

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE MUSTAPHA STAMBOULI DE MASCARA

FACULTÉ DES SCIENCES EXACTES



Chimie Théorique et Modélisation

Présenté par :

Dr : SOUFI-LALOUT Wassila

Revu et complété avec la participation scientifique de :

Mr le professeur Siad GHALEM, Directeur de recherche et directeur de laboratoire des substances naturelles et bio-actives (LASNABIO) -

Equipe : Dynamique des biomolécules.

Année Universitaire : 2021-2022

*Ce polycopie est destiné aux étudiants de première année de Master Chimie Catalyse intitulé «**Chimie Théorique et modélisation**». Il aborde des lois de la chimie, de la physique et de la biologie dans des programmes informatiques spécifiques afin de calculer les structures et les propriétés d'entités chimiques et biochimiques (protéines, acides nucléiques, complexes moléculaires, solides, cristaux etc.) à l'aide des logiciels de la modélisation moléculaire.*

En particulier, nous avons mis l'accent sur l'intérêt du docking moléculaire pour lequel il est primordial de considérer la flexibilité du ligand. Nous avons également introduit la notion de la mécanique moléculaire qui est un outil utilisé pour mieux comprendre les petits systèmes moléculaires et pour répondre à certaines interrogations concernant la structure des molécules.

La notion de dynamique moléculaire a pour but de simuler le mouvement des atomes au cours du temps, à partir de l'énergie potentielle calculée par la mécanique moléculaire. Elle permet de calculer l'énergie cinétique.

Néanmoins, nous avons aussi rappelé les notions de base nécessaires pour pouvoir réaliser des manipulations sur ordinateur et décrit avec précision le protocole opératoire.

Aussi ce manuel est le fruit de plusieurs années d'enseignement et d'encadrement.

Il pourrait faire partie de la documentation de base dans le domaine de la modélisation moléculaire.

Ce polycopié est accompagné d'un CD décrivant des manipulations sur ordinateur.

Contenus pédagogiques

1 ère Année Master Catalyse

Semestre 1

Unité d'Enseignement Fondamentale

Chimie Théorique et Modélisation

Crédits : 4

Coefficients : 2

Contenu de la matière :

- 1- Introduction à la chimie théorique
- 2- Méthodes de la chimie quantique
 - Méthodes empiriques
 - Méthodes semi-empiriques
 - Méthodes non empiriques (*ab initio* et DFT)
 - Applications
- 3- Théories quantiques de la réactivité chimique
 - Théorie des orbitales moléculaires frontières
 - Théorie de l'état de transition
 - DFT conceptuelle et indices de réactivité
 - Applications
- 4- Méthodes de modélisation moléculaire
 - Mécanique moléculaire
 - Dynamique moléculaire
 - Recherche conformationnelle
 - Méthodes QSPR et QSAR
 - Applications
- 5- Introduction à la chimie combinatoire et ses applications

Sommaire

Chapitre I : Introduction à la chimie théorique

I- Introduction	1
II- Chronologie de la naissance de la mécanique quantique	2
II- 1-Hypothèse de Max Planck	2
II- 2- Hypothèse Albert Einstein	2
II-3- Hypothèse Niels Bohr	3
II-4- Hypothèse de Broglie	3
II-5- Equation de Schrödinger	4
II-6- Principe d'incertitude d'Heisenberg	6

Chapitre II : Méthodes de la chimie quantique

II-1. Introduction	7
II-2. Méthodes de calculs	7
a- Méthodes empiriques	7
a-1. Méthode de Hückel simple	8
a-2. Méthode de Hückel Etendue (EHT)	8
b - Méthodes semi-empiriques	9
c - Méthodes non-empiriques (complètes)	11
c-1. Approximation du champ moyen	11
c-2. Approximation OM-CLOA	11

Chapitre III : Méthodes de modélisation moléculaire

III-1. Généralités	14
III-2. La modélisation moléculaire	15

III-3. Les objectifs de la modélisation moléculaire	17
III-4. Notion de la structure moléculaire	18
III-5. Importance de la simulation moléculaire	19
III-6. Représentation empirique de l'énergie Moléculaire	21
III-7. La mécanique moléculaire	21
III-7-1. Principes théoriques	21
III-7-2. Modélisation par champs de force	23
III-7-3. Types de champ de force	26
III-7-4. Minimisation d'énergie	29
a- Minimum local , minimum global	30
b- Minima locaux	30
c- Limitation du champ de force	30
III-8. Les simulations de dynamique moléculaire	31
III-8-1. Principes théoriques	31
III-8-2. Calcul de dynamique moléculaire	31
III-8-3. Recherche conformationnelle à l'aide de la dynamique moléculaire	32
III-9. Docking Moléculaire	32
III-9-1. Approche combinatoire	33
III-9-2. Approche stochastique	33
III-9-3. Approche déterministe	34
III-10. La structure des protéines	34
a- La résolution	35
b- Le facteur R	36
c- Choix d'une structure cristallographique	36
d- Facteur d'agitation thermique	37
e- Détermination du site actif	37
III-11. Protocole général de docking	38
III-12. Docking rigide vers l'introduction de la flexibilité	39
III-13. Un processus de Docking Moléculaire	40
III-14. Programmes de docking moléculaire	40

III-15. Algorithmes de docking	41
III-16. Fonctions de score	42
a- Fonctions de score basées sur un champ de force	44
b- Fonctions de score empiriques	44
c- Les fonctions de score type knowledge-based	45
d- Fonctions de score consensus	45
III-17. Interaction Ligand (Peptide) – Récepteur	46
III-18. La pharmacocinétique - ADME - Tox	47
a. Absorption	49
b. Distribution	49
c. Métabolisme	50
d. Elimination	50
e. Toxicité	50
Références	52

Chapitre I :

*Introduction à la chimie
théorique*

I- INTRODUCTION

On peut définir la chimie théorique comme la description de la chimie à l'aide d'outils mathématiques. En particulier, l'application de la mécanique quantique à la chimie a donné naissance à la chimie quantique.

La chimie quantique consiste à appliquer le formalisme de la mécanique quantique aux systèmes chimiques (molécules, atomes, réactions). Cette branche de la chimie consiste à adopter l'équation de Schrödinger pour décrire un système moléculaire et à déterminer, pour cette équation, des solutions physiquement acceptables.

A partir des années 1950, l'élaboration de nouvelles méthodes quantiques et les progrès accomplis dans la technologie des ordinateurs (surtout après l'apparition des ordinateurs dans les années 1980) rendaient les calculs quantiques accessibles à tous. Aujourd'hui, le chimiste même s'il n'est pas spécialiste peut prévoir les structures et les réactivités moléculaires en utilisant les logiciels commerciaux de chimie quantique comme : HyperChem, Spartan, Mopac, ...etc.

Les concepts quantiques ont imprégnés donc toute la chimie et ils deviennent des outils de base pour la compréhension des phénomènes chimiques à l'échelle moléculaire. Les méthodes de la chimie quantique (semi-empiriques, *ab-initio*, DFT,...) sont actuellement utilisées non seulement pour étudier des molécules à l'état isolé mais aussi et surtout pour étudier l'interaction entre deux (ou plusieurs) molécules réagissant entre elles.

Les programmes de chimie quantique permettent également de tracer les profils énergétiques et d'élucider les mécanismes réactionnels pour différentes classes de réactions chimiques soit en phase gazeuse soit en présence d'un solvant (en solution). Les algorithmes d'optimisation intégrés dans ces programmes permettent également de localiser les états de transition, chose qui est inaccessible par voie expérimentale.

Autrement c'est méthode mathématique et physique qui décrit le comportement d'un système et de son évolution (géométrie énergétique et électronique) dans le temps " t " et l'espace (x,y,z).

Cette théorie a été découverte quand les chercheurs ont voulu étudier la matière profonde (atome) et son interaction avec l'optique et le rayonnement c'est entre 1900-1927.

II- Chronologie de la naissance de la mécanique quantique

II- 1. Hypothèse de Max Planck (1900) En étudiant le corps noir absolu, Max Planck proposa que la fréquence de la lumière émise par un corps noir. La fréquence de la lumière est proportionnelle à l'énergie absorbée. L'énergie radiante émise s'effectue par des quantités finies appelées quanta. La lumière est composée de grains appelés photons.

**" L'échange d'énergie entre la matière et le rayonnement s'effectue par portion
"quantum" "**

$$E = h \cdot \nu \text{ avec } h = 6,625 \cdot 10^{-34} \text{ J.S}$$

en autre terme : Lors de l'interaction matière - rayonnement seuls, les énergies suivantes peuvent être échangées: $\epsilon, 2\epsilon, 3\epsilon, \dots$ Tq: $\epsilon = h \cdot \nu$

ϵ : (morceau d'énergie) énergie d'un quantum

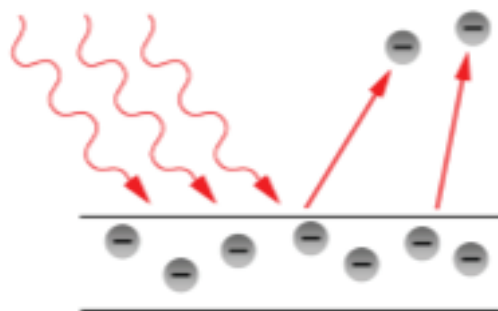
"L'échange de l'énergie est quantifié "

II- 2. Hypothèse Albert Einstein (1905)

Dans le but d'interpréter l'expérience de l'effet photoélectrique à Einstein a énoncé l'hypothèse suivante:

" La lumière est constituée de corpuscule d'énergie $h \cdot \nu$ appelé photon "

Autrement dit **" La lumière est quantifiée "**



II-3. Hypothèse Niels Bohr (1912)

En 1912, Bohr fait la corrélation entre l'énergie et le nombre d'onde.

$$E = h\nu = h \cdot \frac{c}{\lambda}$$

$$= hc \left[\frac{R_H}{n^2} - \frac{R_H}{p^2} \right]$$

Pour interpréter ce phénomène Bohr a introduit les postulats :

Il y a un niveau d'énergie où l'électron ne peut plus rayonner de l'énergie (il y a ni absorption ni émission lorsque l'électron se déplace sur des orbites bien déterminé appelé orbites stationnaire).

- ✓ L'énergie de l'atome subit une variation brusque l'électron passe d'un état à un autre, (sous l'effet d'une absorption ou une perte d'énergie).

$$\Delta E = h\nu = |E_2 - E_1|$$

e : L'électron ne peut prendre que certains valeurs permises

n : ne peut prendre que des valeurs entières

- ✓ Le moment cinétique orbital est quantifié.

$$mvr = n\hbar$$

II- 4. Hypothèse de Broglie (1924)

Louis de Broglie établit une hypothèse : toute particule (corpuscule) en mouvement (électron, neutron, atome) de masse **m** et de vitesse **v** on peut associer une longueur d'onde $\lambda = h/p$

h : est la constante de Planck

p : quantité de mouvement ($p = mv$)

$$E = h\nu = h \cdot \frac{c}{\lambda} \quad (1)$$

$$\Delta E = mc^2 \quad (2)$$

(1) Et (2) $\implies \lambda = h/mc$ (pour la lumière)

II- 5. Equation de Schrödinger (Erwin Schrödinger 1925-1926)

Les équations de la chimie quantique sont basées sur la résolution de l'équation de SCHRÖDINGER qui s'écrit pour les états stationnaires, d'un système moléculaire ou les énergies qui y sont associées à un opérateur Hamiltonien et à leur fonction d'onde.

$$\mathbf{H} \Psi = \mathbf{E} \Psi \quad (1)$$

Où: Ψ : Fonction d'onde de la molécule.

\mathbf{H} : Hamiltonien total d'une molécule comportant \mathbf{N} noyaux et \mathbf{n} électrons, est défini par la somme de cinq termes (terme cinétique des électrons, terme cinétique des noyaux, terme de répulsion entre électrons, terme de répulsion entre noyaux, et le terme d'attraction électrons-noyaux).

\mathbf{E} : Energie totale d'une molécule.

L'hamiltonien d'un système à n électrons et N noyaux est donné par :

$$H_{\text{tot}} = T_E + T_N + V_{EN} + V_{EE} + V_{NN}$$

$$H = -\frac{\hbar^2}{2m_e} \sum_i^n \Delta_i - \frac{\hbar^2}{2M_K} \sum_K^N \Delta_K + \sum_{i>j}^n \frac{e^2}{r_{ij}} + \sum_{K>L}^N \frac{Z_K Z_L e^2}{r_{KL}} - \sum_{K=1}^N \sum_{i=1}^n \frac{Z_K e^2}{R_{Ki}}$$

Avec: T_E : l'énergie cinétique des n électrons.

T_N : l'énergie cinétique des N noyaux.

V_{EN} : l'énergie d'attraction électrons- noyaux.

V_{EE} : l'énergie de répulsion électron- électron.

V_{NN} : l'énergie de répulsion noyau- noyau.

La résolution exacte de cette équation n'est possible que pour l'atome d'hydrogène et les systèmes mono-électroniques. Pour les systèmes poly électroniques, on fait appel aux méthodes d'approximation, pour la résolution approchée de l'équation de SCHRÖDINGER.

En 1927, Born Oppenheimer est proposé de simplifier la résolution de l'équation de SCHRÖDINGER en séparant la partie électronique de la partie nucléaire dans la fonction d'onde Ψ . D'après l'approximation de Born Oppenheimer qui suppose que les

noyaux sont fixes, l'énergie cinétique des noyaux s'annule et l'énergie de répulsion entre noyaux se ramène à une constante.

Dans le cadre de cette approximation l'hamiltonien H peut se réduire à la forme suivante :

$$\mathbf{H}_{\text{tot}} = \mathbf{T}_E + \mathbf{V}_{EN} + \mathbf{V}_{EE} + \mathbf{V}_{NN}$$

$$H = -\frac{\hbar^2}{2m_e} \sum_{i=1}^n \Delta_i - \sum_{K=1}^N \sum_{i=1}^n \frac{Z_K e^2}{R_{Ki}} + \sum_{i>j}^n \frac{e^2}{r_{ij}}$$

$$\mathbf{H}_{\text{tot}} = \mathbf{H}_{\text{électronique}} + \mathbf{H}_{\text{nucléaire}}$$

RQ: En chimie structurale, on s'intéresse aux **é** qui sont responsables à la formation de liaisons chimiques conduisant à une structure définie dans l'espace

$$\text{Avec : } H_{\text{électronique}} = T_E + T_{EN} + V_{EE}$$

$$H_{\text{nucléaire}} = V_{NN}$$

Donc le problème réside dans le calcul de l'énergie électronique et l'équation à résoudre est de la forme :

$$\mathbf{H}_{\text{élec}} \Psi_{\text{élec}} = \mathbf{E}_{\text{élec}} \Psi_{\text{élec}}$$

(les noyaux représentent les cœurs des molécules et les électrons représentent l'extérieures d'une molécule.)

Les propriétés moléculaires qui peuvent être calculées par la résolution de l'équation de Schrödinger sont multiples. On peut citer entre autres :

- Structures et énergies moléculaires
- Energies et structures des états de transition
- Fréquences de vibration, spectres IR et Raman
- Propriétés thermochimiques
- Energies de liaison
- Chemins réactionnels
- Orbitales moléculaires
- Charges atomiques
- Moments multipolaires
- Déplacements chimiques RMN et susceptibilités magnétiques
- Affinités électroniques et potentiels d'ionisation

- Potentiels électrostatiques et densités électroniques etc.

II- 6. Principe d'incertitude d'Heisenberg (1927)

« Il est impossible de déterminer simultanément la quantité de mouvement et la position exacte d'un corpuscule (particule). »

$$\Delta x \cdot \Delta p_x \geq \hbar$$

Δx : incertitude sur la position

$\Delta p_x = m \Delta v_x$: incertitude sur la quantité de mouvement

$$m \Delta v_x \cdot \Delta x \geq \hbar/2\pi$$

En 1927, Born Oppenheimer est proposé de simplifier la résolution de l'équation de SCHRÖDINGER en séparant la partie électronique de la partie nucléaire dans la fonction d'onde Ψ . D'après l'approximation de Born Oppenheimer qui suppose que les noyaux sont fixes, l'énergie cinétique des noyaux s'annule et l'énergie de répulsion entre noyaux se ramène à une constante.

Dans le cadre de cette approximation l'hamiltonien H peut se réduire à la forme suivante :

$$\mathbf{H}_{\text{tot}} = \mathbf{T}_E + \mathbf{V}_{EN} + \mathbf{V}_{EE} + \mathbf{V}_{NN}$$

$$H = -\frac{\hbar^2}{2m_e} \sum_{i=1}^n \Delta_i - \sum_{K=1}^N \sum_{i=1}^n \frac{Z_K e^2}{R_{Ki}} + \sum_{i>j}^n \frac{e^2}{r_{ij}}$$

$$\mathbf{H}_{\text{tot}} = \mathbf{H}_{\text{électronique}} + \mathbf{H}_{\text{nucléaire}}$$

$$\text{Avec : } \mathbf{H}_{\text{électronique}} = \mathbf{T}_E + \mathbf{T}_{EN} + \mathbf{V}_{EE}$$

$$\mathbf{H}_{\text{nucléaire}} = \mathbf{V}_{NN}$$

Donc le problème réside dans le calcul de l'énergie électronique et l'équation à résoudre est de la forme :

$$\mathbf{H}_{\text{élec}} \Psi_{\text{élec}} = \mathbf{E}_{\text{élec}} \Psi_{\text{élec}}$$

Chapitre II :

*Méthodes de la chimie
quantique*

II-1. Introduction

Les méthodes de la mécanique quantique, qui font appel à la distribution des électrons répartis en orbitales autour de la molécule, impliquent des temps de calcul souvent élevés qui limitent leur usage à des petites molécules ou nécessitent le recours à de nombreuses approximations. Elles sont particulièrement adaptées au calcul des charges et des potentiels électrostatiques, à l'approche des mécanismes réactionnels ou à la polarisabilité. L'objectif de la mécanique quantique est principalement de déterminer l'énergie et la distribution électronique.

Les concepts quantiques ont imprégnés donc toute la chimie et ils deviennent des outils de base pour la compréhension des phénomènes chimiques à l'échelle moléculaire. Les méthodes de la chimie quantique (semi-empiriques, *ab-initio*, DFT,...) sont actuellement utilisées non seulement pour étudier des molécules à l'état isolé mais aussi et surtout pour étudier l'interaction entre deux (ou plusieurs) molécules réagissant entre elles.

Chimie quantique:

- on traite que des électrons mais pas des noyaux, donc on regarde la structure électronique et les noyaux restent fixes.
- les électrons sont localisés sur des liaisons.

II-2. Méthodes de calculs

a- Méthodes empiriques

Les méthodes empiriques sont des méthodes de mécanique moléculaire basées sur des concepts de la mécanique classique où les atomes et leurs électrons sont confondus en une collection de points matériels. Ces derniers agissent les uns sur les autres par l'intermédiaire d'un potentiel empirique encore dit champ de forces ne dépendant que de la position relative des atomes dans l'espace.

Une méthode est dit empirique si l'hamiltonien H total peut s'écrire comme la somme de l'hamiltonien monoélectronique si:

$$H(1,2,3,\dots,n) = \sum h(i) \text{ alors } E(1,2,3,\dots,n) = \sum e(i)$$

$$\psi(1,2,3,\dots,n) = \prod \psi(i)$$

Les méthodes empiriques, comme, par exemple, « Hückel Simple » et Théorie de Hückel étendue (Extended Hückel Theory (EHT)) réduisent considérablement les temps de calcul en ne considérons que les parties "nécessaires" ou "intéressantes". Ces méthodes utilisent :

- Beaucoup d'approximations
- Faciles et simples
- Pas d'intégrale à calculer (toutes les intégrales sont remplacées par des paramètres empiriques).
- Très rapides

a-1. Méthode de Huckel simple

Cette méthode à été proposé par Huckel en 1931, elle permet de traiter les systèmes conjugués π c.à.d. que les atomes hybridés sp^2 . Elle est basé sur les approximation :

$$\sum C_{kr} (h_{rs} - e_k S_{rs}) = 0 \quad s = 1, 2, \dots, n$$

$$S_{rs} = \begin{cases} 1 & \text{si } r = s \\ 0 & \text{si } r \neq s \text{ (néglige le recouvrement)} \end{cases}$$

$$H_{rs} = \begin{cases} \alpha & \text{si } r = s \\ \beta & \text{si } r \neq s \text{ (r et s sont chimiquement liés)} \\ 0 & \text{si } r \neq s \text{ (r et s ne sont pas chimiquement liés)} \end{cases}$$

a-2. Méthode de Huckel Etendue (EHT)

La méthode de Huckel étendue est une méthode quanta-chimique qui a été proposé par Hoffmann en 1963, dans le but de traiter tous les électrons de valence porté par toutes les OAs de valence c.à.d. elle traite toutes les liaisons σ et π qui forment une structure organique. Elle est basé sur les approximation :

$$S_{rs} = \begin{cases} 1 & \text{si } r = s \\ S_{rs} > = < 0 & \text{si } r \neq s \text{ (tient compte au recouvrement)} \end{cases}$$

$$H_{rs} = \begin{cases} -I_r & \text{si } r = s \\ \frac{1}{2} K (H_{rr} + H_{ss}) S_{rs} & \text{si } r \neq s \text{ avec } K=1.75 \text{ (d'après Hoffmann)} \end{cases}$$

b - Méthodes semi-empiriques

Une méthode semi empirique est une méthode dans laquelle une partie des calculs nécessaires aux calculs des intégrales et les intégrales qui restent ne sont pas calculées mais elles sont remplacées par des paramètres ajustés sur l'expérience .

- on regarde seulement les électrons de valence.
- avec un Hamiltonien plus simple.
- Choisir une base minimum
- Ces méthodes sont des méthodes incomplètes.
- On ne prend que les électrons et les orbitales atomiques des couches de valences pour former la base, les électrons des couches internes sont inclus dans le cœur nucléaire.
- Dans les méthodes semi empirique on parle des répulsions cœur-cœur au lieu des répulsions noyau-noyau.
- Utiliser des valeurs paramétrées pour les intégrales bi-électroniques restantes
- Un grand nombre d'intégrales bi-électroniques sont négligées (celles à 3 et 4 centres dont la valeur est souvent voisine de zéro).

Parmi les méthodes les plus connues sont dénommées : CNDO/2, INDO, MNDO, elles sont toutes en commun un point essentiel est de négliger un certain nombre de termes bi-électroniques contenus dans l'opérateur de Fock.

Méthode CNDO/2 (Complete Neglecting of Differential Overlap / version 2): Cette méthode a été proposée par *Pople Santry* et *Segal* en 1965-1966. Elle est basée sur l'approximation valentielle, toutes les intégrales possédant un recouvrement différentiel, sont négligées.

Méthode INDO (Intermediate Neglecting of differential Overlap) 1967. Elle permet de distinguer entre les états singlets et les états triplets d'un système en conservant les intégrales d'échange.

Méthode MNDO (Modified Neglect of Diatomic Differential Overlap): Proposée par Dewar en 1977. MNDO ne considérait que les atomes de carbone, d'hydrogène, d'oxygène et d'azote. ne peut pas être appliquée dans des molécules contenant un métal. (Un des désavantages de cette méthodologie est qu'elle décrit mal les liaisons hydrogènes, Cette méthode ne traite pas les métaux de transition et présente des difficultés pour les systèmes conjugués).

Méthode NDDO (Neglect of Diatomic Differential Overlap): Proposée par Pople en 1965, qui consiste à négliger le recouvrement différentiel entre orbitales atomiques sur des atomes différents. Toutes les intégrales bi-électroniques bicentrées sont retenues (nulles).

Méthode AM1 (Austin Model 1): proposée par Dewar en 1985. Elle corrige le terme de répulsion cœur- cœur (le problème de la représentation des liaisons hydrogènes, est la plus utilisée et donne souvent de très bons résultats pour un temps de calcul très réduit par rapport aux méthodes *ab initio*).

Méthode PM3 (Parametric Method 3) : proposée par Stewart en 1989. Elle utilise une procédure de paramétrisation automatique au cours des calculs. Dans cette méthodologie, la paramétrisation atomique a été effectuée par groupe d'élément. Deux fonctions Gaussiennes par atome sont utilisées pour le calcul de la répulsion cœur-cœur. De plus, des paramètres pour les éléments du groupe d font partie de cette méthode.

L'avantage des calculs semi-empiriques est qu'ils sont donc plus rapides que les autres méthodes quantiques. Leur inconvénient réside dans le fait qu'elles sont soumises à de nombreuses approximations. En fait, peu de propriétés peuvent être prédites de manière fiable, en particulier, pour des molécules de structures trop éloignées de celles utilisées pour la paramétrisation des méthodes. En général celle-ci est réalisée afin de reproduire différentes propriétés : géométries, énergies de formation ou encore énergies de réaction, moments dipolaires, potentiels d'ionisation voire des propriétés spécifiques telles que des spectres électroniques ou des déplacements chimiques RMN . Le calcul semi-empirique est la technique le plus utilisée, notamment pour l'étude de molécules de taille intermédiaire.

c - Méthodes non-empiriques (complètes)

Dans les méthodes non-empiriques toutes les intégrales sont calculées d'une manière exacte. Soit une molécule comporter N noyaux et n électrons

$$\mathbf{H}_{\text{tot}} = \mathbf{H}_E + \mathbf{H}_N$$

H_E: Hamiltonien électronique

H_N: Hamiltonien nucléaire

$$H_{\text{tot}} = T_E + V_{EN} + V_{EE} + T_N + V_{NN}$$

La résolution de l'équation de Schrödinger reste toujours impossible à résoudre car la fonction d'état qui représente le mouvement de toutes les particules de systèmes indéfinis. De plus le terme (e^2/r_{ij}) est infini pour cela la mécanique quantique a introduit différentes approximations :

c-1. Approximation du champ moyen (Approximation monoélectronique)

L'approximation du champ moyen, proposée par Hartree en 1927, le principe de cette approximation c'est considérer que chaque électron du système est porté en interaction avec un champ électrostatique donné par le reste des électrons des systèmes.

Ce champ électrostatique est représenté par un potentiel du champ moyen donné par l'expression suivante :

c-2. Approximation OM-CLOA (méthodes *ab initio*)

Les expressions analytiques des orbitales moléculaires ϕ_i n'ont pas été définies dans le cadre de la méthode de Hartree-Fock. C'est Roothaan qui a utilisé la technique OM-CLOA pour construire les OM. Cette méthode consiste à exprimer l'orbitale moléculaire Ψ_i par une combinaison linéaire d'orbitales atomiques :

$$\Psi_k(\mathbf{i}) = \sum C_{kr} \rho_r$$

C_{kr} sont les coefficients à faire varier.

N étant le nombre d'OA pris dans CLOA.

ρ_r OA r porte l'électron (i)

Les meilleurs coefficients sont ceux qui minimisent l'énergie. En procédant par la méthode des variations et après certaines manipulations algébriques, on aboutit aux équations de Roothaan définies par le système séculaire suivant :

$$\sum_{r=1} C_{ir} (F_{rs} - \epsilon_i S_{rs}) = 0 \quad s = 1, 2, 3, \dots, n$$

F_{rs} = élément de la matrice de Hartree-Fock

S_{rs} = intégrale de recouvrement

Les méthodes *ab initio* sont des méthodes non empiriques, toutes les intégrales sont calculées rigoureusement et il n'y a pas d'approximations à faire sauf celle de Born Oppenheimer et l'approximation OM- CLOA.

Dans les méthodes *ab-initio*, toutes les particules (noyaux et électrons) sont traitées explicitement. On n'utilise aucun paramètre empirique dans le calcul de l'énergie.

- * Ces méthodes sont complètes

- * Calcul de toutes les intégrales.

- * Temps de calcul est très élevé

- * Les méthodes *ab initio* sont possibles que pour des molécules moyennes et petites taille.

- * Cette méthode permet d'obtenir des résultats fiables et précis pour des molécules de petite taille.

Cependant il est également possible d'obtenir de bons résultats à partir de méthodes quantiques moins exigeantes en moyens informatiques : les méthodes basées sur la théorie de la fonctionnelle de la densité (DFT) ont atteint un niveau de fiabilité compétitif avec des méthodes *ab-initio* incluant la corrélation électronique, tout en permettant de réduire considérablement la charge de calculs.

La méthode de la fonctionnelle de densité (DFT) calcule l'énergie du système à partir de la densité et non plus des orbitales moléculaires. Cette dernière méthode demande moins de calculs pour des résultats similaires.

L'idée directrice de la théorie de la fonctionnelle de la densité (DFT) est que l'énergie d'un système d'électrons peut être écrite en terme de la densité de probabilité électronique à un électron ρ . Pour un système de n électrons $\rho(\mathbf{r})$ correspond à la densité électronique au point de l'espace \mathbf{r} . L'énergie électronique $E[\rho]$, est une fonctionnelle de la densité électronique car à chaque fonction $\rho(\mathbf{r})$ est associée une seule énergie.

$$E = G[\rho(\mathbf{r})]$$

Chapitre III:

Méthodes de la Modélisation

Moléculaire

III-1. Généralités

Les méthodes théoriques utilisées pour l'obtention de modèles permettant de comprendre et de prédire la structure, les propriétés physico-chimiques et les interactions moléculaires sont connues sous le nom de « Modélisation Moléculaire ». Ces méthodes permettent de fournir des informations complémentaires à celles obtenues par des méthodes expérimentales.

La modélisation par ordinateur d'une molécule implique généralement une présentation graphique de la géométrie ou de la configuration des atomes de la molécule suivie de l'application d'une méthode théorique, comme la mécanique quantique semi-empirique, la mécanique quantique *ab initio*, ou encore la mécanique moléculaire, pour évaluer les propriétés physiques de la molécule à l'étude.

Avec les molécules de taille intermédiaire il faut aussi tenir compte quelles sont plus flexibles que les petites molécules et par conséquent posséderont plusieurs états conformationnels de population variable à la température ambiante ou physiologique pour les molécules d'intérêt biologique. De plus, il est généralement accepté que les interactions intermoléculaires peuvent aussi dépendre des états conformationnels de la molécule et de l'environnement.

La recherche et la synthèse de nouveaux composés chimiques sont aujourd'hui souvent associées à une étude par modélisation moléculaire. La modélisation moléculaire est une technique permettant, non seulement de représenter les propriétés et les réactions chimiques mais aussi de manipuler les modèles des structures en deux ou trois dimensions (figure 1).

La modélisation moléculaire est un ensemble de techniques bioinformatiques et biophysiques qui permettent de représenter et de visualiser la forme des molécules dans l'espace, et de simuler leurs mouvements et leurs comportements.

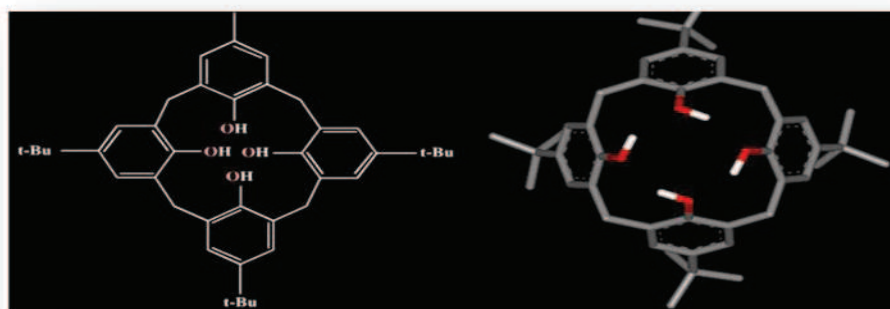


Figure 1 : Structure en deux et trois dimensions

La modélisation moléculaire est le terme générique regroupant les méthodes qui permettent de simuler le comportement d'un système de particules, elle exploite des lois de la chimie, de la physique et de la biologie dans des programmes informatiques spécifiques afin de calculer structures et propriétés d'entités chimiques et biochimiques (protéines, acides nucléiques, complexes moléculaires, solides, cristaux etc).

III-2. La modélisation moléculaire

La modélisation moléculaire met en jeu des méthodes de calcul théorique dont les principales font appel à la mécanique moléculaire, la dynamique moléculaire et à la mécanique quantique afin de déterminer la géométrie des atomes d'une molécule ainsi que ses propriétés physicochimiques.

La modélisation moléculaire basée sur les principes de la mécanique classique de Newton, l'application de la mécanique moléculaire permet de représenter un système à l'échelle atomique et d'en calculer l'énergie. Réaliser ce calcul pour différentes conformations de la structure considérée permet de caractériser différents points de la surface d'énergie potentielle associée au système. Chaque point de la surface est donc associé à une probabilité d'échantillonnage, ce qui induit la création d'une statistique de rencontre des conformations du système.

Cette méthode de calcul se base sur la sommation de différents termes et est très utilisée pour la description de systèmes de taille importante, excédant le millier d'atomes.

La modélisation et la simulation moléculaire permet de donner des représentations visuelles étonnantes des molécules (Drug-design), et de calculer les valeurs de certaines propriétés physico-chimiques, de déterminer précisément sa géométrie et de prévoir ses propriétés pharmacologiques ou toxiques.

Cette méthode a été développée dans le cas de cibles biologiques et plus particulièrement d'enzymes, de récepteurs et des protéines cibles, dont la structure et la conformation (structures tertiaire et quaternaire) ont pu être établies par diverses méthodes appropriées (cristallographie aux rayons X, résonance magnétique nucléaire (RMN) ...).

Ces structures sont représentées sur l'écran d'un ordinateur et, après avoir localisé le site actif de l'enzyme ou de la protéine, la possibilité que présentent certaines

substances (ligands) de s'y insérer est étudiée. De nombreux logiciels actuellement disponibles permettent de visualiser les interactions potentielles entre un ligand et son récepteur.

L'informatique permet de simuler la rencontre d'un médicament et sa cible. L'ordinateur reproduit en trois dimensions la forme compliquée de ces deux molécules, leur flexibilité, tout en tenant compte des forces d'attraction et de répulsion qui s'exercent. La simulation est alors réalisée afin de connaître l'activité du médicament : docking moléculaire.

Ainsi, l'outil informatique permet d'élaborer des hypothèses, de dessiner de nouvelles substances susceptibles de se lier à une cible connue, de tester un très grand nombre de molécules, de choisir les plus actives qui devront être synthétisées, de modifier une molécule pour la rendre plus active et d'éliminer celles qui sont toxiques ou inefficaces. Une représentation "la plus proche possible de la réalité" correspondra à une structure de plus basse énergie.

Cependant, si la molécule proposée apparaît active par la simulation, elle ne sera pas forcément dans la réalité. Elle devra donc être synthétisée pour vérifier ses éventuels effets thérapeutiques.

Ainsi, la modélisation moléculaire trouve de nos jours en effet, les structures tridimensionnelles de ces biomolécules sont le point de départ pour des simulations visant à comprendre leurs interactions, entre elles ou avec des petites molécules, ce qui est à la base d'un grand nombre de phénomènes chimiques, biologiques et biochimiques . Ainsi, la modélisation moléculaire trouve de nos jours d'importantes applications, parmi lesquelles trois exemples classiques sont :

- ✓ L'étude des propriétés des matériaux : à partir de simulations moléculaires, un lien peut être établi entre les résultats de calculs menés à l'échelle microscopique (sur un échantillon représentatif d'atomes en interaction) et des propriétés mesurables à l'échelle macroscopique (sur les systèmes matériels réels que l'on souhaite maîtriser). Ainsi, peuvent être étudiées les propriétés mécaniques des polymères, les propriétés électroniques de solides cristallins ou encore les propriétés thermodynamiques et spectroscopiques d'une large gamme de composés d'intérêt scientifique et technologique.
- ✓ Le développement de nouveaux médicaments : le mécanisme d'action de nombreux médicaments consiste à agir comme inhibiteur (ligand) d'une enzyme (récepteur) impliquée dans le développement de la maladie, que ce soit une protéine

d'un microorganisme pathogène ou d'un organisme humain. L'élucidation de la structure tridimensionnelle de protéines impliquées dans plusieurs pathologies a permis, via des simulations informatiques, la découverte d'inhibiteurs puissants pour ces protéines, en réduisant considérablement le nombre d'essais de screening nécessaires pour aboutir à un nouveau médicament.

- ✓ La rationalisation de l'ingénierie enzymatique : le fondement est le même que celui du développement de médicaments : en permettant d'étudier les interactions entre les substrats (ligand) et les enzymes (récepteurs), la modélisation moléculaire apporte une meilleure compréhension de la sélectivité enzymatique, au niveau moléculaire. Ceci peut rendre possible, par la suite, le développement d'approches prédictives pour la sélectivité des enzymes natives vis-à-vis d'un substrat ou d'une classe de substrats donnée, ou encore, d'identifier des acides aminés dont la mutation ponctuelle permettrait de modifier la sélectivité ou la spécificité enzymatiques.

III-3. Les objectifs de la modélisation moléculaire

- ✓ Visualisation infographique et conception moléculaire assistée par ordinateur de molécules à partir de données structurales.
- ✓ Utilisation de bases de données informatisées pour identifier des systèmes moléculaires avec des points communs.
- ✓ Utilisation de champs de force empiriques pour déterminer les propriétés moléculaires ainsi que les distances intra- et inter- atomiques.
- ✓ Corrélation des propriétés moléculaires avec leurs structures électroniques.
- ✓ Information sur les mouvements moléculaires dynamiques et leurs énergies.
- ✓ Applications au domaine de la reconnaissance moléculaire en chimie organique, bioorganique et médicinale ainsi qu'en science des matériaux.

Les utilisateurs de la modélisation moléculaire se divisent en deux groupes :

- Ceux qui font des calculs relativement précis sur des petites molécules (environ 100 atomes)
- Ceux qui cherchent par des méthodes plus approximatives à déterminer la structure des macromolécules.

III-4. Notion de la structure moléculaire

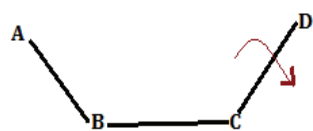
Une structure moléculaire est un enchaînement atomique caractérisé par la nature des constitutifs, leur ordre et leur mode de liaison, par sa géométrie interne (longueurs et angles de liaisons et configuration) et par sa conformation en structure 3D. Cette dernière est donc définie par la position dans l'espace des atomes constituant la molécule et elle dépend essentiellement des angles de rotations autour des liaisons simples.

Dans certains cas particuliers des changements conformationnels peuvent inclure des rotations autour de liaisons doubles ou plus précisément des changements de configurations. Considérons une molécule placée dans un certain environnement. En chimie organique et bio-organique cet environnement correspond généralement à un solvant (mais il peut y avoir d'autres cas: membrane, macromoléculaire, etc...). L'état d'équilibre du système moléculaire-solvant correspond à un minimum de l'énergie libre de Gibbs: $\Delta G = \Delta H - T\Delta S$

Cette énergie contient trois termes:

- ✓ L'énergie interne de la molécule correspond à des interactions interatomiques à l'intérieure de la molécule,
- ✓ L'énergie d'interaction molécule-solvant,
- ✓ L'énergie de solvant en dehors des zones d'interaction avec le soluté.

Considérons le premier terme qui est l'énergie interne :



Dans un fragment ABCD, on distingue l'énergie des liaisons AB, BC, CD.

C'est une énergie élevée de l'ordre de 400 à 600 KJ/mol. Pour deux atomes séparés par deux liaisons tels que les atomes AC, l'énergie d'interaction ne dépend pas de rotation autour de la liaison BC. Par contre, pour les atomes séparés par trois liaisons (ou plus) tels que AD, l'énergie d'interaction dépend de rotation autour de la liaison BC. A une conformation définie par l'ensemble des rotations autour des liaisons telles que BC, va donc correspondre une énergie comportant des termes fixes et des termes variables (Figure 2). C'est donc le dernier terme qu'il est intéressant de considérer pour définir l'énergie conformationnelle dont un minimum correspond à un état stable de la molécule isolée.

L'analyse de ces interactions montrent que leurs effets s'exercent dans un domaine de distances de l'ordre de 2 à 10 Å.

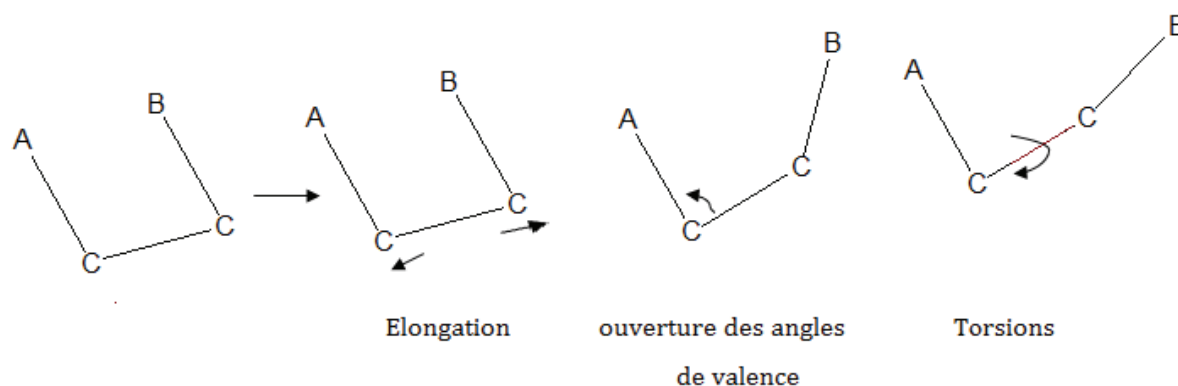


Figure 2 : Différentes possibilités de minimiser les interactions entre A et B

Pour une molécule isolée il existe généralement plusieurs minima d'énergie correspondant à plusieurs conformations stables ou conformères. Le minimum le plus bas définit la forme la plus stable. Les résultats de l'analyse conformationnelle montrent que pour les molécules en solution, la conformation stable (ou les conformations stables) se trouvent généralement dans le jeu des conformations stables de la molécule isolée. Les interactions avec le solvant ont surtout pour effet de modifier les valeurs relatives des énergies minimales ou de déplacer légèrement les positions des minima dans l'espace conformationnel. On s'accorde donc pour considérer que les résultats d'un calcul conformationnel sur la molécule isolée donnent une bonne idée des possibilités conformationnelles de cette molécule et peuvent conduire dans de nombreux cas à la détermination de la forme la plus stable ou d'une forme prédominante.

Il existe de nombreuses méthodes (quantique, semi-empirique, empirique) de calcul de l'énergie interne que nous verrons par la suite. On est aussi capable aujourd'hui de simuler (technique de Monte Carlo, dynamique moléculaire) des effets de solvant à la condition de disposer de logiciels adaptés et d'ordinateurs très puissants. Ces simulations ne sont pas accessibles à tous les laboratoires de routine. Nous mettrons surtout l'accent sur le cas de la molécule isolée et la technique de mécanique moléculaire.

III-5. L'importance de la simulation moléculaire

La chimie assistée sur ordinateur, basée sur des concepts de chimie théorique et informatique, est une discipline récente liée au développement des calculs digitaux et est devenue désormais un complément utile à l'expérimentation classique. Les calculs théoriques ne sont pas voués à remplacer l'expérience, mais destinés à analyser de façon critique les résultats, à proposer une interprétation des phénomènes physico-chimiques, et à suggérer de nouvelles expériences.

Toute simulation d'un processus naturel est appelée modélisation et la manifestation d'une telle simulation est appelée un modèle. Les molécules sont des collections microscopiques d'atomes qui possèdent certaines propriétés liées à l'observation de phénomènes macroscopiques.

Un modèle moléculaire est alors un essai de simulation de certaines de ces propriétés moléculaires sous forme macroscopique afin de mieux les manipuler et les comprendre. Le concept de modèle est aussi un terme à la mode dans les différentes disciplines des sciences naturelles et des sciences humaines. Dans le cadre d'une théorie naturelle, un modèle est la représentation d'une réalité tangible qui permet d'appréhender cette réalité ou bien de tester une théorie. Enfin, un modèle n'est pas une hypothèse qui peut être vraie ou fausse, mais c'est une fiction uniquement valable dans un contexte défini par le concepteur de celle-ci.

La simulation moléculaire correspond à la détermination numérique de grandeurs de la thermodynamique statistique et de propriétés structurales, énergétiques ou dynamiques du modèle d'un système (poly)moléculaire à partir d'une ou de l'ensemble des configurations géométriques engendrées par un ordinateur.

Autrement dit, les simulations sont assimilables à des expériences (numériques) sur un monde imaginaire constitué de modèles.

Ces calculs peuvent avoir différents objets :

- i) puisqu'il n'y a qu'un nombre faible et fini de problèmes exactement solubles en mécanique classique ou quantique, la simulation peut être un moyen de tester un aspect de ces théories sur un problème complexe.
- ii) le développement et l'exploration des conditions d'application de modèles doit permettre de guider la réflexion sur des résultats expérimentaux existants.
- iii) la modélisation des systèmes réels doit avoir un caractère prévisible afin de conduire à l'élaboration de nouvelles expériences et découvertes.

Un des buts actuels de la modélisation moléculaire est de reproduire ou prédire qualitativement, et si possible quantitativement, des phénomènes physico-chimiques et bio-organiques de plus en plus fins, des effets qui peuvent être mesurés par des différences de grandeurs thermodynamiques.

III-6. Représentation empirique de l'énergie Moléculaire

La stabilité de la structure tridimensionnelle d'une molécule est déterminée par les interactions intramoléculaires et les interactions avec le milieu extérieur (solvant).

La recherche des conformations stables d'une molécule consiste à déterminer les minima de l'énergie globale d'interaction. Cette énergie peut être calculée par des méthodes quantiques ou semi-empiriques généralement longues et onéreuses.

Pour faciliter les calculs, on considère généralement que le terme variable de cette énergie dépend de la construction de la molécule et de l'arrangement de ses atomes : c'est le principe des méthodes empiriques (mécanique moléculaire, dynamique moléculaire).

Dans la plupart de ces méthodes, il n'est pas tenu compte des interactions avec le solvant, mais uniquement des interactions entre les atomes constitutifs de la molécule.

III-7. La mécanique moléculaire

III-7-1. Principes théoriques

Les méthodes de mécanique moléculaire (MM) utilisent les lois de la physique classique pour la prédiction des structures et de leurs propriétés comme l'énergie. Au même titre que les méthodes de mécanique quantique est un outil puissant utilisé pour mieux comprendre les petits systèmes moléculaires et pour répondre à certaines questions concernant la structure électronique de molécules. D'autre part, la mécanique moléculaire décrit les molécules en fonction de leurs noyaux et non pas de leurs électrons, elle définit une fonction purement empirique prenant pour variable, les positions des noyaux et impliquant de manière implicite la présence des électrons.

En mécanique moléculaire:

- on positionne les atomes avec une distribution implicite des électrons.
- mais on ne peut pas améliorer, par exemple, un transfert de protons ni une tautomérisation car les longueurs de liaison changent.

En 1ère approximation: MM → info. conformation

MM → info. structure électronique

En MM il faut: types de liaison/propriétés → transférable

fragments moléculaires → transférable

En mécanique moléculaire on considère donc une structure moléculaire comme étant composée de billes et de ressorts (forces harmoniques) associés à une série de fonctions de potentiel. La somme de ces fonctions est exprimée sous la forme d'un **champ de force** (force field) moléculaire (figure 2).

$$E_{\text{total}} = E_{\text{élongations}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{Van der Waals}} + E_{\text{électro}} + E_{\text{hydrogène}}$$

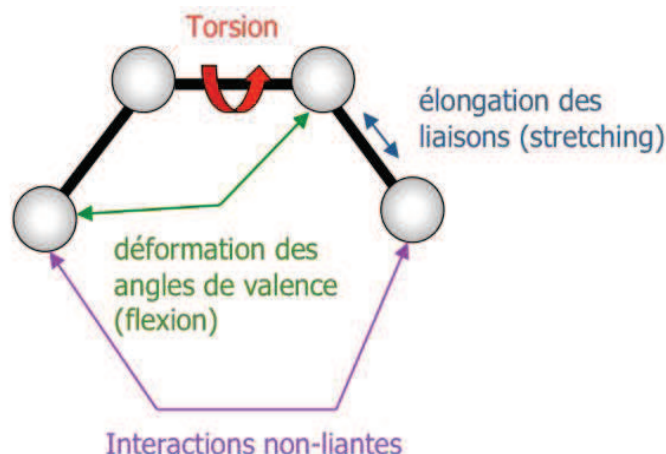


Figure 2 : Décomposition énergétique classique d'un champ de force.

Chacun de ces termes possédant une position d'équilibre préférentielle (longueur de liaison, angle de liaison...). 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. A cet égard, la mécanique moléculaire ressemble aux modèles de type "tiges et boules", mais elle est beaucoup plus quantitative.

En quelque fois appelée énergie stérique ou énergie potentielle. Elle dépend à la différence d'énergie entre la molécule réelle et une molécule aléatoire dans laquelle toutes les valeurs de la structure comme la longueur des liaisons et les angles de liaisons sont exactement à leur valeur idéale.

L'idée directrice de cette méthode est d'établir, par le choix des fonctions énergétiques et des paramètres qu'elles contiennent, un modèle mathématique, le "champ de force", qui représente aussi bien que possible les variations de l'énergie potentielle avec la géométrie moléculaire. Cependant, il n'existe pas encore de modèle unique permettant de simuler tous les aspects du comportement moléculaire, mais un ensemble de modèles.

Au sein de l'ordinateur, la représentation d'une molécule consiste donc en un ensemble de coordonnées atomiques, une liste de liaisons chimiques et un jeu de

fonctions et de paramètres constituant le potentiel d'interaction. Le potentiel empirique qui détermine l'énergie conformationnelle de la molécule est constitué de deux types de termes représentant, respectivement, les interactions entre les atomes liés (longueur de liaison, angle de valence, angle dièdre) et les atomes non liés (Van der Waals, électrostatique).

La recherche d'une conformation consiste alors à faire une minimisation de l'énergie intramoléculaire. Cette énergie potentielle est fractionnée en un certain nombre de terme additifs indépendants. Chacun de ces termes est représenté par une fonction analytique simple justifiée par des calculs quantiques et incluant des paramètres empiriques.

III-7-2. Modélisation par champs de force

Dans l'expression fonctionnelle des champs de forces classiques, l'énergie potentielle totale du système (E_{tot}) est exprimée comme étant la somme des contributions énergétiques des termes liés et de celles des termes non liés :

$$E_{\text{Total}} = E_{\text{stretch}} + E_{\text{bend}} + E_{\text{tors}} + E_{\text{Van der Waals}} + E_{\text{electro}} + \dots$$

$E_{\text{atomes-liés}}$

$E_{\text{atomes-non-liés}}$

Les interactions intramoléculaires ne dépendent que des coordonnées internes des molécules c'est-à-dire des liaisons, des angles de valence, et de torsions ($E_{\text{atomes-liés}}$).

Les interactions intermoléculaires prennent en compte les interactions qui n'interagissent pas par des termes de liaison, d'angle de valence et d'angle de torsion. Le potentiel non liant s'exprime en trois termes : un terme de Van der Walls, terme d'énergie électrostatique et un terme des liaisons hydrogène ($E_{\text{atomes-non-liés}}$).

L'énergie stérique minimale calculée est la somme des contributions apportées par les énergies d'élongation, de flexion, de Van der Waals, électrostatique .

Elle peut se deviser en termes d'interaction intermoléculaire concernant des atomes non liés chimiquement (électrostatiques, Van der Waals, etc.), et un terme d'interaction intramoléculaire concernant des atomes liés chimiquement (liaisons, angles, dièdres, etc.).

Les interactions de Van Der Waals sont des dipôles non permanents de faible rayon d'action (figure 3). Elles sont nombreuses et contribuent essentiellement dans la recherche de la concordance stérique entre le ligand et la protéine réceptrice.

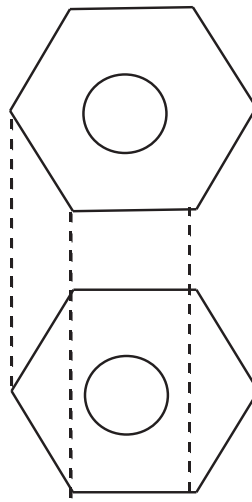


Figure 3: Interactions de Van Der Walls

La liaison hydrogène est formée d'une association entre un atome électronégatif et un hydrogène lié sur un second atome relativement moins électronégatif. Les atomes qui participent généralement dans la formation de ponts hydrogène sont l'oxygène et l'azote (Figure 4). L'énergie d'une liaison hydrogène est relativement élevée, de l'ordre de 2 à 5 Kcal/mol. Dans les systèmes biologiques à très courte distance (0,8 à 2,8 Å) entre un H lié à un O ou un N. Ils sont peu nombreux et s'adaptent très bien à la flexibilité (l'angle peut varier de 120° à 180°).

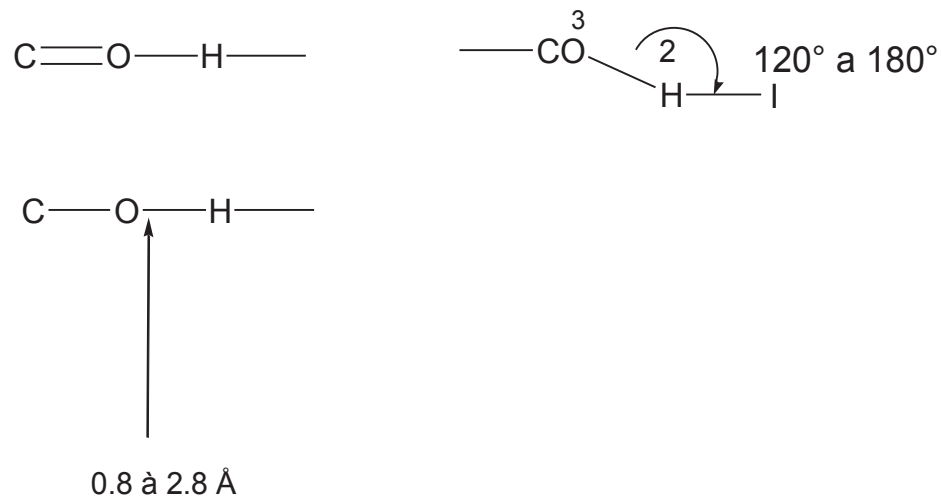


Figure 4 : ponts hydrogène

Les interactions électrostatiques agissent à plus grande distance entre dipôles chargés. La distance pour une liaison saline, par exemple, est de 3,7 à 4,5 Å (Figure 5), et sa force est de l'ordre de 8 Kcal/mol. En présence de solvant, ces interactions sont favorisées du fait de la polarité et de la petite taille de la molécule d'eau.

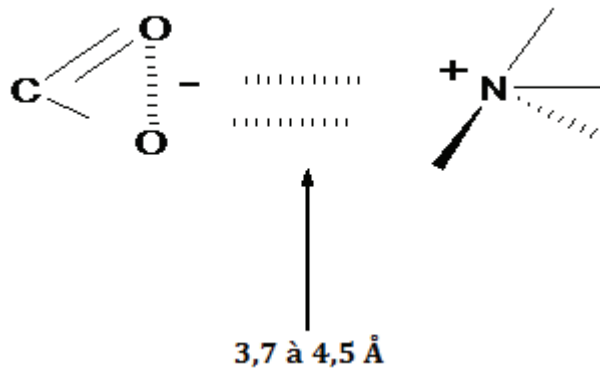


Figure 5 : Interaction électrostatiques

On appelle champ de force le modèle mathématique représentant l'énergie potentielle d'une molécule en mécanique moléculaire.

Sous ce terme sont en fait regroupés deux éléments:

- d'une part l'expression des différentes fonctions contribuant au calcul énergétique .
- d'autre part les valeurs des différentes constantes paramétrant ces fonctions.

Ce potentiel empirique est constitué de deux parties représentant les interactions entre atomes non liés (électrostatique et Lennard-Jones) et les interactions entre atomes liés (représentation harmonique des déformations des liaisons, des angles de valence et des barrières énergétiques des angles de torsion propres et impropres) .

III-7-3. Types de champ de force

Le champ de force est une expression que la mécanique moléculaire a emprunté à la spectroscopie, en modifiant légèrement sa signification. En effet, il est important de remarquer qu'un champ de force décrivant correctement les spectres de vibration de molécules est souvent impropre à décrire les structures moléculaires et vice versa. Ce n'est donc pas une fonction arbitraire de coordonnées cartésiennes, mais bien une somme de contributions d'atomes liés et d'interactions non-liantes.

- ✓ Un champ de force consiste dans un ensemble de fonctions d'énergie potentielle (analytical energy function) associée à une série de paramètres numériques, obtenus expérimentalement (IR, micro-onde...) ou évalués théoriquement. C'est donc une énergie potentielle multidimensionnelle ; c'est-à-dire une somme de contributions liantes et non liantes à N-corps.
- ✓ Pour une seule molécule, on peut compter plusieurs milliers de ces contributions individuelles.
- ✓ Le développement d'un nouveau champ de force ou de paramètres (**paramétrisation**), demeure un défi majeur pour les fonctions chimiques complexes, i.e., cétone conjugué, phosphonate, sulfate...

- ✓ De plus, la compatibilité d'un système de champ de force à un autre n'est pas souvent possible.

Pour rendre compte des phénomènes de couplage des mouvements de déformation de liaison et de déformation des angles de valences, Allinger a introduit un **terme croisé** ("cross term") dans le programme MM2. Il faut aussi tenir compte de déformation "**hors plan**", dont l'énergie est associée à la déformation de la géométrie de liaison des atomes sp^2 .

$$E_{hp} = 1/2 K_{hp} d^2 \quad \text{où } d \text{ est la distance entre l'atome central et le plan de ses substitués,}$$

K_{hp} la constante de déformation "hors plan".

Paramétrisation

- Paramètres de références: Les fonctions d'énergie décrites précédemment contiennent de nombreux paramètres (K_b , K_θ , moments dipolaires, barrières rotationnelles, etc...) qui sont ajustés généralement à partir de données expérimentales déduites de l'analyse de composés modèles (spectroscopie vibrationnelle, cristallographie, thermodynamique, etc...) ou dans quelques cas déduits de calculs de mécanique quantique. Ces paramètres sont optimisés de façon à ce que les propriétés calculées (géométries, énergies, chaleurs de formation...) se rapprochent le plus possible des propriétés expérimentales des composés modèles de la banque de données.

- Paramètres de substitution: Les banques de données des différents champs de force sont généralement très limitées. l'absence de paramètres de référence en particulier pour les systèmes hétérocycliques nécessite un choix de paramètres de substitution .

Conclusion

Le principal défaut des champs de force vient de la description de charges ponctuelles fixes. De plus le traitement de l'énergie électrostatique est discutable. On néglige les effets dus aux charges distribuées en volume et l'évolution de cette distribution en fonction de la conformation. La polarisation est ignorée ainsi que les effets à longue portée. Les autres termes sont une approximation satisfaisante.

On suppose que chaque type de fonction de potentiel est en principe transférable d'une molécule à l'autre. Puisqu'il n'y a pas de règles absolues concernant le nombre et

le type de fonctions d'énergie de potentiel utilisables, on retrouve un grand nombre différent de champ de force en mécanique moléculaire.

Il existe de nombreux champs de force et de nombreuses manières de définir l'énergie potentielle. Parmi les champs de forces le plus connus, figurent :

- ✓ **EMO** : Le programme EMO (Energy Of Molecule), est développé par B. Blaive, il est basé sur le champ de force MM2. Il est articulé autour de trois menus :

Menu 1: entrée de la molécule à l'aide d'un clavier.

Menu 2 : manipulation géométrique sur la molécule.

Menu 3 : minimisation de l'énergie par mécanique moléculaire.

- ✓ **MM2/MM3/MM4** : MM2 est le premier champ de force développé par Allinger et col. Il a été conçu au début pour les molécules simples (alcane, alcène, alcyne non conjugués, amines...), mais ses versions améliorées MM3 (1989) et MM4 (1996) lui permettent de traiter des molécules organiques de plus en plus complexes.
- ✓ **AMBER** (*Assisted Model Building with Energy Refinement*), a été écrit par Kollman, Le champ est paramétré pour les protéines et les acides nucléiques (UCSF, 1994). Il a été utilisé pour les polymères et pour d'autres petites molécules.
- ✓ **CHARMM** (chemistry at harvard using molecular mechanics) , spécialisé dans le traitement de molécules biologiques, telles que les acides nucléiques et les protéines .
- ✓ **CHARM (Bio+)** : Développé par Karplus et col, pour le calcul de biomolécules. son concept est semblable à celui d'AMBER. Bien qu'au début, ce champ de force est conçu pour les aminoacides et les protéines, maintenant il traite d'autres biomolécules.
- ✓ **OPLS** : Le programme OPLS (Optimized Potentials for Liquid Simulations), comme l'indique son nom, est conçu pour optimiser le potentiel qui permet la description des propriétés de solvatation. Il est écrit par W. L. Jorgensen et J. Tirado Rives .
- ✓ **GROMOS** (Groningen molecular Simulation Program Package) est écrit par Van Gunsteren et conçu spécialement pour les biomolécules en milieu aqueux en vue de l'étude des interactions entre les molécules d'eau et les groupements polaires des protéines,
- ✓ **SPASIBA** : (Spectroscopic Potential Algorithm for Simulating biomolecular conformational Adaptability), élaboré par Gérard Vergoten et col. (1995). Il combine le champ de force spectroscopique modifié de Urey-Bradley-Shimanouchi et le champ de force AMBER. Il permet de trouver les structures, les

énergies de conformation et les fréquences vibrationnelle au minimum énergétique d'une molécule .

- ✓ **MMFF** (Merck Molecular Force Field),
- ✓ **CFF** (consistent force-field) Le premier champ de force polarisable utilisé de manière importante dans de nombreuses applications aux systèmes biologiques.
- ✓ **CVFF** (Consistent Valence Force Field), utilisé dans le traitement de petites molécules organiques et de biomolécules,
- ✓ **ECEPP** (Schraga, 1983) : petites et protéines.
- ✓ **DREIDING-II** (Goddard III, 1990) : produits org, peptides et protéines (BIOGRAF).

III-7-4. Minimisation d'énergie

Une simulation typique de mécanique moléculaire est la minimisation d'énergie, le but étant de trouver une conformation de faible énergie d'une molécule ou d'un système moléculaire.

Le rôle de la minimisation est essentiellement de relaxer la structure initiale et d'éliminer les mauvais contacts intra atomiques. Les structures « minimisées » correspondent généralement au minimum local le plus proche de la structure de départ plutôt qu'au minimum global d'énergie. Cette détermination de la conformation moléculaire de plus basse énergie est réalisée par minimisation de l'énergie. Elle peut se faire selon différentes méthodes mathématiques, leur principe étant d'ajuster (à partir d'une conformation initiale) les paramètres géométriques de la structure moléculaire de façon à minimiser la valeur de la fonction énergétique.

Les méthodes de minimisation ne déterminent ainsi que la conformation stable la plus proche en énergie de la conformation initiale de la structure moléculaire, celle-ci pouvant correspondre à un minimum énergétique global ou simplement local : Les minima locaux sont inévitables car les longueurs de liaison et les angles de valence sont rapidement optimisés au stade initial de la minimisation ; mais quand ces derniers ont atteint leur valeur d'équilibre, d'autres mouvements torsionnels sont inhibés puisque pour réduire d'avantage l'énergie totale du système il faut pouvoir déplacer plusieurs atomes(Figure 6).

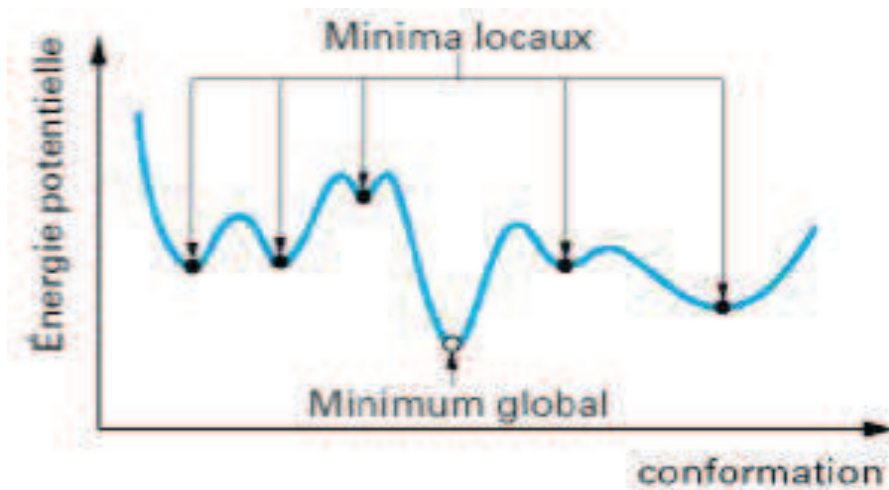


Figure 6 : Déterminer les minimas de l'énergie globale d'interaction

a- Minimum local , minimum global

MM est basée sur la MQ classique plus l'approximation de Born-Oppenheimer. Les électrons sont plusieurs milliers de fois plus légers que les noyaux et en conséquence se déplacent beaucoup plus vite. L'approximation de Born-Oppenheimer stipule que le mouvement des électrons peut être découplé de celui des noyaux, conduisant à deux équations séparées. On ignore donc les électrons et les atomes sont traités comme des points de masse chargés avec une taille définie. Le mouvement des électrons dépend donc paramétriquement des positions des noyaux.

- itération successive d'une mol. initiale → conduit à une géométrie optimisée
- faut des paramètres fiables .
- BUT → trouver un minimum **GLOBAL**
 - chaque paramètre qui détermine la géométrie initiale est modifié par incrément jusqu'à l'obtention d'un min. **LOCAL**
 - faut ensuite une DM pour franchir certaines barrières
($E_{\text{cinétique}}$ sous forme d'augmentation de Température)

b- Minima locaux - quasi inévitable

- car la longueur de liaison et l'angle de liaison sont rapidement optimisés au stade initial de la minimisation;
- mais quand eux ont atteint leur valeur d'équilibre, d'autres mouvements torsionnels sont inhibés puisque pour réduire d'avantage l'énergie totale du système il faut pouvoir bouger plusieurs atomes.

c- Limitation du champ de force

- MM basée sur des méthodes empiriques, applicables à des classes de composés avec des valeurs expérimentales, adéquates pour permettre la paramétrisation.
- pas de valeurs de ΔH disponibles pour des systèmes dont les angles varient de plus de 20° (déformation).
 - cycles à 3,4.....: résultats difficiles à reproduire.

III-8. Les simulations de dynamique moléculaire

III-8-1. Principes théoriques

La dynamique moléculaire est une approche de simulation qui prend en compte les interactions entre atomes et molécules pendant une période de temps, à une certaine température, permettant ainsi une évaluation des mouvements atomiques. On se trouve alors en présence d'un assemblage de masses, de ressorts et de vitesses et donc le problème peut être résolu en intégrant la seconde équation de mouvement de Newton.

$$\vec{F}_i = m_i \vec{a}_i = m_i \frac{d^2 \vec{r}_i(t)}{dt^2}$$

\vec{F}_i : Vecteur force agissant sur l'atome i.

m_i : Masse de l'atome i.

\vec{a}_i : Vecteur accélération de l'atome i.

\vec{r}_i : La position de l'atome i.

Les vitesses sont appliquées de façon aléatoire sur les atomes (sphères) selon la température choisie pour l'expérience. En utilisant la DM, on peut étudier la flexibilité ou la rigidité des biomolécules, mesurer les interactions intermoléculaires entre la protéine et les ligands ou d'autres biomolécules, calculer l'énergie libre ou bien étudier l'effet du solvant sur la structure des biomolécules. Parmi les algorithmes intégrateurs les plus souvent implémentés dans les logiciels se trouvent les algorithmes de Verlet et de Verlet-Leapfrog.

III-8-2. Calcul de dynamique moléculaire

Une simulation de dynamique moléculaire se réalise généralement en quatre étapes :

- 1- Une étape dite de thermalisation au cours de laquelle les vitesses des atomes sont augmentées progressivement afin d'atteindre la température finale choisie.

2- Une étape d'équilibration pendant laquelle les vitesses ne sont plus modifiées. L'énergie cinétique se répartit sur toute la molécule afin d'atteindre son état d'équilibre.

3- Enfin, pendant la dernière étape, appelée dynamique productive, les coordonnées et les vitesses sont sauvegardées pour une analyse de la dynamique.

III-8-3. Recherche conformationnelle à l'aide de la dynamique moléculaire

Une limitation inhérente aux algorithmes classiques de minimisation est que normalement ils localisent un minimum local proche de la conformation initiale, ce qui n'est pas nécessairement le minimum global. Ils ne poussent donc pas le système au-delà des barrières d'énergie, mais descendent plutôt vers le puits le plus près.

Les vraies molécules trouvent naturellement leur conformation au minimum global en fluctuant autour d'un ensemble de conformations ayant des énergies rapprochées. Le minimum global est ainsi échantillonné par simulation en MM et le négatif du gradient du potentiel d'énergie correspond à la force présente. Ainsi la force et la masse de chaque atome (loi de Newton) peuvent être intégrées numériquement pour prédire où les atomes vont se déplacer dans un court intervalle de temps (pas). En utilisant une succession de cet intervalle de temps, on peut construire une trajectoire du mouvement en fonction du temps.

Un ennui cependant de la dynamique est que plusieurs centaines et même milliers d'états conformationnels doivent être échantillonnés durant cette trajectoire. De plus, le phénomène d'intérêt peut très bien se produire dans une échelle de temps qui va de la milliseconde à des minutes et même plus longtemps à 300 K. Actuellement, la DM est limitée à une dizaine de nanosec.

III-9. Docking Moléculaire

Docking (ancrage ou amarrage en français) est le nom donné aux simulations moléculaires dans lesquelles différentes approches sont combinées pour étudier les modes d'interaction entre deux molécules. Dans la plupart des cas, il s'agit d'un récepteur macromoléculaire (cible de docking) dont la structure tridimensionnelle est connue et d'une petite molécule (ligand). Le récepteur macromoléculaire étant le plus souvent une protéine, le terme docking seul est couramment employé pour désigner un « docking protéine-ligand ».

Le plus important problème pour l'étape de Docking est de parcourir le mieux possible l'espace conformationnel. La complexité de ce problème est fonction du nombre de degrés de liberté, de translation, de rotation en plus des conformations de départ possibles du ligand. Afin d'éviter des calculs que les machines ne peuvent résoudre ou

seulement dans des temps bien trop importants, plusieurs approximations sont possibles. Les algorithmes de recherche de la flexibilité du ligand peuvent se classer en trois principes, nommés combinatoire, stochastique, et déterministe.

III-9-1. L'approche combinatoire

Cette approche est basée sur des grilles de valeurs pour chaque degré de liberté, et chacune de ces grilles est explorée de manière combinatoire au cours de la recherche. En raison de l'effet combinatoire, le nombre d'évaluations augmente bien plus rapidement que le nombre de degrés de liberté. Pour cela, des critères de fin sont imposés pour éviter à l'algorithme de parcourir des portions de l'espace qui ne mèneraient qu'à de mauvaises solutions.

Dans un premier temps le ligand est découpé en parties rigides et flexibles. Entre les points où des rotations sont possibles, une ou plusieurs «ancres» rigides sont définies, ensuite une première partie rigide est mise en interaction avec le récepteur puis les parties flexibles sont ajoutées de manière successive avec une exploration des angles de torsion.

Le plus important est le choix des fragments de base à placer en premier dans le site car il est très difficile pour les algorithmes de compenser une mauvaise position initiale. Les différentes implémentations changent par la manière dont le premier segment rigide est placé dans le site et dans les procédures d'élimination quand le nombre d'angles de rotation augmente.

Nous supposons que la position du segment rigide de départ ou ancre fait partie des n positions de plus basse énergie possible. Le nombre n détermine la largeur de l'arbre de recherche.

III-9-2. L'approche stochastique

L'approche stochastique consiste à effectuer des changements aléatoires dans la structure tridimensionnelle du ligand. Habituellement il s'agit de modifier un degré de liberté à chaque fois. L'un des points faibles de cette méthode est l'incertitude de convergence. Pour l'éviter, il faut multiplier les calculs, indépendamment les uns des autres. Un des principaux algorithmes stochastiques est la méthode de Monte Carlo .

Pour la méthode de Monte Carlo le ligand est considéré dans son ensemble et les changements s'effectuent aussi bien sur les translations, les rotations que sur les torsions. A chaque mouvement, la molécule est minimisée et son énergie est calculée. Pour déterminer si une structure doit être conservée les programmes se servent du critère de Metropolis .

Pour les systèmes macromoléculaires le critère de Metropolis peut s'exprimer ainsi : à partir d'une configuration i , on tire au hasard une configuration j . Cette nouvelle

configuration est acceptée avec une probabilité α , qui est la probabilité que le système puisse aller d'un état d'équilibre i vers un état j .

La différence d'énergie entre la nouvelle et l'ancienne configuration est calculée. Si la nouvelle configuration est d'énergie plus basse elle est acceptée. En revanche si elle est d'énergie plus haute, un nombre au hasard entre 0 et 1 est tiré, si ce nombre est inférieur à α la nouvelle configuration est acceptée, sinon l'ancienne configuration est gardée. Ce critère est, dans la plupart des cas, très efficace et simple à mettre en œuvre.

Pour obtenir la structure la plus stable ou quand il est raisonnable de penser que la structure de départ est très éloignée de celle de l'interaction, il est possible d'utiliser le recuit simulé. Cette méthode est une variante de la méthode de Monte Carlo qui s'applique particulièrement bien aux macromolécules. Elle consiste à effectuer plusieurs cycles de calculs avec la particularité que le premier cycle sera simulé à haute température et que dans les cycles suivants ce facteur sera progressivement diminué. Chaque nouveau cycle commence avec la meilleure structure du précédent.

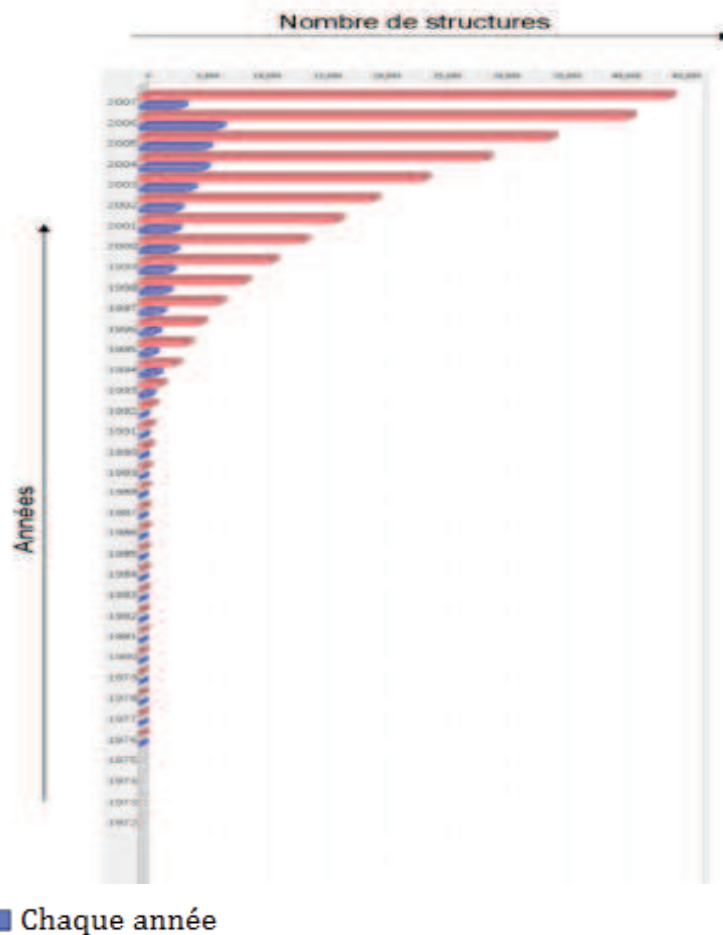
III-9-3. L'approche déterministe

Dans l'approche déterministe, l'état initial détermine les mouvements à effectuer pour générer l'état suivant. Cet état devant être d'énergie égale ou inférieure à celle de l'état initial. Le problème des systèmes déterministes est qu'ils peuvent facilement rester piégés dans un minimum local car leurs capacités à surmonter des barrières énergétiques sont faibles.

III-10. La structure des protéines

Trois méthodes expérimentales permettent aujourd'hui de déterminer la structure des protéines : la résonance magnétique nucléaire (RMN), la microscopie électronique et la cristallographie par rayons X. Cette dernière technique est responsable de la majorité des structures issues d'une base de données de structures accessibles gratuitement appelée la «protein data bank» (PDB),¹² c'est-à-dire plus de 40 000 structures protéiques (juin 2007).

La RMN fournit, quant à elle, 6 200 structures contre seulement 150 dans le cas de la microscopie électronique. Nous avons représenté le nombre de structures définies les 30 dernières années, toutes techniques confondues (Figure 7).



■ Total ■ Chaque année

Figure 7: Evolution du nombre de structures protéiques disponibles dans la PDB

L'histogramme montre une évolution exponentielle du nombre de structures définies chaque année (en bleu). Le nombre de structures totales (en rouge) suit également une courbe exponentielle de croissance, prouvant la progression des techniques. Une autre preuve du succès de ces méthodes est la qualité des structures.

a) La résolution

La résolution en angström de la protéine est une des données reflétant la qualité des structures ayant permis de construire le modèle cristallographique. Généralement, la résolution est limitée par la manière dont les cristaux diffractent, le temps nécessaire pour collecter des données de meilleures résolutions, la dynamique de la protéine, la qualité de l'appareillage et la température du système.

Une résolution proche de 1 Å permet de distinguer tous les atomes y compris les hydrogènes. Une résolution de l'ordre de 6 Å permet seulement de distinguer que des structures de types «hélice α » ou «feuille β » par exemple (Figure 8).

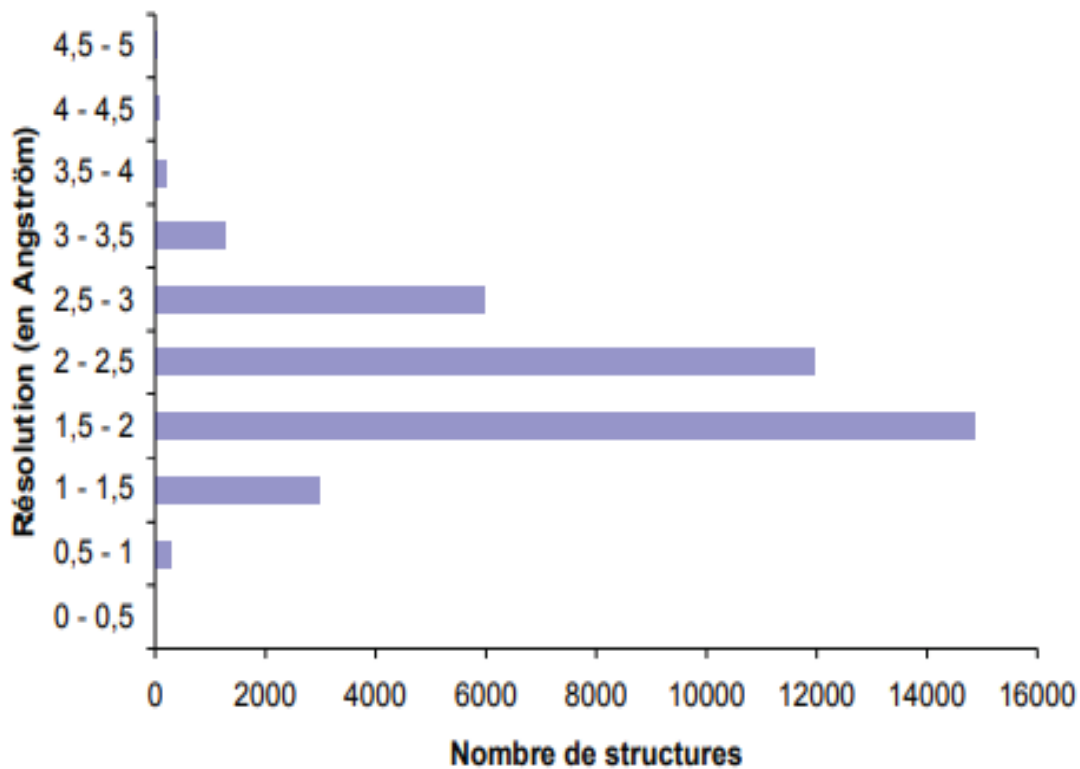


Figure 8. Nombre de structures protéiques par intervalle de résolution donné. L'histogramme précédent illustre parfaitement la répartition de la qualité des structures au sein de la PDB. La majorité des structures ont une résolution comprise entre 1,5 et 2,5 Å. Cet intervalle de résolution prouve que les structures sont de bonne qualité.

b) Le facteur R

Un autre indicateur est le facteur R qui est une grandeur indicatrice de l'écart entre les facteurs de structures observés et calculés. Le facteur R est compris entre 0 et 1 (plus le facteur R est proche de 0 et plus la prédiction est juste).

c) Choix d'une structure cristallographique

La combinaison des deux facteurs évoqués précédemment permet de réaliser un choix pertinent de structure cristallographique. Par exemple, une structure de protéine dont le facteur R est proche de 0 mais possédant une mauvaise résolution n'a aucune signification. Typiquement, une structure dont le facteur R avoisine la valeur de 0,2 et possédant une résolution correcte (inférieure ou égale à 2 Å) correspond la plupart du temps à une structure précise et bien définie. Il existe à ce jour des outils capables d'évaluer la qualité des structures cristallographiques.

d) Facteur d'agitation thermique

Les atomes du cristal bougent autour d'une position atomique moyenne. Les rayonnements ne voient pas les atomes exactement à la même position dans les différentes mailles du cristal. Ce facteur permet d'observer la flexibilité des résidus.

e) Détermination du site actif

Une étape fondamentale dans la stratégie de criblage virtuel est l'identification des résidus d'acides aminés pouvant intervenir dans le processus de reconnaissance du ligand. Un ligand cristallisé avec le récepteur est susceptible de renseigner sur l'emplacement du site actif. Malgré tout, cette information doit être interprétée avec prudence car, pour un même récepteur, le site actif peut être différent selon le profil pharmacologique de la molécule (agoniste, antagoniste, agoniste inverse).

Dans le cas où aucun ligand n'a été cristallisé avec le récepteur, des méthodes capables de détecter les cavités sont une bonne alternative. Nous avons utilisé CASTp pour mesurer le volume des sites actifs. Cet algorithme, disponible en ligne, mesure les surfaces accessibles des cavités accessibles mais également inaccessibles. Les méthodes utilisées consistent à évaluer la surface accessible aux solvants (surface Richards) et la surface moléculaire (surface Connolly).

Enfin, CASTp permet de mesurer le nombre d'ouvertures, leur taille ainsi que leur circonférence. Généralement, ce type d'algorithme identifie plusieurs régions possibles. Il revient alors à l'utilisateur de tester tous ces sites de fixation potentiels et accessibles vis-à-vis du profil pharmacologique de molécules connues.

L'opération d'arrimage du ligand dans le site actif de la protéine est appelée docking moléculaire (figure 9). Une simulation de docking comprend essentiellement deux étapes :

- La première (le docking) est l'étape de sélection, consistant à placer le ligand dans le site actif de la protéine et à échantillonner les conformations, positions et orientations (poses) possibles, en ne retenant que celle qui représentent les modes d'interactions les plus favorables.
- La deuxième (le scoring) est l'étape de classement, qui consiste à évaluer l'affinité entre le ligand et la protéine et de donner un score aux poses obtenues lors de la phase de docking. Ce score permettra de retenir la meilleure pose parmi toutes celles proposées .

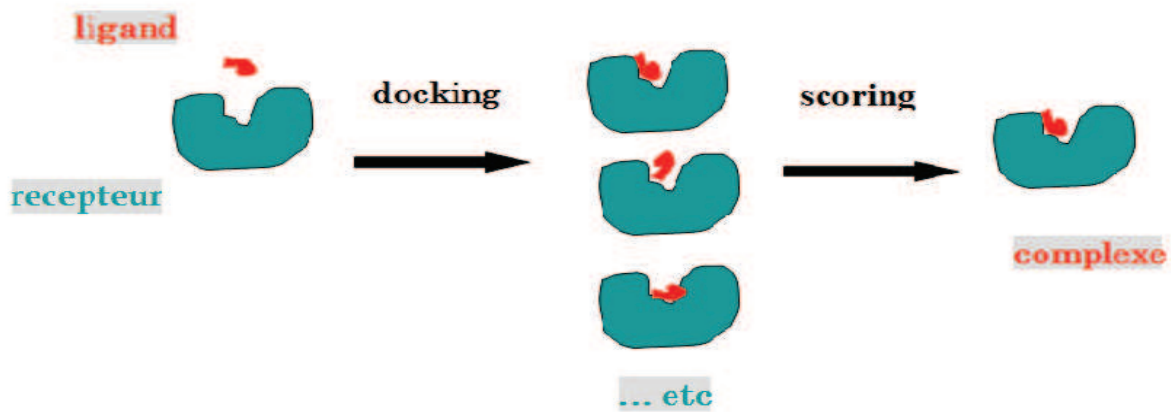


Figure 9 : Principe générale d'un programme de docking

III-11. Protocole général de docking

Les approches utilisées actuellement sont exclusivement calculatoires et évaluées par des outils de visualisation. Ces approches peuvent être décomposées en quatre à cinq phases successives (**figure 10**)

- Choix du mode de représentation des protéines (tout atomes, pseudo-atomes, grille, etc.),
- Exploration conformationnelle (corps-rigide position/orientation du ligand et/ou flexible position/orientation/forme du ligand),
- Minimisation de la fonction d'évaluation de l'énergie d'interaction (ou fonction de score) des conformations issues de l'exploration,
- Regroupement par ressemblances et classification par évaluation plus fine du score , accompagnée d'une étape non automatique d'évaluation visuelle des résultats lorsque le score ne permet pas de discriminer la conformation native des différentes conformations générées.
- Une étape optionnelle d'affinement des complexes sélectionnés par minimisation ou dynamique moléculaire.

Un algorithme de recherche pour explorer les possibilités de modes de liaison, un mécanisme pour placer le ligand dans le site de liaison et une fonction de score pour classer les différents modes de liaison.

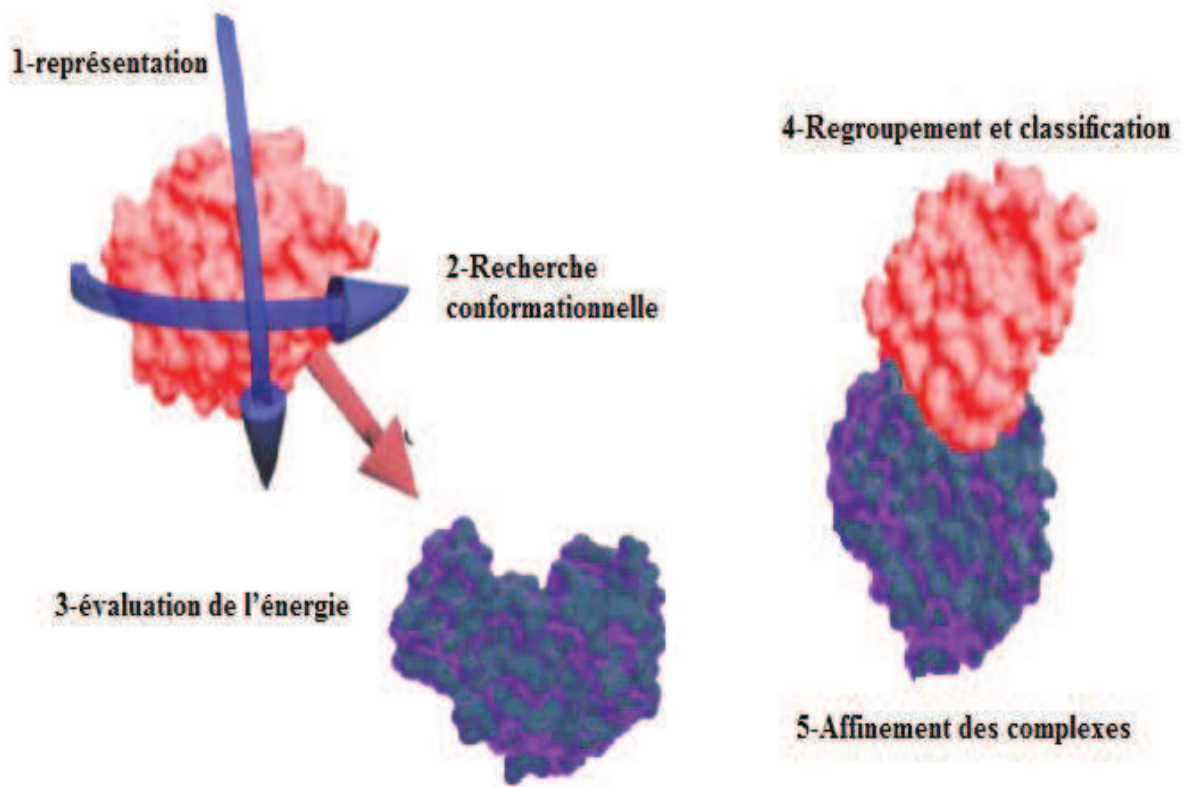


Figure 10 : Schéma du protocole général de docking

III-12. Docking rigide vers l'introduction de la flexibilité

Les premiers logiciels de docking développés au début des années 80 considéraient le récepteur et le ligand rigides en se basant sur le modèle « clé-serrure » (figure 11). C'est sans aucun doute la catégorie la plus simple et rapide de docking.

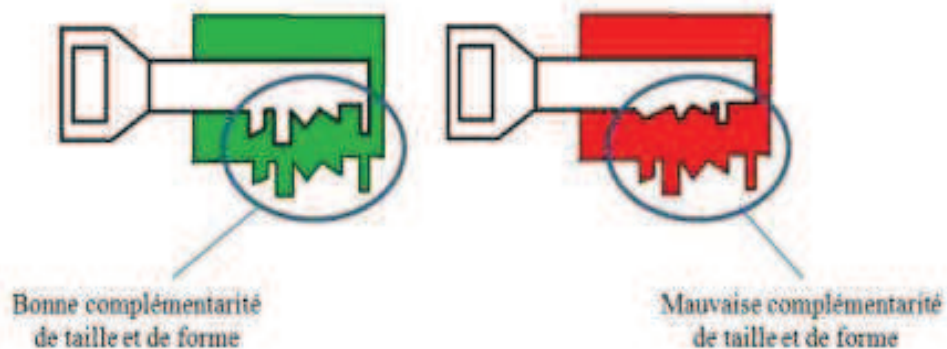


Figure 11 : Schéma illustrant le principe de « clé-serrure ».

Ils ont distingué deux niveaux de docking :

- ✓ Le docking semi-flexible est "asymétrique" et généralement utilisé pour l'amarrage protéine-ligand, le ligand étant considéré comme flexible, la protéine gardée rigide.

- ✓ Le docking flexible enfin, traite la flexibilité des deux molécules, mais la flexibilité permise est limitée pour simplifier les modèles.

L'utilisation de tels programmes mènent à de bons résultats, principalement quand la protéine a une flexibilité limitée. En effet, dans de tels cas, la structure cristallographique peut être considérée comme plus représentative de l'état de la protéine dans son environnement naturel, ce qui augmente les chances de simuler correctement le mode d'interaction protéine ligand.

III-13. Un processus de Docking Moléculaire

L'intérêt du Docking Moléculaire est de déterminer le mode d'interaction et de formation d'un complexe de deux molécules (un ligand et un récepteur) ou de plusieurs molécules, en cherchant des orientations dans l'espace et des conformations favorables pour la fixation d'un ligand à un récepteur ; et la structure obtenue confère les propriétés du complexe ainsi formé.

Un processus de « Docking Moléculaire » peut être résumé en trois étapes :

- L'identification du site de fixation et caractérisation du récepteur.
- L'orientation du ligand dans le site du récepteur.
- L'évaluation de l'orientation du ligand et de son adéquation au récepteur.

A la fin de chaque cycle d'exploration, on va enregistrer la meilleure pose avec le ligand (la plus faible énergie).

III-14. Programmes de docking moléculaire

Un grand nombre de programmes (commerciaux ou non) de docking moléculaire sont d'ores et déjà disponibles. Parmi ceux-ci, nous citerons par exemple AUTODOCK, FLEXX, GOLD, DOCK, SURFLEX, MOLEGRO VIRTUAL DOCKER, MOE et UCSF CHIMERA etc. Ils diffèrent les uns des autres sur la manière de représenter le système moléculaire et la manière de déterminer le score de docking (fonction de score).

L'application des programmes d'amarrage permet de conduire à de nombreuses réussites dans le domaine de la découverte de nouvelles molécules bioactives.

III-15. Algorithmes de docking

En principe, un docking peut être fait de façon « manuelle », directement par le modélisateur, en plaçant le ligand dans le site actif de la protéine à l'aide d'une interface graphique. Ensuite, la géométrie de l'ensemble est optimisée de manière à corriger les problèmes stériques et obtenir un complexe énergétiquement stable.

Cette approche est appliquée quand on a une idée précise du mode d'interaction réel du ligand. Néanmoins, le plus souvent, le mode d'interaction réel n'est pas connu. Dans ce cas, tester manuellement toutes les conformations et orientations des ligands s'avère impossible d'un point de vue pratique, même en considérant la protéine comme un corps rigide. Pour contourner cette difficulté, les algorithmes de docking ont été conçus pour rechercher de façon objective, rapide et efficace les modes d'association protéine-ligand les plus favorables.

Les algorithmes de docking peuvent être séparés en deux grandes classes : ceux qui ne tiennent pas compte de la flexibilité de la protéine, en traitant celle-ci comme un corps rigide, et ceux qui sont capables de prendre en compte, partiellement ou totalement, la flexibilité du récepteur.

- ✓ Dans les procédures de docking considérant la protéine comme un corps rigide, la flexibilité du ligand seule est prise en compte pour l'obtention de complexes. Selon la méthode utilisée pour générer les conformères du ligand et les placer dans la cavité catalytique du récepteur, les algorithmes peuvent être sous-divisés en algorithmes de simulation de MM et DM, de forme, systématiques et stochastiques. Plusieurs articles de revue décrivant le principe, les points forts et les limitations de ces méthodes sont disponibles. Toutefois, cette classification doit être considérée avec précaution, puisque un bon nombre d'algorithmes combinent plus d'une méthode pour la génération et l'échantillonnage de conformères du ligand. Dans la plupart des cas, l'utilisation d'algorithmes considérant la protéine comme un corps rigide mène à de bons résultats, principalement quand la protéine a une flexibilité limitée. En effet, dans de tels cas, la structure cristallographique peut être considérée comme plus représentative de l'état de la protéine dans son environnement naturel, ce qui augmente les chances de simuler correctement la complexation des ligands.
- ✓ Certaines protéines présentent naturellement des régions de grande flexibilité, subissant des réarrangements considérables en présence d'un ligand. Dans ce cas, négliger la flexibilité de la protéine peut mettre en péril la fiabilité des résultats de docking, et rend nécessaire l'utilisation d'approches capables de tenir compte de la flexibilité du système entier. Des méthodes indirectes ou directes, où la flexibilité de la protéine est partiellement ou totalement prise en compte, sont décrites dans la littérature. Néanmoins, ces méthodes ne sont pas souvent utilisées car le gain en précision par rapport aux algorithmes traditionnels est généralement trop petit par rapport à l'augmentation du temps de simulation.

III-16. Fonctions de score

La prédiction de la conformation du ligand dans le site actif de la protéine est le plus grand défi du design des drogues basé sur la structure. Le rôle des algorithmes de docking désignés à exécuter cette tâche est divisé en deux parties essentielles :

- Trouver toutes les conformations, translations, rotations du ligand dans le site actif de la protéine en un temps d'exécution très petit.

- Calculer l'énergie utilisée pour évaluer les interactions entre la protéine et le ligand, et classer les solutions . Donc le défi est de trouver une fonction qui peut incorporer toutes les contributions de l'attachement. Dans ce domaine plusieurs méthodes ont été développées. Les méthodes de champ de force calculent l'énergie libre ΔG d'un complexe protéine-ligand en utilisant les fonctions développées pour l'affinement de la structure 3D et les méthodes de la dynamique moléculaire.

La procédure de docking permet de générer une liste de complexes représentant les modes d'association favorables entre le ligand et le récepteur macromoléculaire.

L'étape suivante consiste à évaluer ces complexes, afin de trouver celui ou ceux les plus susceptibles de reproduire au mieux le mode d'association réel. L'association entre protéines et ligands est gouvernée par plusieurs paramètres thermodynamiques : les interactions hydrophobes, les interactions électrostatiques, les liaisons hydrogène, les effets de solvatation et les effets d'entropie. Théoriquement, le complexe est favorable si la variation d'énergie libre globale de complexation est négative ($G_{\text{complexation}} < 0$).

La fonction de score est une donnée numérique utile pour quantifier le degré avec lequel un ligand se complexe à un récepteur. C'est globalement une approximation de l'énergie libre résultant du passage de la forme libre de la protéine et du ligand à l'association sous forme de complexe. Le principe thermodynamique est le suivant:

$$\Delta G = \Delta G_{\text{complexe}} - \Delta G_{\text{ligand}} - \Delta G_{\text{protéine}}$$

En pratique, l'évaluation de l'énergie libre des complexes est une tâche souvent coûteuse d'un point de vue informatique, ce qui limite son utilisation en routine. De ce fait, des méthodes approximatives ont été développées pour distinguer (évaluer et classer) les meilleurs complexes parmi ceux générés par une procédure de docking : les fonctions de score.

L'établissement d'une bonne fonction de score est un important problème du docking. Il arrive souvent que la solution évaluée comme étant la plus probable ne soit pas la forme native attendue. Ceci peut être dû au fait que le complexe natif n'est pas forcément celui qui présente la plus grande surface d'accès, ou encore le plus grand

nombre de liaisons hydrogène disponibles. Pour cela il existe différents types de fonctions de score selon les critères sur lesquels celles-ci sont basées. Les interactions entre une protéine et un ligand sont en général de nature non covalente. L'énergie libre de Gibbs se calcule selon l'Équation 1.

$$\Delta G = -RT \ln K_i = \Delta H - T\Delta S$$

Équation 1. Energie libre de Gibbs

R: constante des gaz (8,314 J. K⁻¹ mol⁻¹)

T: Température (K)

K_i: constante d'inhibition

La grandeur du ΔG en solution aqueuse est comprise entre -10 et -70 KJ/mol. Les composantes de l'énergie libre sont:

- la complémentarité géométrique, stérique et de surface entre le ligand et le récepteur
- les contacts présents entre deux régions lipophiles des deux entités.
- les liaisons hydrogène Les molécules d'eau dans les cavités des protéines peuvent parfois être un élément fondamental. Elles sont capables d'assurer le relais entre le récepteur et le ligand et ainsi créer des réseaux de liaisons hydrogène.

Le calcul de l'énergie libre étant cependant lourd, son application devient impossible lorsqu'il s'agit d'évaluer un grand nombre de complexes. L'estimation de l'affinité repose généralement sur un calcul d'énergie contenant un certain nombre d'approximations. Pour cela il existe différentes fonctions de score. Ces dernières ajustées dans les programmes de docking ont fait l'objet de nombreuses publications. Ces fonctions de score sont généralement classées en quatre familles:

- ✓ les fonctions de score basées sur un champ de force,
- ✓ les fonctions de score empiriques,
- ✓ les fonctions de score consensus,
- ✓ les fonctions de score de type knowledge-based.

Ces fonctions de score sont construites à partir de règles fondées sur une analyse statistique des complexes protéine-ligand résolus expérimentalement. La fonction PMF (Potential of Mean Force) fait partie de cette classe de fonction .

Il s'agit d'une technique de classement de complexes qui combine les résultats individuels de plusieurs fonctions de score afin de compenser les faiblesses des unes avec les forces des autres, en augmentant ainsi la fiabilité du résultat final.

a) Fonctions de score basées sur un champ de force

Les fonctions de score basées sur un champ de force calculent, par mécanique moléculaire, l'énergie d'interaction du complexe et l'énergie interne du ligand. G-Score 30 par exemple, fait appel au champ de force de Tripos, et AutoDock à celui d'AMBER.

Les interactions entre récepteur et ligand comprennent souvent des termes de Vander Waals et électrostatiques. L'énergie interne du ligand est généralement écrite de manière similaire.

Les principales limitations des fonctions de score basées sur des champs de forces proviennent du fait qu'elles ont été écrites pour des modèles en phase gazeuse et ne contiennent donc pas d'effet de solvation ni de terme d'entropie. Des extensions incluant un terme d'entropie pour le ligand (dans G-Score) et les liaisons hydrogène protéine-ligand (dans Gold et Autodock) ont été ajoutées récemment.

b) Fonctions de score empiriques

Les fonctions de score empiriques sont utilisées pour interpréter l'énergie d'interaction d'un complexe récepteur-ligand à partir d'une équation de sommation d'interactions chimiques localisées.

Ce type de fonction est basé sur la régression multiple pour ajuster les coefficients de fonctions selon la physique du système. Ajustement à partir d'un jeu de données de complexes récepteurs-ligands avec des affinités mesurées.

Les fonctions de score empiriques contiennent usuellement des termes décrivant les interactions ioniques, interactions hydrophobiques, les ponts ou liaisons hydrogène et les interactions engendrées par le changement d'entropie (pénalité d'entropie). Ceci dit, ces fonctions de score somment ces différents termes en les pondérant à des termes décrivant les différents types d'interactions moléculaires. La plupart des logiciels de docking utilise ce type de fonction à cause de son efficacité de point de vue rapidité et précision. Cependant, leur principal inconvénient est leur forte dépendance aux données de paramètre calibration.

c) Les fonctions de score type knowledge-based

Ces fonctions proviennent de l'analyse des structures tridimensionnelles de complexes ligand-protéine déterminés expérimentalement.

Des règles définissant la géométrie préférentielle des interactions sont déduites de ces structures grâce à des moyens statistiques. Cette alternative aux fonctions empiriques est plus tolérante quant aux interactions présentes au sein du complexe et leurs expressions sont moins strictes.

Ces fonctions statistiques, dépendent de leurs groupes d'apprentissage (expérience), elles ne peuvent modéliser que les interactions qui existent dans leurs bases de données expérimentales. Par conséquent, avec ce type de fonctions, il est difficile de modéliser les interactions qui ne dévient pas trop de leur modèle. L'un des potentiels les plus simples est le PLP (Piecewise Linear Potential). Il est basé sur quatre types d'atomes et prend en compte les interactions stériques et les ponts hydrogènes.

Les distances interatomiques sont prises en compte en dessous de 5 Å. De même l'état de référence est défini comme étant un complexe entre les deux entités chimiques qui n'interagissent pas.

Les fonctions de score type knowledge-based ont été utilisées avec succès dans des études d'amarrage de différentes cibles protéiques et ont montré une certaine amélioration dans la prédiction des modes de liaison corrects et le classement des complexes protéine-ligand par rapport aux fonctions de notation empiriques et celles basées sur le champ de force.

d) Fonctions de score consensus

L'idée principale de ces fonctions est combiner les informations obtenues à partir des différents scores obtenus et qui sont insuffisants des autres⁴⁰. Plusieurs études ont montré que les énergies libres des complexes sont performées par ces fonctions, ainsi les interactions protéine-ligand mieux que les fonctions individuelles. Un exemple de fonction de score consensus est X-CSCORE qui combine un PMF, et ChemScore. Le tableau suivant récapitule les avantages et les désavantages des différentes fonctions de score.

Tableau 3. Avantage et inconvénients des différentes fonctions de score.

Fonction de score	Avantages	Inconvénients
Basé sur les champs de force	<ul style="list-style-type: none"> - Les termes des FFs sont bien étudiés et ont une base physique. - Transférable, rapides quand elles sont utilisés sur une grille pré-calculée. 	<ul style="list-style-type: none"> - Ne tient compte que d'une partie de l'énergie potentielle. - Elles sont parfois augmentée par des termes de solvatation et d'entropie. - L'électrostatique est souvent surestimée.
Empirique	<ul style="list-style-type: none"> - Estimation rapide et directe de l'énergie de liaison. 	<ul style="list-style-type: none"> - Il existe peu de complexes avec à la fois des structures et des énergies de liaison connues précisément. - Forte dépendance sur l'orientation des atomes d'hydrogène. - Pas de véritables pénalités pour les mauvaises structures.

Knowledge-based	-Similaire aux méthodes empiriques, mais plus générales. Existence de données d'inter-distances plus que d'énergies de liaison expérimentales.	Les PMF sont "pairwise", alors que la probabilité de trouver les atomes A and B à une distance r n'est pas "pairwise" et dépend aussi des autres atomes.
------------------------	--	--

III-17. Interaction Ligand (Peptide) – Récepteur

* **Spécificité** \Rightarrow centre actif "message"
 \Rightarrow \Rightarrow Concept

* **activité** \Rightarrow site auxiliaire "adresse"

- différent du 2eme messenger (cAMP): effet amplificateur via la protéine G

* **sélectivité** \Rightarrow dominée par adresse

Tous les résidus d'un peptide non pas la même (également) importance pour l'activité biologique.

Donc action possible à plusieurs niveaux.

- Exigences du récepteur:

$$\Delta G_{\text{total}} = \Delta G_{\text{lipophile}} + \Delta G_{\text{stérique}} + \Delta G_{\text{électrostatique}} + \Delta G_{\text{polarisabilité}} + \Delta G_{\text{résiduel}}$$

"extrathermodynamique" de Hansch (non applicable aux peptides)

- Exigences membranaires:

$$\Delta G_{\text{total}} = \Delta G_{\text{électrostat}} + \Delta G_{\text{transfert}} + \Delta G_{\text{immobilisation}} + \Delta G_{\text{amphiphile}} + \Delta G_{\text{dipôle}}$$

Phase lipidique + ligand = médiateur (catalyseur) d'interactions ligand-récepteur

- Exigences spatiales du récepteur

Composante topologique

- Conformations préférentielles du ligand (peptide)

- Conformation active (ou induite) du peptide

III-18. La pharmacocinétique -ADME - Tox

Pharmacocinétique est traditionnellement définie comme l'étude du sort des médicaments dans l'organisme, depuis leur absorption jusqu'à leur élimination. Cette discipline permet ainsi de relier les doses de médicaments administrées aux concentrations sanguines observées, et de décrire leur évolution en fonction du temps.

Parmi les notions fréquemment utilisées en pharmacocinétique, citons : La biodisponibilité désigne la fraction de la dose de médicament administrée qui atteint la

circulation systémique, et la vitesse à laquelle elle l'atteint. Elle reflète principalement la phase d'absorption.

Pour être efficace en tant que médicament, une molécule puissante doit atteindre sa cible dans le corps en concentration suffisante, et y rester sous une forme bioactive suffisamment longtemps pour que les événements biologiques attendus se produisent. Le développement de médicaments implique l'évaluation de l'absorption, la distribution, le métabolisme et l'excrétion (ADME) de plus en plus tôt dans le processus de découverte.

Une grande variété de méthodes *in silico* partagent l'objectif de prédire les paramètres ADME à partir de la structure moléculaire. À noter, le travail pionnier de Lipinski et al examiné des composés actifs par voie orale pour définir des intervalles physicochimiques pour une probabilité élevée d'être un médicament oral. (C'est-à-dire la ressemblance médicamenteuse). C'est ce que l'on appelle *la Rule-of-five* qui définit la relation entre les paramètres pharmacocinétiques et physicochimiques.

En outre, la pharmacocinétique (comme l'absorption, gastro-intestinale) et la pharmacodynamie sont des prédictions telles que les règles de Lipinski, et le score de biodisponibilité. La règle de Lipinski est une règle appliquée pour évaluer la drug likeness pour prédire si un composé est susceptible d'être bioactif selon certains paramètres importants tels que le poids moléculaire, LogP, le nombre de HBA et HBD. L'outil suisse ADMET a utilisé le vecteur algorithmique de machine (SVM) avec de grands ensembles de données soigneusement nettoyés d'inhibiteurs / non-inhibiteurs connus ainsi que substrats / non-substrats.

ADMET (Absorption, Distribution, Métabolisme, Excrétion et Toxicité) dépistage aide à détecter la probabilité de drogue de composés. Le format SMILES des ligands a été chargé dans le serveur SwissADME, et l'ADMET le criblage a été effectué aux paramètres par défaut.

- ✓ Il est possible de décomposer le circuit du médicament dans l'organisme en quatre phases une phase de résorption, ou d'absorption, qui correspond au passage du médicament de son site d'administration vers la circulation systémique.
- ✓ Une phase de distribution, correspondant à la diffusion du médicament dans les différents tissus de l'organisme.
- ✓ Une phase de métabolisme, décrivant la transformation du médicament par l'organisme.
- ✓ Une phase d'élimination, s'intéressant aux processus d'élimination du médicament par l'organisme.

L'ensemble de ces quatre phases est fréquemment décrit par le sigle ADME. Certains auteurs élargissent ce concept en incluant une phase de libération du principe actif depuis la forme galénique administrée, et une phase de réponse au principe actif, décrivant l'enchaînement suivant :

Libération- Absorption – Distribution – Métabolisme – Elimination – Réponse On utilise alors le sigle LADMER pour désigner l'ensemble de ces phases .

a. Absorption

Dans la majorité des situations rencontrées en clinique, le site d'administration du médicament est distinct de son site d'action : administration par voie orale d'un médicament dont le site d'action pourra être le cœur, les vaisseaux, les reins par exemple.

Le site de l'administration, le tube digestif dans la situation présentée ici, n'est qu'un lieu de transit, le médicament étant ensuite véhiculé dans l'organisme par la circulation sanguine. L'absorption correspond donc au passage du médicament de son site d'administration à la circulation sanguine générale, au travers de membranes biologiques.

Avant toute absorption, il est souvent nécessaire que le principe actif soit libéré de la forme galénique : dissolution du comprimé, ou de la paroi de la gélule. Plusieurs phénomènes interviennent ensuite dans l'absorption des médicaments : d'une manière schématique, on peut distinguer des phénomènes d'absorption actifs et des phénomènes d'absorption passifs.

Cela consiste à analyser la faculté d'une molécule à pénétrer au sein de l'organisme après administration. Une faible solubilité (par exemple due à une trop grande hydrophobicité) ou une forte polarité ont un impact drastique sur l'absorption intestinale d'un composé.

b. Distribution

La distribution, qui correspond à la diffusion du médicament dans les différents tissus de l'organisme, est également influencée par le caractère plus ou moins lipophile du principe actif, ainsi que par son affinité pour différents tissus : os, tissu adipeux, ...

Le médicament peut se fixer de manière réversible aux protéines plasmatiques : cette fixation se fait principalement sur l'albumine et l'alpha 1 glycoprotéine acide, l'albumine fixant préférentiellement les médicaments acides, alors que l'alpha 1 glycoprotéine acide fixe davantage les médicaments basiques.

On distingue alors la fraction de médicament liée aux protéines et la fraction libre. Le plus souvent, seule la fraction libre peut agir sur son site d'action, diffuser dans les tissus, ou subir les processus de métabolisme ou d'élimination, si bien que l'on peut

considérer que la fraction du médicament qui est liée aux protéines plasmatiques constitue une forme de stockage du médicament.

Ce critère mesure la capacité d'une molécule à diffuser, par exemple via le flux sanguin, à travers l'organisme. En effet, une molécule doit pouvoir passer d'un compartiment à un autre, afin de pouvoir arriver in fine à l'endroit où sa cible doit être atteinte. La forte liaison à des protéines plasmatiques a un impact négatif sur la distribution d'une molécule.

c. Métabolisme

Le métabolisme correspond à la transformation du médicament par l'organisme. Cette transformation est essentiellement réalisée par des réactions impliquant des enzymes. Pour les médicaments, il est usuel de séparer en deux groupes les réactions impliquées dans le métabolisme : les réactions de phase 1 et les réactions de phase 2. Les réactions de phases 1 précèdent habituellement celles de phase 2 dans la chronologie du métabolisme d'un médicament.

Le filtre métabolisme vise à détecter (a) la stabilité de la molécule dans l'organisme qui impacte sur son temps d'action et (b) les métabolites de la molécule initiale, à savoir les composés résultant de sa dégradation ou de modifications enzymatiques ayant lieu au sein de l'organisme.

Chez l'homme, les cytochromes P450 du foie sont les principales enzymes modifiant les xénobiotiques. Ces derniers sont notamment rendus plus hydrophiles par l'introduction d'atomes d'oxygène. Les métabolites peuvent être inactifs, plus actifs que le composé original et bien entendu potentiellement toxiques, d'où la nécessité de les caractériser et de les étudier.

d. Elimination

Afin d'éviter les phénomènes d'accumulation, souvent synonyme de toxicité, il faut veiller à ce que les composés administrés, ainsi que leurs métabolites, soient bien excrétés de l'organisme, par exemple via l'urine ou les selles .

e. Toxicité

Comme son nom l'indique, ce filtre sert à mesurer la toxicité d'un composé et de ses métabolites. Désormais, la toxicité et le manque d'efficacité des candidats médicaments sont les deux plus grandes causes d'échecs dans le développement d'un médicament. Différents types de toxicité sont évalués, entre autre la cancérogénicité .

Un contributeur majeur dans le domaine permettant d'identifier rapidement et à grande échelle des molécules à caractère « drug-like » est communément appelées « règles de Lipinski » ou « la règle de 5 » permettant d'estimer la biodisponibilité d'un composé par voie orale à partir de sa structure bidimensionnelle (2D). Ces règles

concernant les propriétés physico-chimiques ont été définies après l'analyse de 2245 médicaments commercialisés ou en phase finale de développement .

- Le poids moléculaire du composé ne doit pas être supérieur à 500 daltons (Da).
- Le logarithme décimal du coefficient de partage eau / 1-octanol, noté logP, doit être inférieur à 5.
- Le nombre de donneurs de liaisons hydrogène doit être inférieur à 5.
- Le nombre d'accepteurs de liaisons hydrogène doit être inférieur à 10.

D'autres critères ont été mis en place pour compléter et ajuster les règles de Lipinski dans la sélection de composés « drug-like ».

Ainsi, Veber Choisissent d'utiliser critères suivants la surface polaire (PSA, polar surface area) du composé doit être inférieur à 140 Å, le nombre de liaisons de rotation ("rotatable bonds" en anglais) doit être inférieur à 10 Sont souvent employés en complément de la « règle des 5 » Ces critères ont été établis par l'étude de la biodisponibilité orale candidats médicaments .

Référence

- [1] Hermann Dugas. Principes de base en modélisation moléculaire. La librairie de l'université de Montréal 2000(Québec).
- [2] Defranceschi M (2011). Chimie et médicament. Ellipses éd. France. 25 : 160-161.
- Roskoski R (2008). VEGF receptor protein-tyrosine kinases: structure and regulation. Biochem Biophys Res Commun. 375 : 287-291.
- [3] Munagala R, Aqil F, Gupta RC (2011). Promising molecular targeted therapies in breast cancer. Indian J Pharmacol. 43 : 236-245.
- [4] Bresch D, Colin T, Ribba B, Saut O (2009). Computational modeling of solid tumor growth: the avascular stage. SIAM Journal on Scientific Computing. 25 : 160-161
- [5] Carmeliet P, Jain RK (2011). Molecular mechanisms and clinical applications of angiogenesis. Nature. 473 : 298-307.
- [6] Rachid Hmamouchi, , Mohammed Bouachrine et Tahar Lakhlifi. Tentative Pratique du Relation Quantitatives Structure-Activité/Propriété (QSAR/QSPR). Revue Interdisciplinaire. Vol1, n°1,3-4 . 2016
- [7] Brut M, Nouvelle approche méthodologique pour la prise en compte de la flexibilité dans les interactions entre molécules biologiques : Les Modes Statiques Université Toulouse III - Paul Sabatier, 2009.
- [8] Gohlke H. Klebe G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. Angewandte Chemie International Edition; 41 (15): 2644-2676 , 2002
- [9] Ferrara P, Gohlke H, Price DJ, Klebe G. Assessing scoring functions for protein-ligand. Interactions Journal of Medicinal Chemistry; 47: 3032-3047, 2004.
- [10] Lafond M , Docking et scoring. Schrödinger, 2015.
- Wang RX, Lai LH, Wang SM, Further development and validation of empirical scoring functions for structure-based binding affinity prediction. Journal of Computer-Aided Molecular Design; 16: 11-26, 2002.
- [11] Coupez, R.A. Lewis. Docking and Scoring - theoretically easy, practically impossible? Current Medicinal Chemistry; 13: 2995-3003, 2006.
- [12] A.R. Leach, B.K. Shoichet, C.E. Peishoff. Prediction of protein-ligand interactions. Docking and scoring : successes and gaps. Journal of Medicinal Chemistry; 49: 5851-5855 , 2006.
- [13] N. Moitessier, P. Eglebienne, D. Lee, J. Lawandi, C.R. Corbeil. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. British journal of Pharmacology; 153, 2008.
- [14] C.R. Corbeil, P. Englebienne, N. Moitessier. Docking ligands into flexible and

- solvated macromolecules-1. Development and validation of FITTED 1.0. Journal of Chemical Informatic Modelling; 47: 435-449, 2007.
- [15] N. Foloppe, R. Hubbard. Towards predictive ligand design with free-energy based computational methods? Current Medicinal Chemistry;13: 3583-3608, 2006.
- [16] A. Jain. Scoring functions for protein-ligand docking. Current Protein and Peptide Science; 7: 407-420, 2006.
- [17] M. Feher. Consensus scoring for protein–ligand interactions. Drug Discovery Today; 11: 421-428, 2006.
- [18] Jay A, Jain N. Surflex-Dock 2.1: Robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. J Comput Aided Mol Des; 21 :281–306, 2007.
- [19] S .Tian., et al. Ke application of in silico drug-likeness predictions in pharmaceutical research., Adv Drug Deliv Kev 86, 2–10 ,2015.
- [20] C. A. Lipinski, F. Lombardo, B. W .Dominy. & P. J .Feeney,. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug . Del iv. Kev. 46, 3–26 ,2001.
- [21] D.F .Veber, S.R .Johnson, H.Y .Cheng, B.R. Smith, K.W .Ward, et al. Molecular properties that influence the oral bioavailability of drug candidates. J. Med Chem 45: 2615-2623. 2002.
- [22] H. Fatoki.,D. Sanni., O .Adeoyo., B. Faleye. J Nat Prod Plant Resour., 6(6), 21-25. 2016.
- [23] Véronique MALLO. Sous-traitance et développement pharmaceutique d'un médicament : application a la conception d'un comprimé a croquer , diplôme d'état de docteur en pharmacie, Collège sciences de la sante université de bordeaux ,2015.
- [24] SALAH DEHBIA .Mémoire de Master. Modélisation de la structure d'une protéine impliquée dans le cancer et de ses interactions avec quelques ligands. Université Mouloud Mammeri Tizi Ouzo, 2014.
- [25] Patrick Darveau , mémoire .Étude des propriétés des dendrimères pour le transport de molécules bioactives .université Québec, Canada, 2018.
- [26] Corinne HAZAN, thèse de doctorat. Recherche d'inhibiteurs de haute affinité de L'ADN polymérase BETA par criblage virtuel et RMN. Université Toulouse III, 2008 .
- [27] Duan Y, Wu C, Chowdhury S, Lee MC, Xiong G, Zhang W, Yang R, Cieplak P, Luo R, Lee T , Caldwell J, Wang J, Kollmann P. Un champ de force au point de frais pour la

mécanique simulations moléculaires de protéines à base de phase condensée calculs de mécanique quantique. J Comput ChemCME; 24 (16) :1999-2012, 2003.

[28] Ewing T.J, Makino S., Skillman A. G., Kuntzi. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases., J. Comput. Aided. Mol. Des; 15 : 411-428, 2001.

[30] C.R. Corbeil, P. Englebienne, N. Moitessier. Docking ligands into flexible and solvated macromolecules-1. Development and validation of FITTED 1.0. Journal of Chemical Informatic Modelling; 47: 435-449, 2007.