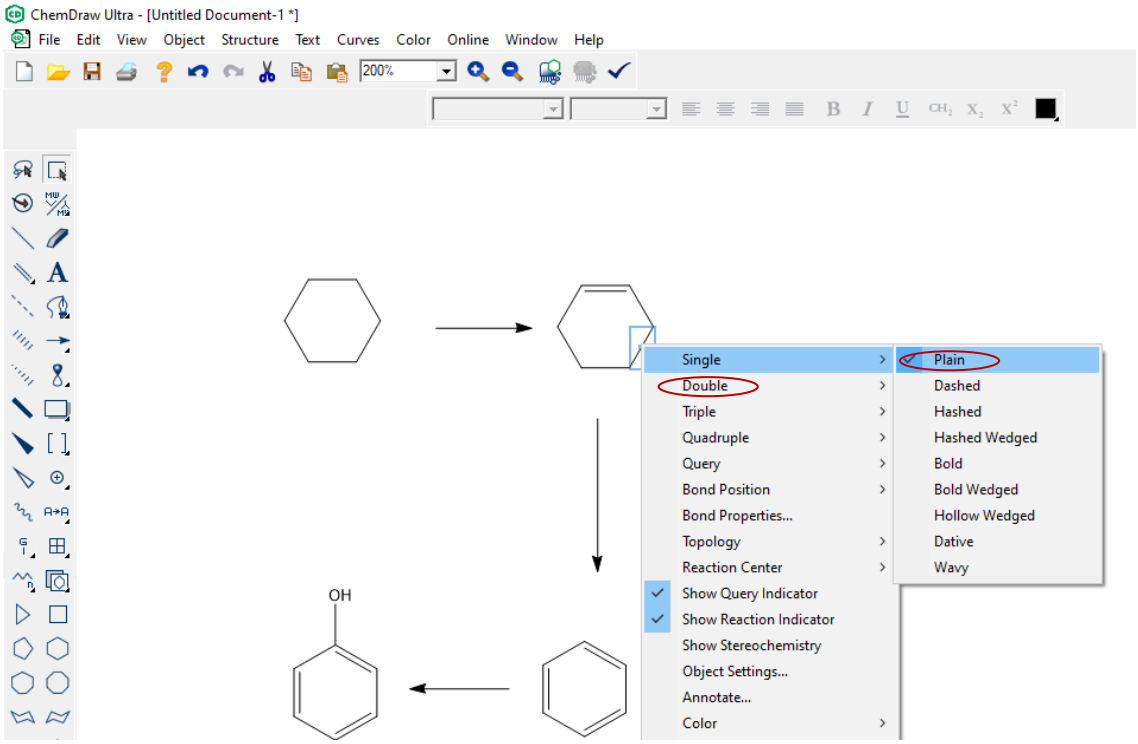
- Cliquer sur liaisons  et à l'aide de l'outil Texte  , tapez «O» puis «H»

La structure développée apparaît :



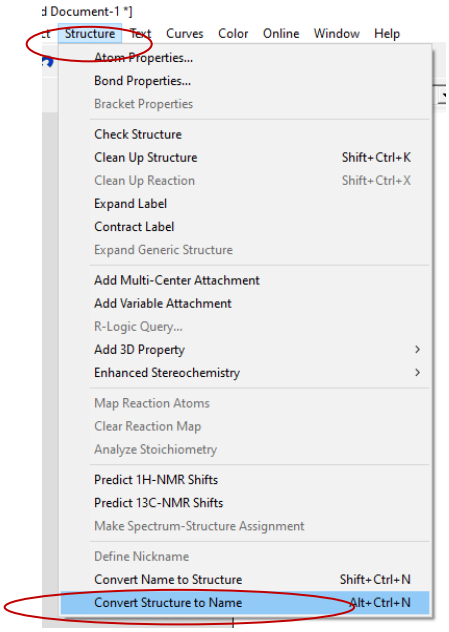
-Ou par Clique sur liaisons  Introduire le cycle de carbone (voir la figure ci-dessous). Pour ajouter des doubles liaisons à notre structure cyclique, on doit cliquer avec le bouton droit de la souris et sélectionnez **Double**>**Plain**

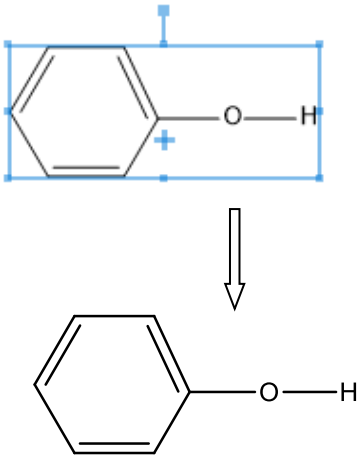
- Un champ de texte apparaît à l'aide de l'outil Texte, tapez «OH » dans le champ de texte.



1.2. Définir une nomenclature selon les standards IUPAC

Sélectionner la molécule avec  , cliquer sur Structure→Convert structure to name



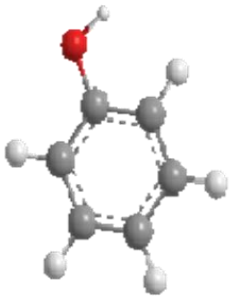


phenol

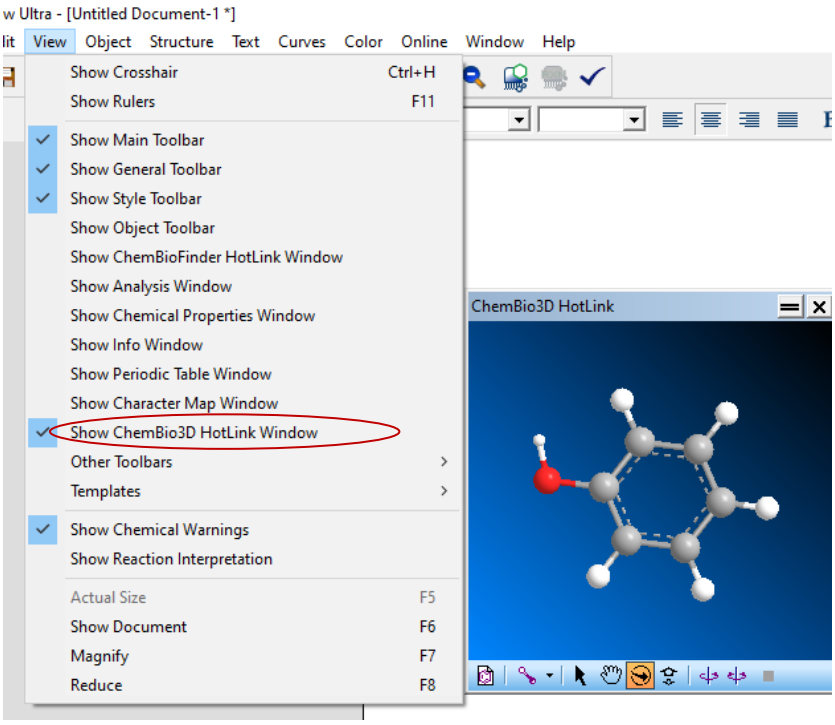
1. 3. Visualisation de la molécule en 3D

Pour obtenir une représentation en 3d :Sélectionner la molécule à représenter avec lasso  , cliquer sur Edit→Get 3D Model→La structure 3D apparaît.

La structure suivante apparaît :

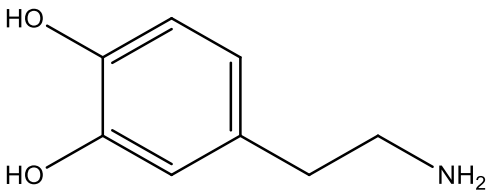


-Pour afficher le modèle dans Chem3D, cliquez sur View→Show Chem3D Hotlink Window



2. Application :

La molécule de dopamine :



TP relatif à la dopamine à préparer dans un document word :
En utilisant le logiciel ChemDraw :
1- Dessiner la molécule de la dopamine
2- Donner sa formule brute et sa masse moléculaire
3- Donner le pourcentage théorique de chaque atome présent dans la molécule
4- Utiliser une fonction du logiciel pour donner toutes les propriétés physico-chimiques de la dopamine.

5- En utilisant le logiciel, nommer en nomenclature systématique la dopamine
6- Prédire le spectre de RMN du proton et voir les différents déplacements des protons
7- Prédire le spectre de RMN du proton et voir les différents déplacements des carbones.
8- visualiser la molécule en 3D.Y a-t-il plusieurs possibilités ?

Annexe

Logiciel «ChemDraw»

Il existe des nombreux logiciels de représentation de molécules. Le plus populaire de ces logiciels se nomme ChemDraw.

ChemDraw s'est imposé depuis longtemps comme la référence des logiciels de dessin de structures moléculaires. Afin de proposer à chaque utilisateur l'outil adapté à ses besoins. ChemDraw est un logiciel développé par CambridgeSoft, Cambridge (PerkinElmer). C'est un logiciel exhaustif qui offre :

-La référence du dessin moléculaire

-Trois versions pour choisir à votre convenance

-De nombreux outils innovants

-Une application 3D

-Un carnet de laboratoire intégré

-L'utilisateur un outil exceptionnel de modélisation et une interface intuitive et facile à utiliser.

Le logiciel existe maintenant en trois versions : ChemDraw Prime, ChemDraw Professional et ChemOffice Professional.

1. ChemDraw Prime est la version comprenant les fonctionnalités essentielles. En plus de fournir les éléments usuels tels que les cycles, liaisons, chaînes, atomes et les groupes fonctionnels, ChemDraw Prime comprend des calculateurs de propriétés, des modèles d'équipements chimiques et de laboratoire ainsi que des outils de dessin de plaque CCM et gel d'électrophorèse.

2. ChemDraw Professional est l'outil complet destiné aux chimistes et biologistes, intégrant toute une gamme d'outils intelligents permettant de faciliter les travaux des chercheurs au quotidien. En plus des fonctionnalités de ChemDraw Prime, il inclut de nombreux outils innovants, tels que la prévision RMN ou la fonction nom=structure. Il permet également d'interroger les bases de données en ligne, d'explorer, organiser et traiter des données de structure grâce à ChemDraw pour Excel, ChemFinder standard et ChemScript, et s'intègre aux carnets de laboratoire.

Il inclut également Chem3D avec une interface d'application tiers et ChemFinder Ultra. Il aide les chimistes et biologistes à mieux visualiser leurs travaux, permet de gagner en compréhension et de corréler l'activité biologique avec les structures chimiques.

3. Chemdraw Professional est la suite complète et intelligente de dessin moléculaire permettant aux scientifiques et chercheurs de capter, stocker, retrouver, analyser et partager les données et informations sur les composés, réactions et propriétés.

Un logiciel pour dessiner des molécules. Une version gratuite est disponible en ligne : https://chemdrawdirect.perkinelmer.cloud/js/sample/index.html

Guide d'utilisation : Prise en main du logiciel **«ChemDraw»** Guide animé accessible à cette adresse : ChemDraw 19.0 User Guide : http://www.cstf.kyushu-u.ac.jp/~furutalab/pdf/ChemDraw-E.pdf (Manuel d'utilisation de Chemdraw en anglais)

Description générale de l'environnement (interface).

Lancer le logiciel **«ChemDraw»**. Voici la fenêtre principale du logiciel,

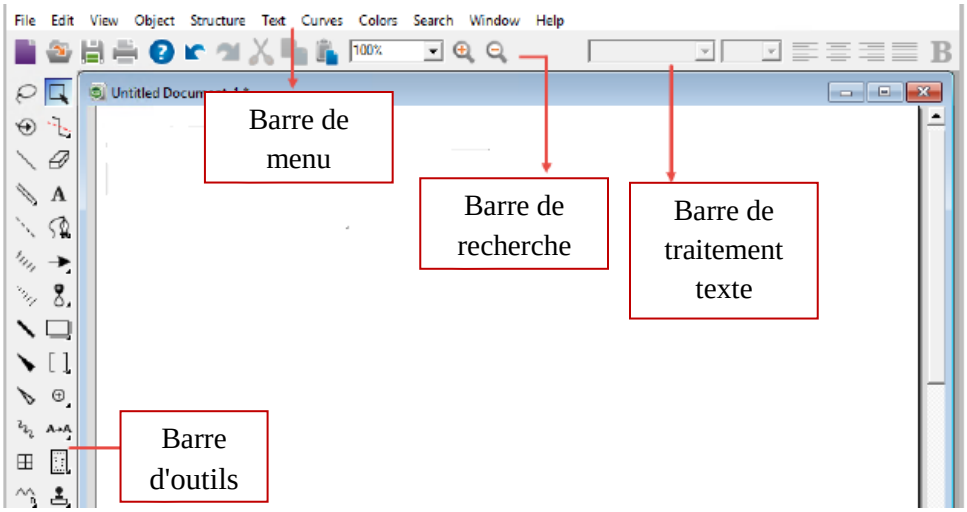


Figure 1 : *Interface utilisateur graphique Chemdraw*

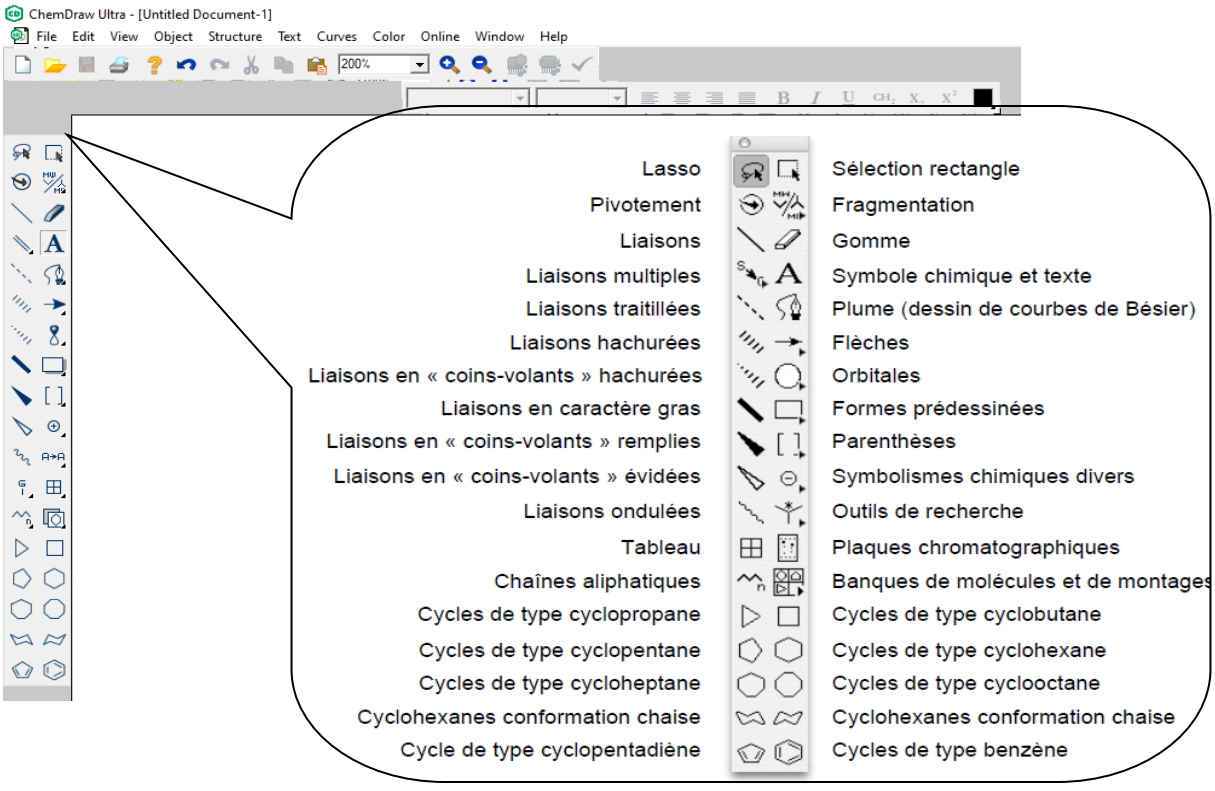


Figure 2 : *Les commandes les plus utilisées.*

Initiation à la modélisation moléculaire: Utilisation des logiciels Gaussian et GaussView

TP6

But:

L'objectif principal de ce TP est de se familiariser avec les outils de calculs de modélisation disponibles dans les 2 logiciels Gaussian 09/ GaussView (Gaussian 09 est un logiciel de calcul de chimie quantique. GaussView servira comme interface graphique pour celogiciel).

Pour cela, un exemple simple de molécule (H_2O) est traités afin de focaliser le TP sur les méthodes, les base et de limiter les temps de calculs.

-On mesurera les distances entre les atomes et l'angle HOH que l'on comparera aux valeurs expérimentales.

1. Optimisation de la géométrie par la méthode Hartree-Fock/3-21G.

Nous avons choisi d'étudier une molécule très simple, la molécule d'eau.

On donne quelques données structurales expérimentales concernant la molécule de l'eau.

Les paramètres géométriques sont : d(O-H) = 0,958 Å et Θ = 104,5°.

1.1. Construire la molécule H2O

Dans la fenêtre « logiciel GaussView » Construisez la molécule d'eau H2O.

-Pour construire la molécule H2O dans **GaussView,** ouvrez une nouvelle molécule en utilisant le menu « **Fichier** »→ « **Nouveau** » → « **Créer MolGroup** ».

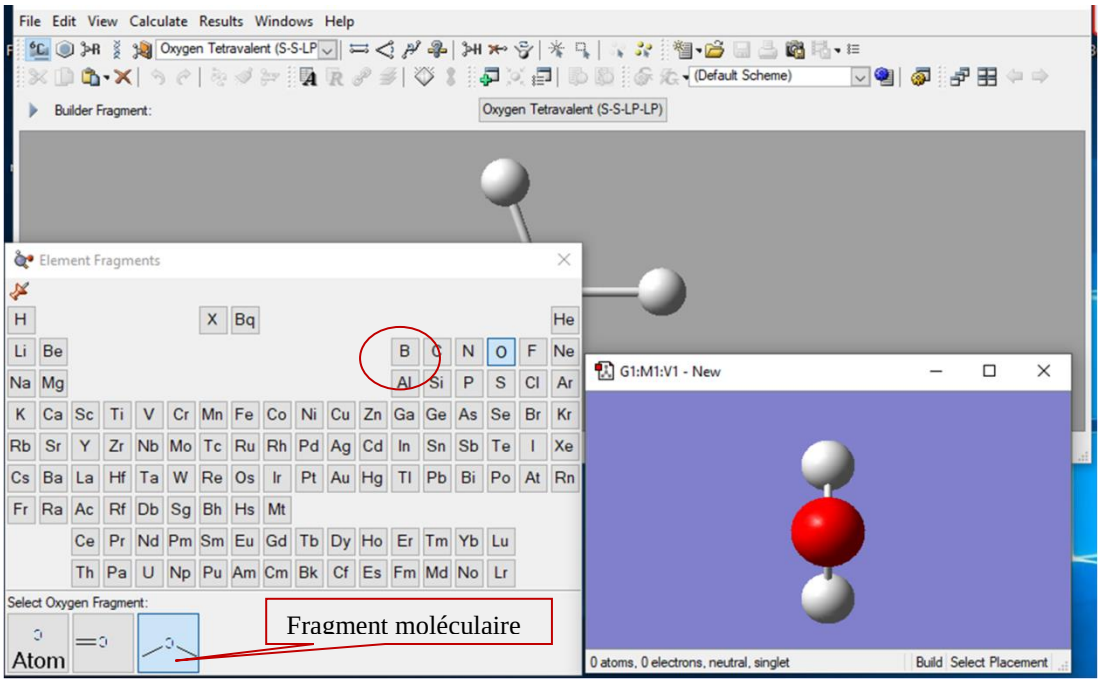
Un nouveau "Voir" fenêtre s'ouvrira.

-Sélectionnez l'icône nouvel atome,  Sélectionnez l'atome apparaît actuellement.

-Pour changer le type d'atome de configuration de liaison cliquez sur le bouton avec le type d'atome en cours.

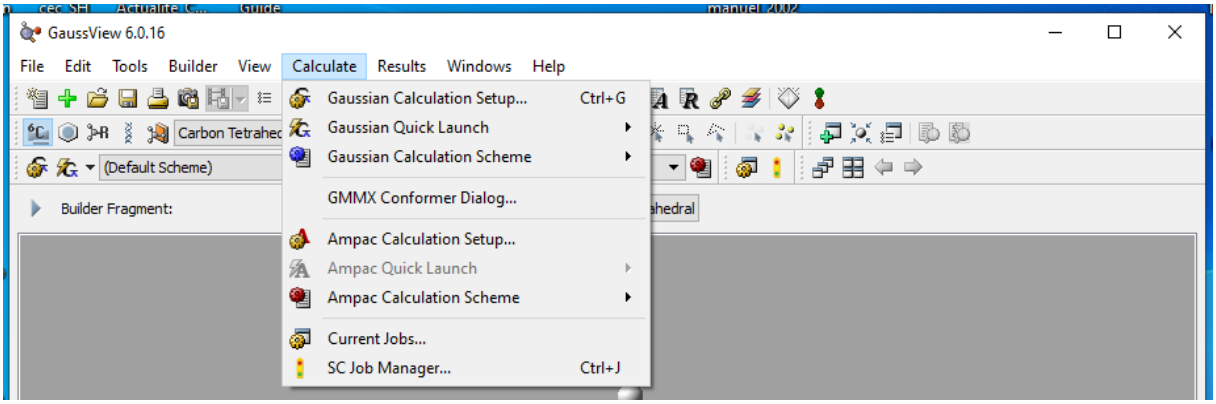
-Un tableau périodique des éléments est ouvert. L'atome de nouveau type et la configuration de liaison peuvent être choisis parmi la liste en cliquant sur les boutons appropriés.

-Pour ajouter un fragment moléculaire existant, cliquez sur l'un des atomes d'hydrogène, qui sera remplacé par la sélection actuelle atome.



1.2. Lancement du calcul

Le calcul peut être configuré en utilisant les « **calculate**»→ « **Gaussian calculation Setup**…» (voir la figure ci-dessous).

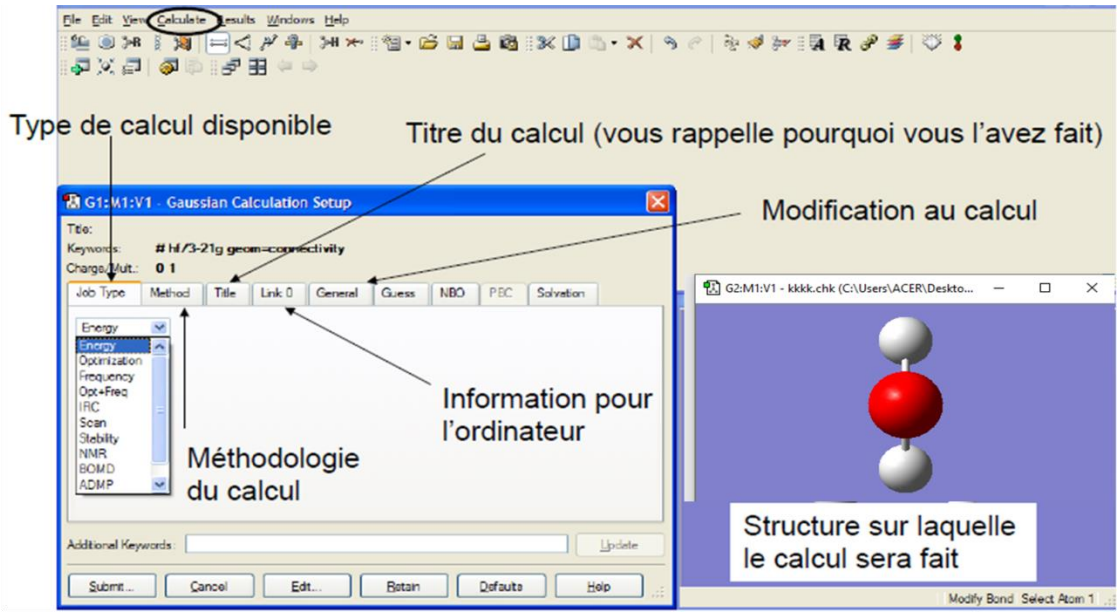


- « **Job Type** » onglet vous permettra de mettre en place l'un des différents types de calculs, y compris: un seul point, optimisation de la géométrie, les calculs de fréquence, la dynamique moléculaire ...

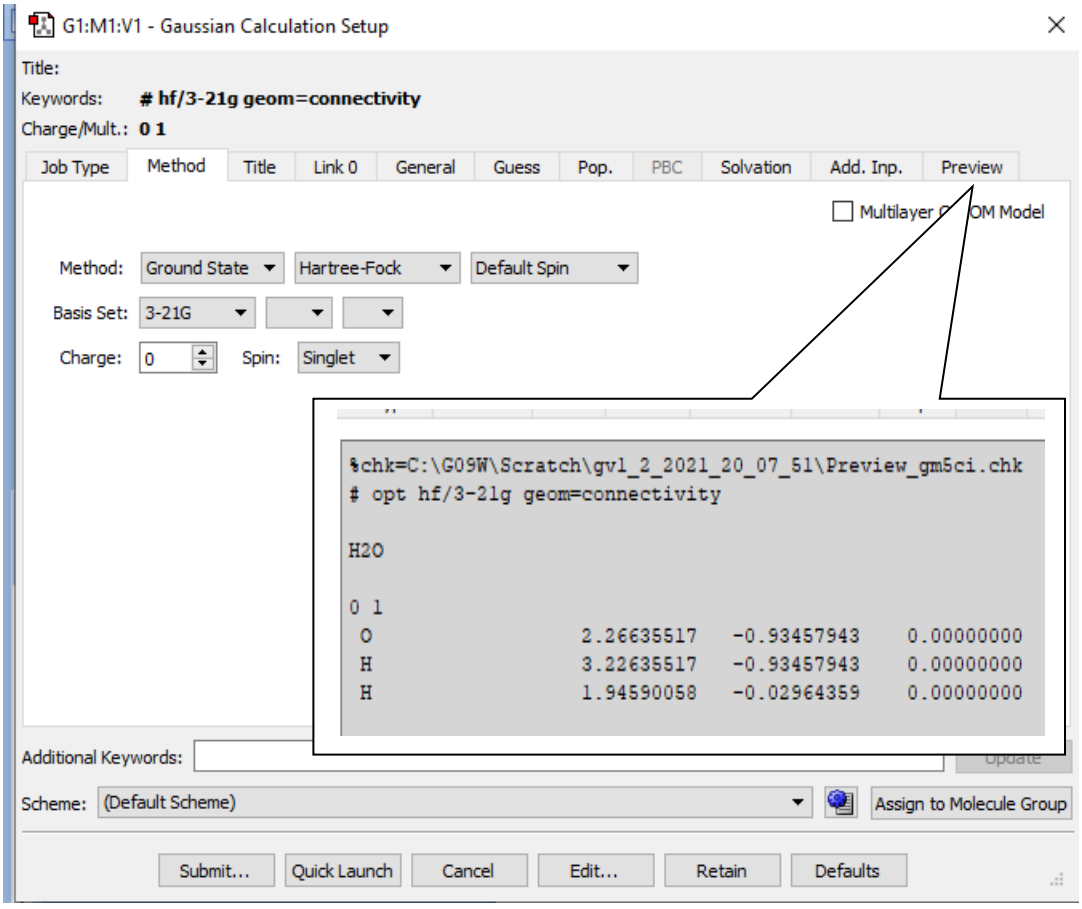
« **Method** » onglet vous permettra de définir la méthode, un ensemble de base, la charge et la multiplicité de spin.

« **Titre** » vous permet d'entrer un titre pour le travail et les commentaires que vous souhaitez inclure dans le fichier de sortie.

-Les autres onglets de fournir des informations spécifiques utilisés dans plusieurs types de calculs. Enfin, d'autres "mots clés" pour contrôler la nature des calculs peuvent être saisis. Lorsque les paramètres corrects ont été fixés, le dossier est soumis (cliquer sur l'icône «**Submit**»). Une fois terminé, le logiciel vous demandera de poursuivre l'action.



-Un fichier d'entrée « **input**» pour un Hartree-Fock calcul d'optimisation de la géométrie sur une molécule d'eau à l'aide de la 3-21G un ensemble de base gaussienne est ci-dessous:



 # opt hf/3-21g geom=connectivity

H2O

0 1

O 2.26635517 -0.93457943 0.00000000

H 3.22635517 -0.93457943 0.00000000

H 1.94590058 -0.02964359 0.00000000

-La première ligne indique le type de base définie et calcul.

-Tapisser deux doit être laissée en blanc.

-La troisième ligne est une ligne de commentaire.

-Ligne quatre est également vide.

-Les lignes restantes contiennent la charge moléculaire, la multiplicité de spin, types d'atomes, et de coordonnées cartésiennes.

- Comparer les valeurs calculées avec les valeurs expérimentales.
- Noter les énergies E_{tot} de molécule d'eau.
- Comment expliquez-vous les valeurs obtenues ? Commenter les différences observées.

2. Choix de la base et de la méthode dans un calcul (molécule d'eau) :

1. Nous allons maintenant optimiser la géométrie de la molécule d'eau par la méthode « HF » avec les bases STO-3G, 6-31G, 6-31G** , 6-311G, cc-pvdz. Compléter le tableau suivant :

	Energie (ua)	d(O-H) (Å)	Θ(HOH) (°)
Expérience			
STO-3G			
6-31G			
6-31G**			
6-311G			
cc-pvdz			
cc-ptdz			
cc-pqdz			

-Vérifier que la longueur de la liaison tend vers la valeur expérimentale en fonction de l'étendue de la base utilisée.

Quelle est la forme la plus stable ?

2. En partant de cette forme, optimiser la géométrie par les méthodes suivantes : DFT/B3LYP/6-31G** et MP2/6-31G** et compléter le tableau suivant :

	Energie (ua)	d(O-H) (Å)	Ɵ(HOH) (°)
Expérience			
B3LYP/6-31G**			
MP2/6-31G**			

- Laquelle des méthodes donne le résultat le plus proche de l'expérience ?

3. Géométrie moléculaire

Nous allons maintenant optimiser la géométrie en partant de H_2O linéaire : Cliquer sur « Draw Geom ». Cliquer sur l'icône « Measure ». Remplacer la valeur de l'angle (proche de 104) par 180. Créer le fichier de donner et lancer le calcul en optimisant la géométrie par la méthode B3LYP/6-31G*. La molécule reste-elle linéaire ? Comparer les énergies entre la structure trouvée ici avec celle trouvée en partant de la forme coudée.

.

1. J. Lonchamp, Introduction aux systèmes informatiques, InfoSup, Dunod, 2017.
2. J. Delacroix, A. Cazes, J. Delacroix, A. Cazes, Architecture des machines et des systèmes informatiques, 3e édition, InfoSup, Dunod, 2008.
3. H. Delalin, Systèmes d'Exploitation du 1ère année SRC. Université d'Artois, 2005.
4. A. Dalalyan, Statistique Numérique et Analyse des Données. Paris Thech, 2011.
5. M. Nekri, Recueil et Traitement Statistique des Données: Introduction Générale à la Statistique CERIST, 2011
6. M.-G. Ricard, Guide méthodologique pour les études et la recherche en Sciences de la nature. Cégep Trois- Rivières.
7. M. Ayadim, Chimie organique structurale: Manipuler les molécules pour les comprendre. Presses universitaires de Louvain, 2014.
8. D. Young, Introduction to Computational Chemistry. Chem. Aust. 11, 5.1998
9. G. H. Grant, W. G. Richards, Computational Chemistry. Oxford, 1995
10. D. Rogers, Computational Chemistry Using the PC.3rd Edition, John Wiley & Sons, 2003
11. D. Young, Computational Chemistry: A Practical Guide for Applying Techniques to Real World Problems. John Wiley & Sons, 2001
12. Georges Gardarin, Base de données, 5ème édition 2003.
13. A. Spicher, Base de données, Traduction modèle E/A schéma relationnel, L3 Informatique.
14. P. Clemente, Cours de Bases de données, SQL (Structured Query Langage), ENSI Bourges, Filière STI 2ème année, 2003-2004.
15. El-M. Daoudi, Cours Système d'Exploitation I, Université Mohammed Premier, Faculté des Sciences d'Oujda, Oujda – Maroc, 2015.
16. Hadji Djebar, Informatique pour la chimie, Université Dr. Moulay Tahar de Saïda, 2019

17. J. Vaillant, Eléments de Statistique descriptive, 2015.

18. M. Guidère, Méthodologie de la recherche, Ellipses, Paris, 2003.

19. N. Doumi, Cours Base de données, Master Chimie Théorique et Computationnelle, Université de Saïda,2020

20. Analyse chimique : Méthodes et techniques instrumentales, cours et exercices corrigés, 7ème éd, Paris (France) : Dunod, 2009

21. T. Tung Nguyen-Dang, Chimie quantique, Université Laval, Quebec A-2005

22. P. Chaquin, Pratique de la chimie théorique, LCT-UPMC

23. F. Rabilloud, Les Methodes Ab-Initio ´Hartree-Fock Et Post-Hartree-Fock, Universite de Lyon