

L'utilisation des outils informatiques chez les chimistes est devenue obligatoire afin de bien manipuler les informations moléculaires qui ont été, au cours des dernières années, stockées numériquement sur les ordinateurs dans des bases de données en très grandes quantités. De plus, la multiplication des données exploitables par les chimistes a donné lieu à une obligation de la numérisation, afin d'être capable de stocker, visualiser et traiter ces mêmes données aisément.

La discipline décrivant l'utilisation des outils informatiques pour traiter et résoudre des problèmes à la fois dans le domaine chimique et biologique est désignée par la Chémoinformatique [8]. Ses utilisations sont très variées et vont de la création et l'utilisation de base de données de petites molécules à la manipulation de fichiers en passant par les études statistiques. Cependant, son application la plus communément admise est dans le domaine de la recherche de nouveaux médicaments (Drug eDiscovery), domaine dans lequel elle joue un rôle central dans l'analyse et l'interprétation des données de structures et de propriétés collectées au cours des criblages à haut débit (technique se fait par les biologistes et visant à identifier des molécules nouvelles et potentiellement actives dans des bases de données de composés).

L'émergence de cette discipline peut être mise en parallèle avec la multiplication des données chimiques stockées numériquement. En effet, les quantités de données générées par les nouvelles approches de Drug design n'ont eu de cesse d'augmenter et il s'est avéré nécessaire, pour traiter les résultats de criblage à haut débit ou encore de la chimie combinatoire, de développer et d'utiliser des techniques informatiques [9].

Les avancées technologiques de la dernière décennie ont rendu possibles de nombreuses découvertes et applications inaccessibles auparavant. Par exemple, le nombre de composés disponibles dans les études de criblage a augmenté de manière exponentielle. En parallèle, les développements techniques dans le domaine de l'informatique et des technologies de communication ont permis la création de bases de données de composés comportant des millions d'entrées. L'exemple le plus pertinente pour illustrer ces avancées est la base de données «PubChem» développée par le NIH (National Institute of Health) [10]. Avec un contenu de plus de 31 millions de composés reliés à leurs activités biologiques, ce genre de bases de données nécessite le développement et l'utilisation d'outils mathématiques et statistiques afin de pouvoir accéder à de nouvelles découvertes en termes de développement

de nouveaux médicaments et à la compréhension «des relations entre structure et activité (QSAR)» ou bien «des relations entre structure et propriétés (QSPR)».

Les premiers essais de modélisation d'activités de molécules datent de la fin du 19<sup>ème</sup> siècle, lorsque Crum-Brown et Frazer [11] postulèrent que l'activité biologique d'une molécule est une fonction de sa constitution chimique. Mais ce n'est qu'en 1964 que furent développés les modèles de «contribution de groupes», qui constituent les réels débuts de la modélisation QSAR. Depuis, l'essor de nouvelles techniques de modélisation par apprentissage, linéaires d'abord, puis non linéaires, ont permis la mise en place de nombreuses méthodes ; elles reposent pour la plupart sur «la recherche d'une relation entre un ensemble de nombres réels, descripteurs de la molécule, et la propriété ou l'activité que l'on souhaite prédire».

### II.1 Définition des QSAR/QSPR

Lors d'une étude de «QSAR (Qualitative Structure-Activity Relationships)» ou de «QSPR (Qualitative Structure-Property Relationships)», on étudie les relations entre la structure et l'activité (propriété) d'un composé ou molécule, par exemple les effets d'une variation chimique locale sur une molécule à l'activité connue. En effet, certains changements chimiques sur certaines parties d'une molécule peuvent entraîner des variations de son activité biologique en agissant sur l'interaction avec la cible [12].

Ainsi, la méthodologie QSAR/QSPR permet de trouver un modèle mathématique qui met en corrélation l'activité (propriété) et la structure au sein d'une famille de composés. De nombreuses méthodes conceptuellement différentes peuvent être utilisées pour mettre en place les modèles mathématiques permettant de détecter des relations de type QSAR. Ainsi, les études QSAR/QSPR sont basées sur des méthodes informatiques, celles de modélisations, déjà exploitées dans différents domaines.

Les grandes phases de la mise en place d'un modèle QSAR/QSPR peuvent être décrites comme suit : Extraire les descripteurs à partir de la structure moléculaire, choisir les descripteurs adaptés à l'étude par rapport à l'activité (propriété) analysée, utiliser les descripteurs comme variables explicatives pour définir une relation qui les corrèle à l'activité en question, et enfin chaque modèle doit être validé sur des jeux de données de test [13-14].

### II.2 Les descripteurs moléculaires

De nombreuses recherches ont été menées, au cours des dernières décennies, pour trouver la

meilleure façon de représenter l'information contenue dans la structure des molécules, et ces structures elles-mêmes, en un ensemble de nombres réels appelés descripteurs ; une fois que ces nombres sont disponibles, il est possible d'établir une relation entre ceux-ci et une propriété ou activité moléculaire, à l'aide d'outils de modélisation classiques.

### II.2.1 Définition

Les descripteurs numériques réalisent un codage de l'information chimique en un vecteur de réels. Tout simplement, un descripteur moléculaire est une représentation mathématique d'une molécule, qui contient à la fois des informations sur la structure, et donc, implicitement ou explicitement, sur ses propriétés physico-chimiques. Ces informations peuvent être encodées par des valeurs scalaires, des vecteurs ou des chaînes de bits [13 - 15].

### II.2.2 Types de descripteurs moléculaires [8]

On dénombre aujourd'hui plus de 10000 descripteurs moléculaires, qui quantifient des caractéristiques physico-chimiques ou structurelles de molécules. Ils peuvent être obtenus de manière empirique ou non-empirique, mais les descripteurs calculés, et non mesurés, sont à privilégier : ils permettent en effet d'effectuer des prédictions sans avoir à synthétiser les molécules, ce qui est un des objectifs de la modélisation. Il existe cependant quelques descripteurs mesurés : il s'agit généralement de données expérimentales plus faciles à mesurer que la propriété ou l'activité à prédire (coefficient de partage eau-octanol [16], polarisabilité, ou potentiel d'ionisation). Les descripteurs moléculaires sont fréquemment classés par rapport à la dimensionnalité de la représentation moléculaire sur laquelle ils sont calculés : On parlera alors de descripteurs 1D, 2D ou 3D [12].

#### a) Les descripteurs 1D :

Sont appelés «descripteurs constitutionnels» et sont faciles et rapides à calculer. Ils sont accessibles à partir de la formule brute de la molécule (par exemple  $C_6H_6O$  pour le phénol), et décrivent des propriétés globales du composé. Il s'agit par exemple de sa composition, c'est-à-dire les atomes qui le constituent, ou de sa masse molaire. On peut remarquer que ces descripteurs ne permettent pas de distinguer les isomères [13-14].

#### b) Les descripteurs 2D :

Les descripteurs moléculaires utilisent la représentation des molécules comme des

graphes sont dits «descripteurs 2D» et contiennent des informations à propos de la connectivité ou à propos de certains fragments moléculaires, mais aussi des estimations des propriétés physico-chimiques. C'est à partir de ce niveau que l'on peut espérer la capture d'informations chimiques pertinentes pour la prédiction de la majorité des propriétés moléculaires. On trouvera dans cette catégorie les descripteurs suivants :

**Les indices topologiques**, qui considère la structure du composé comme un graphe, les atomes étant les sommets et les liaisons sont les arêtes. De nombreux indices quantifiant la connectivité moléculaire ont été développés en se basant sur cette approche, comme par exemple l'indice de Wiener [17], qui compte le nombre total de liaisons dans les chemins les plus courts entre toutes les paires d'atomes (en excluant les hydrogènes), et qui sera également l'axe central de cette thèse.

**Les indices constitutionnels**, qui se basent sur des motifs sous-structuraux. Par exemple, les empreintes BCI (Barnard Chemical Information Ltd) [18] sont des ensembles de bits indiquant la présence ou l'absence de certains fragments dans une molécule. Les fragments prennent en compte les atomes et leurs plus proches voisins, les paires d'atomes et les séquences ou encore les fragments basés sur des cycles. L'approche des clés MDL est une approche similaire comprenant la recherche des 166 fragments MDL [13, 14, 19].

Ces descripteurs 2D reflètent bien les propriétés physiques dans la plupart des cas, mais sont insuffisants pour expliquer de façon satisfaisante certaines propriétés ou activités, telles que les activités biologiques.

### c) Les descripteurs 3D :

Les descripteurs 3D d'une molécule sont évalués à partir des positions relatives de ses atomes dans l'espace, et décrivent des caractéristiques plus complexes ; leurs calculs nécessitent donc de connaître, le plus souvent par «modélisation moléculaire empirique» ou «*ab initio*», la géométrie 3D de la molécule. Ces descripteurs s'avèrent donc relativement coûteux en temps de calcul, mais apportent davantage d'informations, et sont nécessaires à la modélisation de propriétés ou d'activités qui dépendent de la structure 3D. On distingue plusieurs familles importantes de descripteurs 3D :

**Les descripteurs géométriques** : parmi ceux qui sont les plus importants sont le volume moléculaire, la surface accessible au solvant et le moment principal d'inertie.

**Les descripteurs électroniques** : ils permettent de quantifier différents types d'interactions inter et intramoléculaires, de grande influence sur l'activité biologique de molécules. Le

calcul de la plupart de ces descripteurs nécessite la recherche de la géométrie pour laquelle l'énergie stérique est minimale, et fait souvent appel à la chimie quantique. Par exemple, les énergies de la plus haute orbitale moléculaire occupée et de la plus basse vacante sont des descripteurs fréquemment sélectionnés. Le moment dipolaire, le potentiel d'ionisation, et différentes énergies relatives à la molécule sont d'autres paramètres importants.

**Les descripteurs spectroscopiques :** les molécules peuvent être caractérisées par des mesures spectroscopiques, par exemples par leurs fonctions d'onde vibrationnelles. En effet, les vibrations d'une molécule dépendent de la masse des atomes et des forces d'interaction entre ceux-ci ; ces vibrations fournissent donc des informations sur la structure de la molécule et sur sa conformation. Les spectres infrarouges peuvent être obtenus soit de manière expérimentale, soit par calcul théorique, après recherche de la géométrie optimale de la molécule. Ces spectres sont alors codés en vecteurs de descripteurs de taille fixe. Les descripteurs de type MORSE [20] (Molecule Representation of Structures based on Electron diffraction) font appel au calcul des intensités théoriques de diffraction d'électrons.

### II.3 Modèles QSAR de la toxicité (QSTR) :

Les QSAR pour la toxicité remonte au 19<sup>ème</sup> siècle. En 1863, A.F.A. Crois à l'Université de Strasbourg a observé que la toxicité des alcools à des mammifères augmente lorsque la solubilité des alcools dans l'eau diminue. Dans les années 1890, Hans Horst Meyer de l'Université de Marburg et Charles Ernest Overton de l'Université de Zurich, ont noté que la toxicité des composés organiques était tributaire de la lipophilie [21]. Au début des années 1960 Corwin Hansch a proposé un modèle mathématique pour corrélérer l'activité biologique et la structure chimique [22], cette date est considérée comme étant la naissance des méthodes QSAR. Depuis, l'utilisation des QSAR en toxicologie n'a pas cessé d'évoluer.

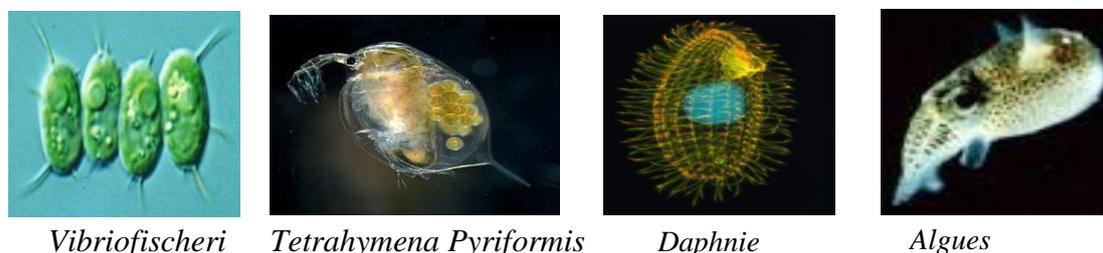
Des modèles QSAR sont maintenant mis au point en utilisant une variété d'approches, de méthodes d'analyse de données et de paramètres [23].

Un grand nombre d'études QSAR de toxicité et, en particulier la toxicité aiguë ont été publiées dans la littérature. La plupart des données de toxicité pour l'environnement ont été obtenues en utilisant des espèces aquatiques (représentées dans la figure 3) en l'occurrence les poissons, les daphnies, les protozoaires « *Tetrahymena Pyriformis* », *Vibrio fischeri*, les algues ...

Cronin et Dearden [24- 25] ont examiné la littérature concernant la modélisation QSAR de la toxicité aquatique. Plusieurs modes d'action ont été identifiés chez les espèces aquatiques, à savoir narcose non polaire, narcose polaire, découplage de la phosphorylation

## II. Modélisation QSAR/QSPR

oxydative, irritation de la membrane respiratoire, inhibition d'acétylcholinestérase, saisie de système nerveux central, inhibition de photosynthèse, et l'alkylation. Cependant, ceux-ci sont généralement plus largement regroupés comme : narcose non polaire, narcose polaire, réactifs sélectifs (électrophiles), et les molécules à mécanismes d'action spécifiques.



**Figure 3** : Différentes espèces aquatiques utilisées pour l'étude de la toxicité aquatique [1].

Verhaar *et al.* [26] ont développé un système sur la base de la présence de groupes fonctionnels pour classer les produits chimiques dans ces quatre groupes **Tableau 3** [1].

**Tableau 3.** Mécanismes possibles pour la toxicité [26]

Mécanisme (Mode d'action)	Structures déterminantes
Narcose non-polaire	Par exemple les alcanes saturés avec halogène et /ou substituants alcoxy (alcools aliphatiques, les cétones, les éthers, des amines)...
Narcose polaire	Les phénols, phénols et anilines avec trois ou moins d'atomes d'halogène, et/ou substituants alkyle...
Formation des radicaux libres	Phénols et anilines avec quatre ou plusieurs atomes d'halogène, ou plus d'un groupe nitro, ou seul un groupe nitro, et plus d'un groupe d'halogène...
Electrophiles /pro-électrophiles	Certains nitrobenzènes ; des noyaux benzéniques sans aniline ou de phénol qui ont deux groupes nitro sur un noyau ; phénols avec un seul groupe nitro et un halogène ; composés aromatiques ayant deux ou plusieurs groupes hydroxy dans la position ortho ou para, quinines ; aldéhydes ; composés aromatiques avec deshalogènes; cétènes; époxydes...

La dépendance de la toxicité des narcoses avec le coefficient de partage, en particulier le coefficient de partage octanol-eau, a été montrée par de nombreux auteurs [27-35]. Le QSAR représentant une toxicité de référence a été dérivé pour un groupe de narcotiques non polaires (alcools saturés, cétones, nitriles, esters et des composés

contenant du soufre). Ils ont conclu que le coefficient de partage octanol-eau est suffisant pour expliquer la toxicité des narcoses apolaires tandis que pour les narcoses polaires il faut la présence d'un descripteur supplémentaire qui explique le caractère électronique des molécules.

En outre, des modèles de combinaison des deux groupes de composés (narcoses apolaires et polaires) et les deux types de descripteurs ont été développés. Freidig et Hermens [28] ont conclu qu'en utilisant des modèles QSAR distincts pour les composés agissant par des mécanismes différents, y compris un descripteur qui caractérise le mécanisme de toxicité particulière, donne de meilleurs résultats que l'utilisation d'un modèle unique qui combine tous les composés avec les mêmes descripteurs.

Certains auteurs ont utilisé l'approche surface de réponse sur la base de l'hydrophobie et l'électrophilie des composés. Dans cette approche, les QSAR comprennent un descripteur qui caractérise la bio-absorption et la distribution (généralement partage octanol-eau ou coefficients de distribution ( $\log P$  ou  $\log D$ ) et un descripteur de réactivité électrophile (habituellement LUMO ou la super delocalisabilité maximale ( $A_{\max}$ )). Cette approche a été appliquée à des différentes espèces aquatiques, y compris la bactérie *Vibriofischeri* [36], les protozoaires *Tetrahymena pyriformis* [37-39], les algues de *Scenedesmus* [40] et *Chlorellavulgaris* [41]. L'avantage de l'approche surface de réponse est qu'elle est simple et a une interprétation mécanistique. Alors que certains auteurs (par exemple, Cronin et Schultz [37], Cronin *et al.* [36], Cronin *et al.* [39]) ont utilisé LUMO comme descripteur de réactivité électrophile entraînant des interactions covalentes dans les systèmes biologiques. Dimitrov *et al.* [42-43] ont suggéré que LUMO peut être également utilisé pour décrire l'interaction électrophile non covalente des produits chimiques narcotiques avec le site d'action. Certains auteurs ont prolongé la démarche réponse-surface en ajoutant un indicateur supplémentaire et d'autres paramètres afin d'améliorer l'ajustement statistique qui est difficile des modèles [44-47]. Toutefois, selon Schultz *et al.* [48] la modélisation QSAR des composés électrophiles en raison de la limitation des données et des descripteurs par rapport à la modélisation QSAR des composés agissant par d'autres mécanismes toxiques. xxx. Cependant, des QSAR basés sur des indices topologiques pour l'étude de la toxicité ont fait l'objet de nombreux travaux [32, 49-52].

### III-1 Généralités

Les techniques de calcul qui peuvent fournir la valeur de l'énergie d'une géométrie, aussi particulière que l'état fondamental, appartiennent à plusieurs catégories [53]:

- méthodes *ab initio*,
- méthodes semi-empiriques,
- méthodes empiriques,
- mécanique moléculaire.

Concernant les deux premières méthodes, elles sont fondées sur l'évaluation des interactions électroniques complètes ou partiellement négligées. Le terme *ab initio* est réservé aux calculs déduits directement des principes théoriques, sans faire intervenir de données expérimentales. Deux méthodes fondamentales sont proposées pour la résolution de l'équation de Schrödinger à partir des principes de base. La théorie des orbitales moléculaires (OM) tend à établir une expression pour la fonction d'onde  $\psi$ , alors que dans la théorie de la fonctionnelle de la densité (DFT), la distribution de la densité électronique ( $\rho$ ) joue ce rôle. Le fondement de la DFT est associé à un théorème dû à Hohenberg et Kohn [54] qui ont démontré que toutes les propriétés d'un système dans un état fondamental non dégénéré sont complètement déterminées par sa densité électronique.

Le type le plus courant de calcul *ab initio*, ou calcul Hartree-Fock (HF), repose sur l'approximation principale du champ central. Le calcul variationnel mis en œuvre conduit à des énergies supérieures aux énergies réelles (Théorème de Eckart) et tendent vers une valeur limite appelée limite de Hartree Fock. La seconde approximation dans les calculs HF consiste à décrire la fonction d'onde par une « fonction utile » qui est connue exactement pour quelques systèmes mono-électroniques. Les fonctions les plus souvent utilisées sont des combinaisons linéaires d'orbitales de type Slater ( $e^{-a x}$ ) ou d'orbitales gaussiennes ( $e^{-a x^2}$ ), dont les abréviations sont, respectivement, STO (pour Slater Type Orbitals) et GTO (pour Gaussian Type Orbitals). La fonction d'onde est obtenue à partir de combinaisons linéaires d'orbitales, ou plus souvent à partir de combinaisons de fonctions d'un ensemble de base. A cause de cette approximation, la plupart des calculs HF conduisent à des énergies supérieures à la limite HF.

L'utilisation de bases de fonctions gaussiennes permet de calculer toutes les intégrales de la méthode sans autres approximations que celles inhérentes à la méthode elle-même.

Réservées initialement au traitement de petites molécules (une dizaine d'atomes), les méthodes *ab initio* ont été étendues, ces dernières décennies, à des systèmes de quelques centaines d'atomes, comme conséquence de l'augmentation de la puissance des ordinateurs (hardware et software).

Une approximation sur l'hamiltonien est considérée comme une méthode semi-empirique.

Les méthodes semi-empiriques sont moins contraignantes en moyens de calculs. De plus, l'incorporation de paramètres déduits des données expérimentales dans certaines de ces méthodes permet de prédire quelques propriétés avec une meilleure précision que celle obtenue avec les méthodes *ab initio* les plus élaborées.

Les méthodes de champ de force ne demandent pas de temps excessifs de calcul pour donner des informations sur l'énergie de la molécule étudiée. La mécanique moléculaire (M M), appelée parfois : calcul par champ de force empirique, (empirical Force Field, EFF, en anglais), permet le calcul de la structure et de l'énergie d'entités moléculaires [55-57]. D'une part, les distributions électroniques ne sont pas explicitement détaillées (à quelques exceptions près), d'autre part, la recherche de l'énergie minimale par optimisation de la géométrie joue un rôle primordial.

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. Les variables de calcul sont alors les coordonnées internes du système : longueur de liaison, angles de valence, angles dièdres et distances entre les atomes non liés. Un calcul de MM aboutit à une disposition des noyaux telle que la somme de toutes les contributions énergétiques est minimisée ; ses résultats concernent surtout la géométrie et l'énergie de système [58].

#### III-2 Méthodes semi-empiriques utilisées

Les méthodes AM1 et PM3 utilisées étant des re-paramétrisations de la méthode MNDO, nous présenterons ces trois méthodes, en rappelant au préalable le cadre des équations (*ab initio*) HFR (Hartree-Fock-Roothaan) sur lequel elles sont basées et les approximations supplémentaires auxquelles il est fait recours.

##### III-2.1 Le cadre Hartree - Fock – Roothaan

Les méthodes *ab initio* utilisent l'équation de Schrödinger électronique obtenue après séparation des mouvements électroniques et nucléaires (approximation de Born-Oppenheimer) [59, 60].

Dans la méthode Hartree – Fock la fonction d'onde  $\psi$  d'un système à N électrons est représentée par un déterminant de Slater  $\psi_0$  de spin orbitales  $\phi$  unique. Les spin orbitales consistent en des produits d'orbitales moléculaires (OM)  $\phi$  et de fonctions de spin ( $\alpha$  ou  $\beta$ ),  $\phi_a = \phi_a \alpha$ ,  $\bar{\phi} = \phi_a \beta$ .

On représentera  $\psi_0$  par :

$$\Psi_0 = |\Phi_1 \bar{\Phi}_1 \Phi_2 \bar{\Phi}_2 \dots \Phi_M \bar{\Phi}_M \rangle \quad (1)$$

pour un système à couches complètes comportant N électrons (auquel cas  $M = \frac{N}{2}$ ).

Chaque OM est développée sous forme d'une combinaison linéaire de fonctions de base, appelées conventionnellement orbitales atomiques (OM-CLOA, combinaison linéaire d'orbitales atomiques), quoiqu'elles ne soient pas généralement, solutions du problème HF atomique.

$$\phi_a = \sum_{\mu}^m c_{\mu a} \chi_{\mu} \quad (2)$$

En tenant compte de (1), on obtient après multiplication à gauche par une fonction spécifique, intégration et application du principe variationnel, un système d'équations linéaires, ou équations de Roothaan – Hall (pour un système à couches complètes) [61,62].

Signalons que la résolution des équations de Roothaan – Hall fournit un total de m (= nombre de fonctions de base) orbitales moléculaires (OM) dont n sont occupées et (m - n) libres ou virtuelles. Celles-ci sont orthogonales à toutes les orbitales occupées, mais n'ont pas d'interprétation physique directe exceptée comme affinité électronique (via le théorème de Koopmans [63]). Elles servent dans la description des états excités.

L'équation (3) condense, sous forme matricielle, les équations de Roothaan – Hall.

$$\mathbf{F} \mathbf{C} = \mathbf{S} \mathbf{C} \boldsymbol{\varepsilon} \quad (1)$$

où:

- la matrice  $\mathbf{F}$  de Fock est l'opérateur hamiltonien effectif,
- $\mathbf{C}$  est la matrice des coefficients des OM,  $c_{\mu a}$ ,
- $\mathbf{S}$  est la matrice de recouvrement,
- et  $\boldsymbol{\varepsilon}$  une matrice diagonale comportant les énergies orbitales.

La matrice de Fock,  $\mathbf{F}$ , comporte toutes les informations relatives au système quantomécanique, c'est – à – dire toutes les interactions prises en compte dans les calculs. Sa formulation *ab initio* est la suivante :

$$\mathbf{F}_{\mu\nu} = \mathbf{H}_{\mu\nu} + \mathbf{J}_{\mu\nu} - \frac{1}{2} \mathbf{K}_{\mu\nu}$$

$$\mathbf{F}_{\mu\nu} = \mathbf{H}_{\mu\nu} + \sum_{\rho}^n \sum_{\sigma}^m \mathbf{P}_{\rho\sigma} \left[ \langle \mu\nu/\rho\sigma \rangle - \frac{1}{2} \langle \mu\sigma/\rho\nu \rangle \right] \quad (2)$$

Avec :

$$\mathbf{H}_{\mu\nu} = \int \chi_{\mu}^*(1) \hat{h} \chi_{\nu}(1) d\tau_1 \quad (3)$$

$$\langle \mu\nu/\rho\sigma \rangle = \iint \chi_{\mu}^*(1) \chi_{\nu}(1) \frac{1}{r_{12}} \chi_{\rho}^*(2) \chi_{\sigma}(2) d\tau_1 d\tau_2 \quad (4)$$

### III. Bases de modélisation moléculaires

$$\text{et } P_{\rho\nu} = 2 \sum_a^m C_{\rho a}^* C_{\nu a} \quad (7)$$

où  $\mu, \nu, \rho$  et  $\sigma$  désignent des orbitales atomiques, et  $H_{\mu\nu}$  des intégrales mono-électroniques représentant les valeurs moyennes de l'opérateur associé à l'énergie cinétique et l'opérateur énergie potentielle d'interaction noyau – électron ( $\widehat{V}_{en}$ ). Les  $\langle \mu\nu/\rho\sigma \rangle$  sont des intégrales de répulsion bioélectroniques représentant  $\widehat{V}_{ee}$  (opérateur d'interaction entre les électrons eux - mêmes), et les  $P_{\rho\nu}$  sont les éléments de la matrice densité  $\mathbf{P}$ .

$\mathbf{J}_{\mu\nu}$  et  $\mathbf{K}_{\mu\nu}$  sont les représentations matricielles des opérateurs coulombien  $\hat{J}$  et d'échange  $\hat{K}$  respectivement.

L'énergie électronique ( $E_{el}$ ) peut être exprimée au moyen des valeurs propres  $\varepsilon_a$  :

$$E_{el} = 2 \sum_a^m \varepsilon_a - \frac{1}{2} \sum_{\mu\nu}^m P_{\mu\nu} \left( \mathbf{J}_{\mu\nu} - \frac{1}{2} \mathbf{K}_{\mu\nu} \right) \quad (5)$$

Comme la matrice de Fock dépend des coefficients des orbitales, les équations de Roothaan doivent être résolues de façon itérative en utilisant la procédure du champ auto-cohérent ou SCF (pour : Self Consistent Field) [64].

Une étape importante de la procédure SCF est la conversion de l'équation générale aux valeurs propres (3) en une équation ordinaire par une transformation orthogonale (méthode d'orthogonalisation de Löwdin) [65,66].

$$\mathbf{F}^\lambda \mathbf{C}^\lambda = \mathbf{S}^{-1/2} \mathbf{F}$$

$$\text{Avec } \mathbf{F}^\lambda = \mathbf{S}^{-1/2} \mathbf{F} \mathbf{S}^{-1/2} \quad (9)$$

$$\text{et : } \mathbf{C}^\lambda = \mathbf{S}^{1/2} \mathbf{C}$$

Notons que  $\mathbf{S}^{-1/2}$  qui est obtenue à partir de la matrice de recouvrement  $\mathbf{S}$  qui n'est jamais singulière, n'est jamais singulière non plus.

Signalons que les approximations électroniques et CLOA (utilisation d'un nombre limité d'orbitales atomiques) et un problème de corrélation limitent la méthode HFR. On dépasse ces limitations par l'utilisation de fonctions corrélées ou en faisant intervenir l'interaction de configuration.

#### III.2.2 Analyse de population de Mulliken [67, 68].

L'analyse de population répond à la question de savoir comment, lors de la formation d'une molécule à N électrons, répartir équitablement ces électrons entre les atomes, de telle sorte que la notion d'atome dans la molécule ne disparaisse pas complètement, sans enfreindre toutefois les principes de la mécanique ondulatoire et en tenant compte de la géométrie particulière de la molécule.

Soit  $d\tau (= d\tau_1 d\tau_2 \dots d\tau_i)$  un élément de volume,  $\psi$  la fonction d'onde déterminantale, la probabilité :

$$dp = \psi^* \psi d\tau \tag{6}$$

représente, à un facteur près, la quantité d'électricité dans  $d\tau$  à un moment donné, puisque :

$$dQ = edP \tag{7}$$

En prenant pour unité de charge, la charge de l'électron on n'introduit plus e avec dP.

Ainsi, si on intègre dP sur l'espace de configuration E, il vient :

$$\int_E dp = N \quad (\text{N électrons}) \tag{8}$$

Le résultat serait une somme sur les orbitales occupées ( les  $\psi$  du déterminant de Slater), c'est -à- dire :

$$\int_E \psi^* \psi d\tau = N = \sum_{i=1}^m 2 \int \psi^*(\nu) \psi_i(\nu) d\tau_\nu \tag{9}$$

Les orbitales étant normées :  $\langle \psi_i(\nu) / \psi_i(\nu) \rangle = 1$

En décomposant sur la base atomique, on obtient :

$$\begin{aligned} \sum_{i=1}^m 2 \int \left( \sum_{l=1}^n c_{li}^* \varphi_l(\nu) \right) \left( \sum_{m=1}^n c_{mi} \varphi_m(\nu) \right) d\tau_\nu &= 2 \sum_{i=1}^m \sum_{l=1}^n \sum_{m=1}^n c_{li}^* c_{mi} \int \varphi_l(\nu) \varphi_m(\nu) d\tau_\nu \\ &= \sum_{l=1}^n \sum_{m=1}^n p_{lm} s_{lm} = \sum_{l=1}^n \left[ \sum_{m=1}^n p_{lm} \right] = N \end{aligned} \tag{10}$$

Avec :

$$s_{lm} = \int \varphi_l^*(\nu) \varphi_m(\nu) d\tau_\nu$$

et :

$$P_{Lm} = \sum_{i=1}^m 2 c_{li}^* c_{mi}$$

### III. Bases de modélisation moléculaires

En posant :  $q_l = \sum_{m=1}^n P_{lm} S_{lm}$ , on se donne un moyen de répartir le nombre d'électrons de la molécule :  $q_l$  est la quantité d'électricité qui peut être attribuée à la 1<sup>ère</sup> orbitale atomique de base.

Remarque : la relation  $\int_E \psi^* \psi d\tau = N$  met en exergue les principes de la mécanique ondulatoire, alors que la géométrie particulière de la molécule ressort dans  $p_{lm}$  et  $S_{lm}$ . La quantité d'électricité attribuée à l'atome L est la somme des  $q_{l(L)}$  ( $l \in L$ ) :

$$Q_L = \sum_{l(L)} q_{l(L)} \quad (11)$$

La relation d'identité initiale  $\sum_l \sum_m p_{lm} S_{lm} = N$  peut être écrite sous la forme :

$$\sum_l q_l = N \quad (12)$$

$$\sum_L Q_L = N$$

$q_l$  est la densité électronique de l'orbitale, et  $Q_L$  celle de l'atome L.

La charge,  $C_L$ , de l'atome L dans la molécule est :

$$C_L = Z_L - Q_L \quad (13)$$

$Z_L$  étant le nombre d'électrons de l'atome isolé, et  $Q_L$  la quantité d'électricité qu'il possède dans l'atome.

#### III.2.3 Les méthodes semi-empiriques.

Dans les méthodes semi-empiriques on simplifie l'approche Hartree-Fock - Roothaan.

1) Dans la construction de  $\Psi_0$  : seuls les électrons de valence sont traités de façon explicite en utilisant un ensemble de base minimal. Ce qui signifie que les atomes H sont décrits par une fonction 1s, les éléments Li à F par un ensemble {2s, 2p}, les éléments Na à Cl par un ensemble {3s, 3p}, Ca, K, et Zn à Br avec un ensemble {4s, 4p}, Sc – Cu avec un ensemble de base {4s, 4p, 3d} ; etc...

On tient compte des électrons de cœur soit en corrigeant la charge nucléaire, soit en introduisant des fonctions pour modéliser les répulsions simultanées entre noyaux d'une part et entre électrons de cœur d'autre part.

2) Dans la construction de  $F^\lambda$  on néglige une grande part des interactions, en particulier dans la partie bi-électronique  $\langle \mu\nu/\rho \sigma \rangle$ . Toutes les intégrales mettant en jeu des orbitales atomiques centrées sur plus de 2 noyaux sont négligées. Certaines classes d'intégrales sont remplacées par des paramètres. C'est le cas, en particulier, des intégrales mono-électroniques bi-centres  $H_{\mu\nu}$  qui sont, pour une large part, responsables de la liaison chimique.

La façon d'introduire ces simplifications dans le modèle permet de distinguer entre les différentes méthodes.

Une autre façon de réduire les intégrales bi-électroniques est l'approximation du recouvrement différentiel nul (RDN) dans laquelle on néglige tous les produits des fonctions de base dépendant des coordonnées d'un même électron localisé sur des atomes différents. Cela signifie que tous les produits des fonctions orbitales atomiques  $\chi_\mu \chi_\nu$  sont posés égaux à zéro et l'intégrale de recouvrement se réduit à  $S_{\mu\nu} = \delta_{\mu\nu}$  ( $\delta_{\mu\nu}$  est le symbole de Kronecker ;  $\delta_{\mu\nu} = 0$  si  $\mu \neq \nu$  et  $\delta_{\mu\nu} = 1$  si  $\mu = \nu$ ).

Dans l'approximation RDN, toutes les intégrales tri et tétra-centres s'annulent ce qui transforme la matrice de recouvrement en une matrice unité. Les intégrales mono-électroniques tri-centres sont égalées à zéro. Toutes les intégrales bi-électroniques tri et tétra-centres sont négligées.

Les paramètres sont imposés pour compenser les approximations. Ainsi toutes les intégrales restantes sont remplacées par des paramètres convenables ajustés sur des grandeurs fournies par l'expérience.

Toutes les méthodes semi-empiriques modernes sont basées sur l'approche MNDO (Modified Neglect of Differential Overlap) [69], dans laquelle des paramètres sont assignés aux différents types d'atomes puis ajustés de telle sorte à reproduire certaines propriétés comme les chaleurs de formation, les variables géométriques, les moments dipolaires et les énergies de première ionisation.

Les paramètres sont conçus séparément pour des classes de composés tels que les hydrocarbures, les systèmes CHO, les systèmes CHN, etc...

Les méthodes AM1 et PM3 [70] appartiennent aux dernières versions de la méthode MNDO.

Dans la méthode MNDO les paramètres associés aux intégrales bi-électroniques mono-centres sont basés sur des données spectroscopiques relatives aux atomes isolés et l'évaluation des autres intégrales bi-électroniques repose sur les interactions multipole-multipole de l'électrostatique classique. Dans cette méthode, des composés contenant H, Li, Be, B, C, N, O, F, Al, Si, Ge, Sn, Pb, P, S, Cl, Br, I, Zn, et Hg ont été paramétrés.

### III. Bases de modélisation moléculaires

L'hamiltonien associé aux électrons de valence est donné par :

$$\hat{H}_{\text{Val}} = \sum_{i=1}^{n(\text{val})} \left[ -\frac{1}{2} \nabla_i^2 + V(i) \right] + \sum_{i=1}^{n(\text{val})} \sum_{j \neq i} \frac{1}{r_{ij}} \quad (19)$$

qui se simplifie en :

$$\hat{H}_{\text{Val}} = \sum_{i=1}^{n(\text{val})} \hat{H}_{\text{val}}^c(i) + \sum_{i=1}^{n(\text{val})} \sum_{j \neq i} \frac{1}{r_{ij}} \quad (20)$$

$$\text{où : } \hat{H}_{\text{val}}^c(i) = \left[ -\frac{1}{2} \nabla_i^2 + V(i) \right] \quad (21)$$

(val) désigne le nombre d'électrons de valence du système,  $V(i)$  est l'énergie potentielle de l'électron  $i$  dans le champ des noyaux et des électrons de cœur,  $\hat{H}_{\text{val}}^c(i)$  est la contribution mono-électronique à  $\hat{H}_{\text{Val}}$ .

Les éléments de la matrice de Fock sont calculés à l'aide de l'équation :

$$F_{\text{val},rs} = H_{\text{val},rs}^c + \sum_{t=1}^b \sum_{u=1}^b P_{tu} \left[ (rs|tu) - \frac{1}{2} (ru|ts) \right] \quad (22)$$

Dans la méthode MNDO les éléments de la matrice de Fock peuvent être calculés comme suit.

Les éléments de la matrice de cœur (intégrale de résonance de cœur)  $H_{\mu_A \mu_B}^c = \langle \mu_A(1) | \hat{H}_{(1)}^c | \mu_B(1) \rangle$ , avec des orbitales atomiques centrées sur les atomes A et B sont donnés par :

$$H_{\mu_A \mu_B}^c = \frac{1}{2} \left( \beta_{\mu_A} + \beta_{\nu_B} \right) S_{\mu_A \nu_B} ; A \neq B \quad (23)$$

où les  $\beta$  sont les paramètres de chaque orbitale. Par exemple, le carbone avec les orbitales atomiques de valence 2s 2p, centrées sur le même atome de carbone, aura les paramètres  $\beta_{C2s}$  et  $\beta_{C2p}$ .

Les éléments de la matrice de cœur à partir d'orbitales atomiques différentes centrées sur le même atome sont fournis par l'équation (24) :  $H^c(1) = -\frac{1}{2} \nabla_1^2 + V(1)$ , où  $V(1)$  est l'énergie potentielle de l'électron de valence 1 dans le champ du cœur. Décomposant  $V(1)$  en contributions individuelles de cœurs atomiques, il vient :

$$H^c(1) = -\frac{1}{2} \nabla_1^2 + V_A(1) + \sum_{B \neq A} V_B(1) \quad (24)$$

Ainsi :

$$H_{\mu_A \nu_B}^c = \langle \mu_A | -\frac{1}{2} \nabla_1^2 + V_A | \nu_A \rangle + \sum_{B \neq A} \langle \mu_A | \nu_B | \nu_A \rangle \quad (25)$$

Des considérations de la théorie des groupes [71] permettent d'annuler  $\langle \mu_A | -\frac{1}{2} \nabla_1^2 + V_A | \nu_A \rangle$ , de telle sorte que :

$$H_{\mu_A \nu_B}^c = \sum_{B \neq A} \langle \mu_A | \nu_B | \nu_A \rangle \quad (26)$$

Si l'on considère que l'électron 1 interagit avec un point du cœur de charge  $C_B$ , alors :

$$V_B = -\frac{C_B}{r_{1B}} \quad (27)$$

$$\langle \mu_A | \nu_B | \nu_A \rangle = -C_B \langle \mu_A | \frac{1}{r_{1B}} | \nu_A \rangle \quad (28)$$

Dans la méthode MNDO,  $\langle \mu_A | \nu_B | \nu_A \rangle = -C_B \langle \mu_A \nu_A | s_B s_B \rangle$ , où  $s_B$  est l'orbitale de valence s centrée sur l'atome B :

$$H_{\mu_A \nu_B}^c = \sum_{B \neq A} \langle \mu_A | \nu_B | \nu_A \rangle = -\sum_{B \neq A} C_B \langle \mu_A \nu_A | s_B s_B \rangle; \mu_A \neq \nu_A \quad (29)$$

Les éléments de la matrice de cœur :  $H_{\mu_A \mu_A}^c = \langle \mu_A(1) | \hat{H}^c | \mu_A(1) \rangle$  sont calculés en utilisant la relation :

$$H_{\mu_A \mu_A}^c = \langle \mu_A | -\frac{1}{2} \nabla^2 + V_A | \nu_A \rangle + \sum_{B \neq A} \langle \mu_A | \nu_B | \nu_A \rangle = U_{\mu_A \mu_A}^c + \sum_{B \neq A} \langle \mu_A | \nu_B | \nu_A \rangle \quad (30)$$

$U_{\mu_A \mu_A}^c$  est évalué à partir de paramètres tirés de spectres atomiques (les paramètres utilisés pour l'atome de carbone :  $U_{ss}$  et  $U_{pp}$ ). Donc :

$$H_{\mu_A \nu_A}^c = U_{\mu_A \mu_A} + \sum_{B \neq A} C_B \langle \mu_A \nu_A | s_B s_B \rangle \quad (31)$$

L'évaluation de  $\langle \mu_A \nu_A | s_B s_B \rangle$  est réalisée comme suit :

- 1) Toutes les intégrales tri et tétra – centres sont annulées dans la méthode RDN.
- 2) Les intégrales de répulsion électroniques mono-centres sont soit des intégrales coulombiennes  $g_{\mu \nu} \langle \mu_A \mu_A | \nu_A \nu_A \rangle$ , soit des intégrales d'échange  $h_{\mu \nu} \langle \mu_A \nu_A | \mu_A \nu_A \rangle$ .

Pour l'atome de carbone, par exemple, les intégrales sont  $g_{ss}$ ,  $g_{sp}$ ,  $g_{pp}$ ,  $g_{pp'}$ ,  $h_{sp}$  et  $h_{pp'}$ , p et p' étant portées par des axes différents.

- 3) Les intégrales de répulsion bi-centres sont calculées à partir des valeurs d'une intégrale mono-centre et la distance inter - nucléaire en utilisant une procédure d'expansion multipole[72].

4) Le terme de répulsion cœur – cœur est donné par :

$$V_{CC} = \sum_{B>A} \sum_A [C_A C_B (s_A s_B / s_B s_B) + f_{AB}] \quad (32)$$

où :

$$f_{AB} = f_{AB}^{MNDO} = \left[ C_A C_B (s_A s_B / s_B s_B) \left( e^{-\alpha_A R_{AB}} + e^{-\alpha_B R_{AB}} \right) \right] \quad (33)$$

$\alpha_A$  et  $\alpha_B$  sont les paramètres des atomes A et B. Pour les paires O-H et N-H, par exemple, on aura :

$$f_{AH}^{MNDO} = \left[ (C_A C_H (s_A s_H) s_H s_H) \left( R_{AH} e^{-\alpha_A R_{AH}} + e^{-\alpha_H R_{AH}} \right) \right] \alpha_A \alpha_H \quad (34)$$

où A désigne soit N soit O.

Dans la méthode MNDO, les paramètres suivants doivent être optimisés :

- 1) Les intégrales mono-électroniques mono-centres  $U_{ss}$  et  $U_{pp}$ .
- 2) L'exposant  $\xi$  de la STO. Pour la MNDO  $\xi_s = \xi_p$ .
- 3)  $\beta_s$  et  $\beta_p$ . La méthode MNDO suppose que  $\beta_s = \beta_p$ .

Dans la méthode AM1,  $\xi_s \neq \xi_p$ .

Des composés comportant différents atomes (H, B, Al, C, Si, Ge, Sn, N, P, O, S, F, Cl, Br, I, Zn et Hg) ont été paramétrés dans AM1.

On a :

$$f_{AB}^{AM1} = f_{AB}^{MNDO} + \frac{C_A C_B}{R_{AB}} \left[ \sum_k a_{kA} \exp \left[ -b_{kA} (R_{AB} - C_{BA})^2 \right] \right] + \frac{C_A C_B}{R_{AB}} \left[ \sum_k a_{kB} \exp \left[ -b_{kB} (R_{AB} - C_{KB})^2 \right] \right] \quad (35)$$

Stewart a re-paramétré les valeurs pour générer la série PM. Celle qui dérive de AM1 est connue sous l'appellation PM3 (Parametric Method 3).

Dans la méthode PM3, les intégrales de répulsion mono-centres sont paramétrées par optimisation. La fonction de répulsion de cœur contient seulement deux fonctions gaussiennes par atome. Des composés comportant des atomes parmi : H, C, Si, Ge, Sn, Pb, N, P, As, Sb, Bi, O, S, Se, Te, F, Cl, Br, I, Al, Ga, In, Te, Be, Mg, Zn, Cd et Hg ont été paramétrés dans PM3.

#### III.3. Champ de force.

##### III.3.1 Définition :

La mécanique moléculaire est une méthode d'analyse conformationnelle basée sur l'utilisation de champs de forces empiriques et la minimisation d'énergie.

Dans un sens général, la mécanique moléculaire traite les atomes (ou les noyaux) d'une molécule comme des masses ou des sphères reliées par des ressorts de différentes forces représentant les liaisons.

Les interactions entre particules (de type atomique) sont traitées à l'aide de fonctions de potentiel tirées de la mécanique classique : fonctions de potentiel individuelles pour décrire les différents types d'interactions.

Les fonctions d'énergie potentielle comportent des paramètres empiriques décrivant des interactions entre des ensembles d'atomes. La paramétrisation est faite à partir de données expérimentales (RMN, RX, calculs *ab initio*) sur le plus grand ensemble possible de molécules. Le choix des données expérimentales est important et le modèle obtenu en dépend étroitement. Les constantes sont ajustées pour rendre l'expression de l'énergie potentielle,  $E$ , la plus générale possible.

Les fonctions de potentiel et les paramètres exploités pour l'évaluation des interactions sont désignés par 'champ de force'.

Une hypothèse importante est que le champ de force déterminé à partir d'un ensemble de molécules est transférable à d'autres molécules.

Notons qu'un ensemble de paramètres développés et testés sur un nombre relativement petit de cas est encore applicable à une plus large gamme de problèmes. En outre, les paramètres développés à partir de données relatives à de petites molécules peuvent être utilisés pour étudier des molécules plus grandes tels que les polymères.

##### III.3.2 Quelques exemples :

Parmi les champs disponibles nous citerons les plus répandus largement utilisés pour le traitement de petites molécules.

- **MM2, MM3 et MM4**, [http:// enropa. Chem. uga. edu/allinger/mm2 mm3 chtml](http://enropa.chem.uga.edu/allinger/mm2%20mm3.html) introduit par Allinger *et al* [73-76], largement utilisé pour le traitement de petites molécules.

- **AMBER**: [http:// amber. Scripps.edu](http://amber.scripps.edu)

(Assisted Method Building and Energy Refinement) introduit par Cornell *et al* [77], très largement utilisé dans le traitement des protéines et des acides nucléiques.

- **CHARMM:** <http://yuri.harvard.edu>

(Chemistry at Harvard Macromolecular Mechanics) développé par Mackerall, Karplus *et al* [75] qui est largement utilisé pour la simulation de petites molécules et jusqu'aux complexes solvatés de grandes macromolécules biologiques.

- **MMFF :** (Merck Molecular Force Field) développé par Halgren [78-80], il est similaire à MM3 dans la forme, mais en diffère par son application focalisée sur les processus de condensation de phases (en dynamique moléculaire). Il reproduit l'exactitude MM3 pour les petites molécules, et est applicable aux protéines et autres systèmes d'importance biologique.

Les différents champs de force se distinguent par trois aspects principaux :

- 1) La forme de la fonction de chaque terme énergétique.
- 2) Le nombre de termes croisés (qui reflètent le couplage entre coordonnées internes) inclus.
- 3) Le type d'information utilisé pour ajuster les paramètres.

#### III.3.3 Représentation simple d'un champ de force :

Beaucoup de champs de force utilisés actuellement pour la modélisation moléculaire peuvent s'interpréter en termes d'une représentation relativement simple à quatre composantes des forces intra et intermoléculaires internes au système considéré. Les pénalités énergétiques sont associées aux écarts des liaisons et des angles par rapport à leurs valeurs de 'référence' ou 'd'équilibre', il y a une fonction qui décrit la façon dont l'énergie change lors de la rotation des liaisons, et finalement le champ de force contient des termes décrivant l'interaction entre des parties non liées du système. Des champs de forces plus sophistiqués peuvent comporter des termes additionnels, mais ils présentent invariablement ces quatre composantes. Un fait intéressant lié à cette représentation est qu'on peut imputer les variations à des coordonnées internes spécifiques telles que les longueurs et les angles de liaisons, la rotation des liaisons ou les mouvements des atomes relativement les uns aux autres. Ce qui permet de comprendre facilement comment les changements des paramètres du champ de force affectent ses performances, et aident également dans le processus de paramétrisation.

Pour des molécules seules ou des ensembles d'atomes et / ou de molécules, un tel champ de force a la forme fonctionnelle suivante :

$$\begin{aligned}
 \mathbf{V}(\mathbf{r}^N) = & \sum_{\text{liaisons}} \frac{k_i}{2} (l_i - l_{i,0})^2 + \sum_{\text{angles}} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 + \sum_{\text{torsions}} (1 - \cos(n\omega - \gamma)) \\
 & + \sum_{i=1}^N \sum_{j=i+1}^N \left( 4 \varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) \quad (36)
 \end{aligned}$$

$\mathbf{V}(\mathbf{r}^N)$  représente l'énergie potentielle qui est fonction des positions ( $\mathbf{r}$ ) des N particules (habituellement les atomes)

Les diverses contributions sont représentées schématiquement sur la figure 4.

Le premier terme de l'équation (36) modèle l'interaction entre paires d'atomes liés, représentée dans ce cas par un potentiel harmonique qui fournit la variation de l'énergie lorsque la longueur de liaison  $l_i$  dévie de sa valeur de référence (à l'équilibre)  $l_{i,0}$ . Le second terme est une sommation sur tous les angles de valence de la molécule, encore modélisé par un potentiel harmonique (un angle de valence est l'angle formé par 3 atomes A-B-C où A et C sont tous deux liés à B). Le troisième terme dans l'équation (36) renseigne sur la variation d'énergie lorsqu'une liaison tourne. La quatrième contribution est le terme de non liaison, qui est calculé entre toutes les paires d'atomes (i et j) localisés sur différentes molécules ou appartenant à la même molécule mais séparés par au moins 3 liaisons (c'est – à – dire avec une relation 1, n où  $n \geq 4$ ). Dans un champ de force simple le terme de non liaison comprend un terme de potentiel coulombien pour les interactions électrostatiques et le potentiel de Lennard – Jones pour les interactions de van der Waals.

- Elongation des liaisons : MM2 a été paramétré pour ajuster les distances moyennes calculées à partir du mouvement vibrationnel à température ambiante à celles obtenues par diffraction électronique.

Pour mieux reproduire la courbe de Morse, MM2 fait intervenir un terme quadratique et un autre cubique :

$$v(l) = \frac{k}{2} (l - l_0)^2 \left[ 1 - k'(l - l_0) \right] \quad (37)$$

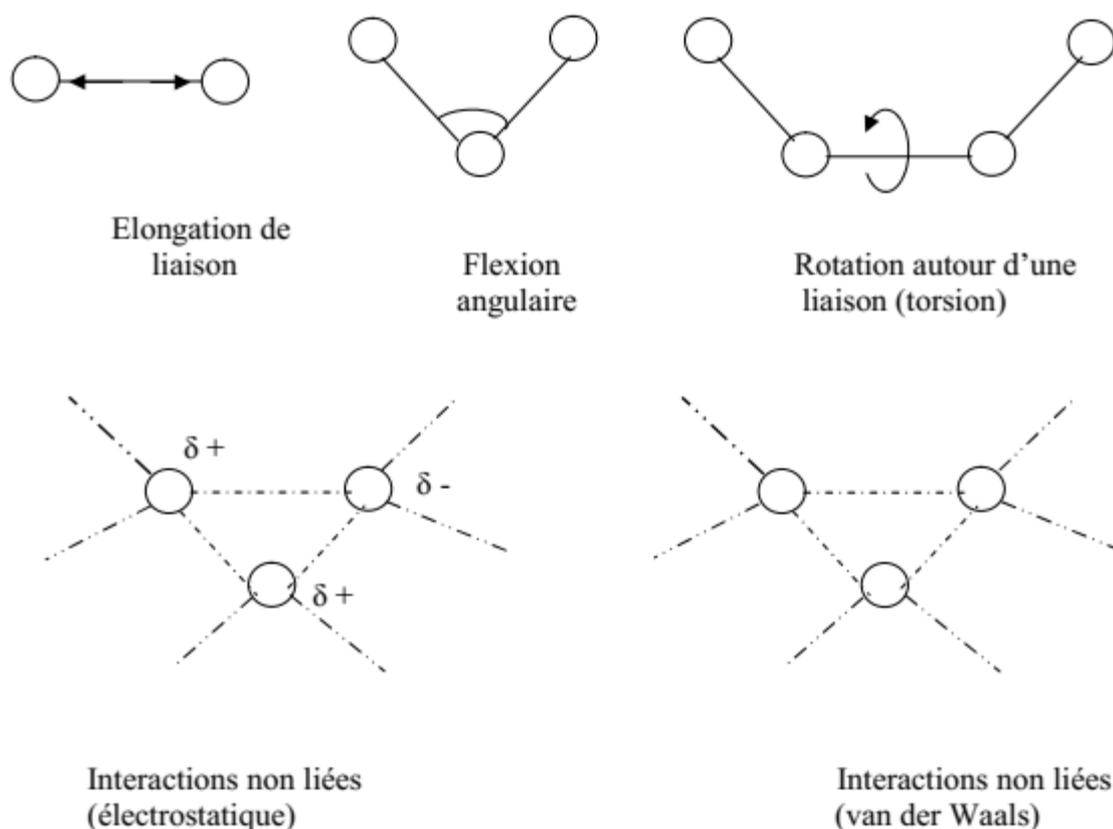
- Variation des angles : Les déviations des angles de leurs valeurs de référence sont souvent écrites en utilisant la loi de Hooke ou potentiel harmonique :

$$v(\theta) = \frac{k}{2} (\theta - \theta_0)^2 \quad (38)$$

Comme pour les termes d'élongation des liaisons, la précision du champ de force peut être augmentée par l'incorporation de termes d'ordres supérieurs. MM2 contient un terme d'ordre 4 en plus du terme quadratique.

$$v(\theta) = \frac{k}{2} (\theta - \theta_0)^2 \left[ 1 - k'(\theta - \theta_0)^2 \right] \quad (39)$$

### III. Bases de modélisation moléculaires



**Figure 4 :** Représentation schématique des quatre contributions à un champ de force de MM : élongation de liaison, flexion angulaire [53].

- Torsion des angles dièdres : correspond à la rotation d'une liaison selon l'angle dièdre  $\omega$  formé par 4 atomes. Le champ de force MM2 utilise 3 termes d'une série de Fourier :

$$v(\omega) = \frac{V_1}{2} (1 + \cos \omega) + \frac{V_2}{2} (1 - \cos 2\omega) + \frac{V_3}{3} (1 + \cos 3\omega) \quad (40)$$

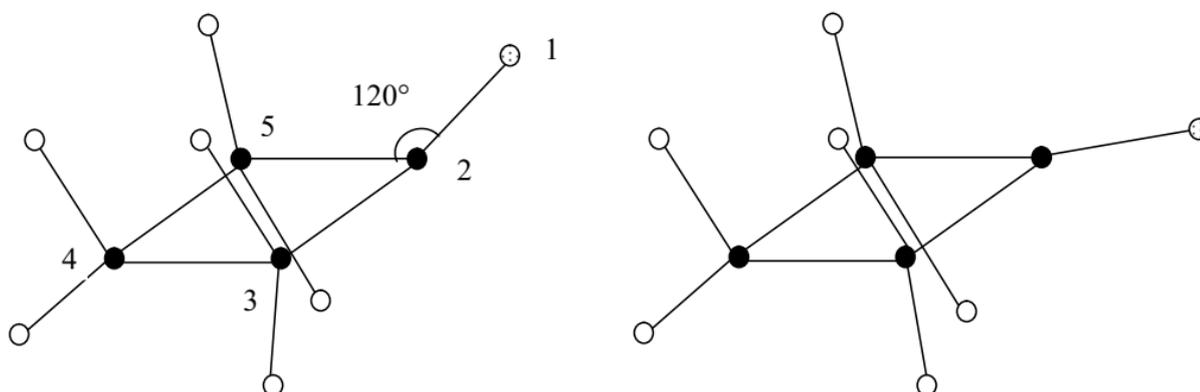
Une interprétation physique a été attribuée à chacun des 3 termes à partir de l'analyse de calculs *ab initio* effectués sur des hydrocarbures fluorés simples.

- Angle dièdre impropre ou déviation extra - planaire :

Examinons comment la cyclobutanone serait modélisée en utilisant uniquement un champ de force comportant des termes standards pour l'élongation des liaisons et les variations angulaires du type de ceux apparaissant dans l'équation (40). La structure d'équilibre obtenue avec un tel champ de force sera caractérisée par la localisation de l'atome d'oxygène hors du plan formé par l'atome de carbone contigu (à l'oxygène) et les 2 carbones qui lui sont liés (figure 5).

Dans cette configuration, les angles vers l'oxygène adoptent des valeurs proches de la valeur de référence  $120^\circ$ . Expérimentalement, il est établi que l'atome d'oxygène demeure dans le plan du cyclobutane bien que les angles C-C=O soient grands ( $133^\circ$ ). Ceci parce que l'énergie de liaison

$\pi$ , qui est maximalisée dans l'arrangement coplanaire, sera plus réduite si l'atome d'oxygène est dévié



**Figure 5:** Sous un terme extra - planaire, l'atome d'oxygène de la cyclobutanone est prévu de se situer hors du plan du cycle / gauche plutôt que dans le plan [53].

hors du plan. Pour réaliser la géométrie désirée il est nécessaire d'ajouter un (ou des) terme(s) additionnel(s) dans le champ de force qui maintienne(nt) le carbone  $sp^2$  et les 3 atomes qui lui sont liés dans le même plan. La plus simple façon de le réaliser est d'utiliser un terme de variation angulaire extra – planaire.

Il y a plusieurs façons d'incorporer des termes de variation extra – planaire dans un champ de force. Une approche possible est de traiter les 4 atomes comme un angle de torsion impropre pour lequel les 4 atomes (figure 6) ne sont pas liés dans la séquence 1 – 2 – 3 – 4. Une façon de définir, dans ce cas, une torsion impropre consiste à impliquer les atomes 1 – 5 – 3 – 2 de la figure 6.

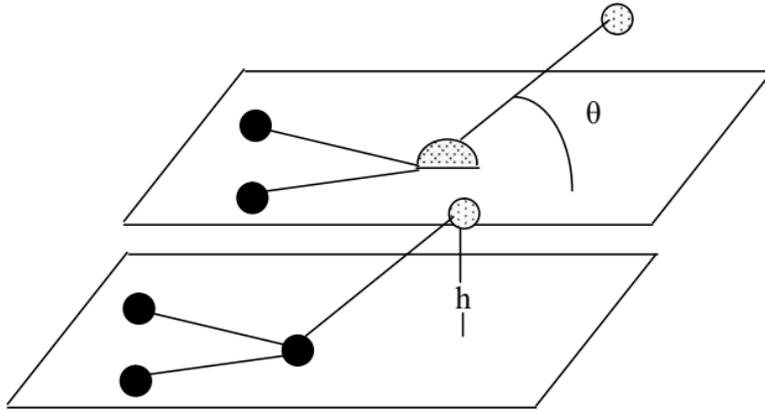
Un potentiel de torsion de la forme suivante :

$$v(\omega) = k (1 - \cos 2\omega) \quad (41)$$

peut être utilisé pour maintenir l'angle de rotation impropre à  $0^\circ$  ou  $180^\circ$ .

Il existe diverses autres façons pour inclure la contribution de la variation d'angle extra – planaire. Par exemple, une définition qui est la plus proche de la notion de « variation extra – planaire » implique le calcul de l'angle entre une liaison à partir de l'atome central et le plan défini par l'atome central et les 2 autres atomes (figure 6). La valeur  $0^\circ$  correspond aux 4 atomes coplanaires. Une troisième approche consiste à calculer la hauteur de l'atome central au – dessus du plan défini par les 3 autres atomes (figure 6). Avec ces 2 définitions la déviation de la coordonnée extra – planaire (angle ou distance) peut être modélisée à l'aide d'un potentiel harmonique de la forme :

$$v(\theta) = \frac{k}{2} \theta^2 \quad ; \quad v(h) = \frac{k}{2} h^2 \quad (42)$$



**Figure 6:** Deux façons pour modéliser les contributions de la variation d'angle extra – planaire [53].

- Termes de croisement : Les termes de croisement permettent de tenir compte des interactions entre l'élongation des liaisons, la variation des angles et la torsion des angles dièdres. Par exemple, si les liaisons C-O et O-H de l'angle C-O-H sont étirées, alors la distance entre les atomes terminaux (C et H) est augmentée, ce qui rend plus facile la diminution de l'angle COH. Pareillement, la diminution de l'angle COH tend à être accompagnée d'une augmentation des longueurs des liaisons O-H et C-O. Pour tenir compte de cette interaction, on peut ajouter un terme de croisement « élévation – variation angulaire ». (stretch - bend) de la forme :

$$v_{\Delta \theta} = \frac{1}{2} k_{12} (\Delta l_1 + \Delta l_2) \Delta \theta \quad (43)$$

avec  $\Delta l_1 = l_1 - l_{10}$  ;  $\Delta l_2 = l_2 - l_{20}$  et  $\Delta \theta = \theta - \theta_0$

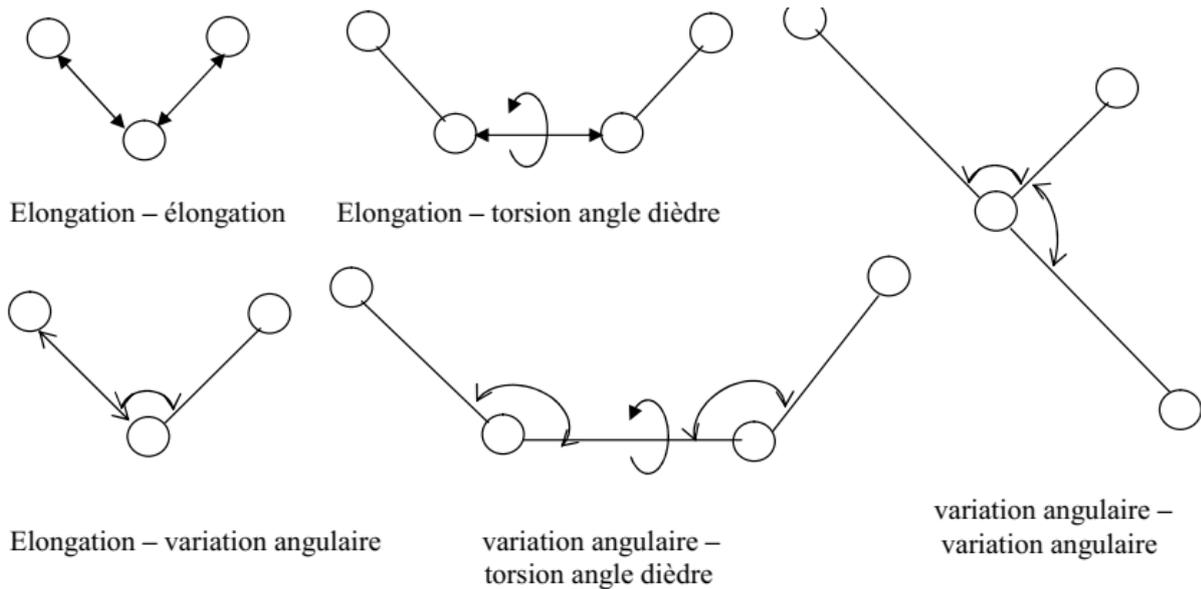
$l_{10}$ ,  $l_{20}$  et  $\theta_0$  représentent les valeurs de référence pour  $l_1$ ,  $l_2$  et  $\theta$  respectivement.

Les termes de croisement les plus utilisés sont (figure 7) :

- élévation – élévation et élévation – variation angulaire, pour deux liaisons à un même atome ;
- élévation – torsion angle dièdre, variation angulaire - torsion angle dièdre et variation angulaire - variation angulaire pour 2 angles avec un atome central commun.

Le champ de force MM2 fait intervenir uniquement un terme de croisement élévation – variation angulaire.

### III. Bases de modélisation moléculaires



**Figure 7:** Illustration schématique des termes de croisement supposés les plus importants pour les champs de forces [53].

- Interactions électrostatiques : Le terme électrostatique  $v_{es}$  est généralement défini par la somme des interactions électrostatiques entre toutes les paires d'atomes exceptées les paires 1, 2 et 1, 3 :

$$v_{es} = \sum_{1, \geq 4} v_{es, ij}, \text{ où les atomes } i, j \text{ vérifient la relation } (1, \geq 4v_{es, ij}).$$

$V_{es}$  est généralement calculé en affectant des charges atomiques partielles à chaque atome et en appliquant un potentiel coulombien.

MM2 n'utilise pas cette procédure mais associe un moment dipolaire avec chaque liaison puis calcule  $V_{es}$  comme somme des énergies potentielles d'interactions entre moments de liaisons. Chaque moment dipolaire étant localisé au centre, et dirigé le long de cette liaison. Les valeurs des moments de liaisons de certains types de liaisons ont été choisies pour ajuster les moments dipolaires expérimentaux de petites molécules.

L'équation (44) décrit un modèle de charge, basé essentiellement sur les dipôles centrés, utilisé dans MM2 [73].

$$v_{es} = \frac{\mu_i \mu_j}{k r^3} (\cos \chi - 3 \cos \alpha_i \cos \alpha_j) \quad (44)$$

$\chi$  et  $\alpha_i, \alpha_j$  désignent respectivement les angles entre les dipôles et les angles entre chaque dipôle et le vecteur de connexion.

- Interactions de van der Waals : La plupart des champs de force utilisent le potentiel 12 – 6 de Lennard Jones.

MM2 utilise le potentiel de Buckingham, avec un terme attractif proportionnel à  $r^{-6}$  et un terme répulsif proportionnel à  $e^{-\alpha r}$  où  $\alpha$  est un paramètre :

$$V_{vdw} = A e^{-\alpha r} - \frac{B}{r^6} \quad (45)$$

- Liaisons conjuguées : MM2 utilise une procédure générale qui consiste à effectuer des calculs semi-empiriques sur les électrons  $\pi$  pour en tirer les ordres de liaisons, qui sont ensuite utilisés pour affecter des longueurs de liaisons et des constantes de force de référence pour les liaisons conjuguées.

#### III.3.4 Champ de force MM+.

C'est une extension du Champ de force MM2, avec l'ajout de quelques paramètres additionnels. MM+ est un champ de force robuste, il a l'aptitude de prendre en considération les paramètres négligés dans d'autres champs de force et peut donc s'appliquer pour des molécules plus complexes tels que les composés inorganiques.