

## Conception Rationnelle de molécules biologiquement actives.

Pr. Raphaël Terreux

Faculté de Pharmacie (ISPB)

Département pédagogique des Sciences Physico-Chimiques et Pharmacie Galénique  
8 avenue Rockefeller, 36373 Lyon Cedex 08, France



Groupe EC-Mo , Laboratoire de Biologie tissulaire et Ingénierie Thérapeutique (LBTI)

UMR N°5305 CNRS UCBL

7 Passage du Vercors, 69367 Lyon Cedex 07



# Plan général



## 1. Partie I : Concepts & Théories

1. Conception d'une molécule active
2. Modèles et Modélisation Moléculaire: définitions et objectifs
3. Méthodes de calcul

## 2. Partie II : Applications

1. Approche « récepteur »
  1. Docking
  2. Denovo design
2. Approche « sans récepteur »
  1. QSAR
  2. ADMET



# 1.1 Conception d'une molécule active

- Médicament : mélange de molécules actives (MA) et d'excipients
- Découvrir un médicament, une très vieille préoccupation
  - Antiquité :
    - Observation et empirisme
  - Avant 1980 :
    - Modifications chimiques basées sur l'empirisme
  - Depuis 30 ans :
    - Recherche de molécules hyperactives
    - Nouveaux domaines, pas de MA connues
    - Minimisation des effets secondaires



# 1.1 Conception d'une molécule active

## La plupart des médicaments connus

- Savoir-faire ancestral
- Approche basée sur la similitude
- Points communs à tous les médicaments
  - Concept de "drug-likeness"
  - Règles de Lipinski (Cf Cours Pr. M. Leborgne)

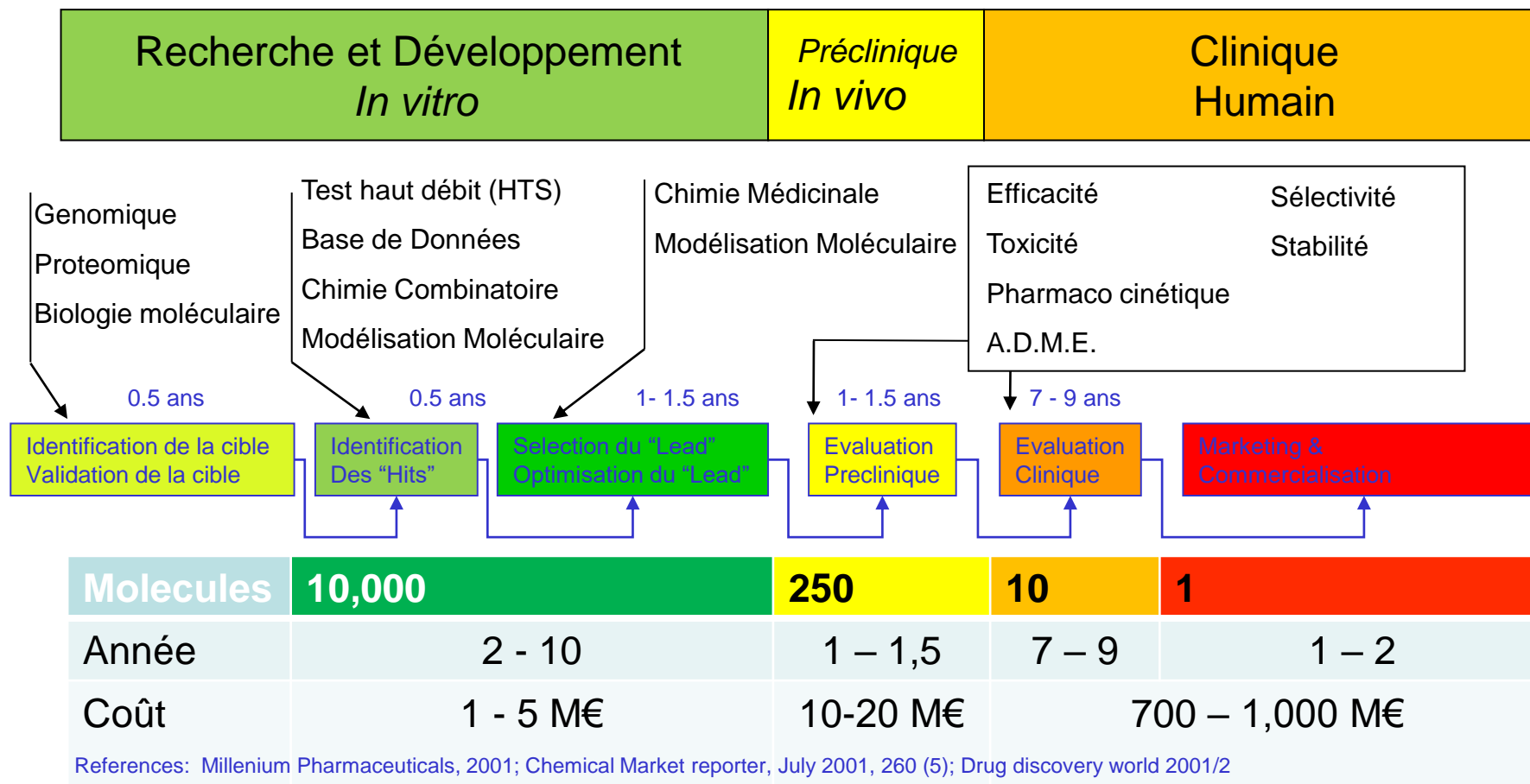
## Identification d'un « hit »

1. Approche classique à partir de substances naturelles
2. Chimie combinatoire / Criblage à haut débit (High Throughput Screening : HTS):  
Synthèse automatique et rapide de milliers de molécules  
Criblage expérimental à haut débit: HTS  
Capacité +100.000 molécules / jour
3. Modélisation moléculaire
4. ...

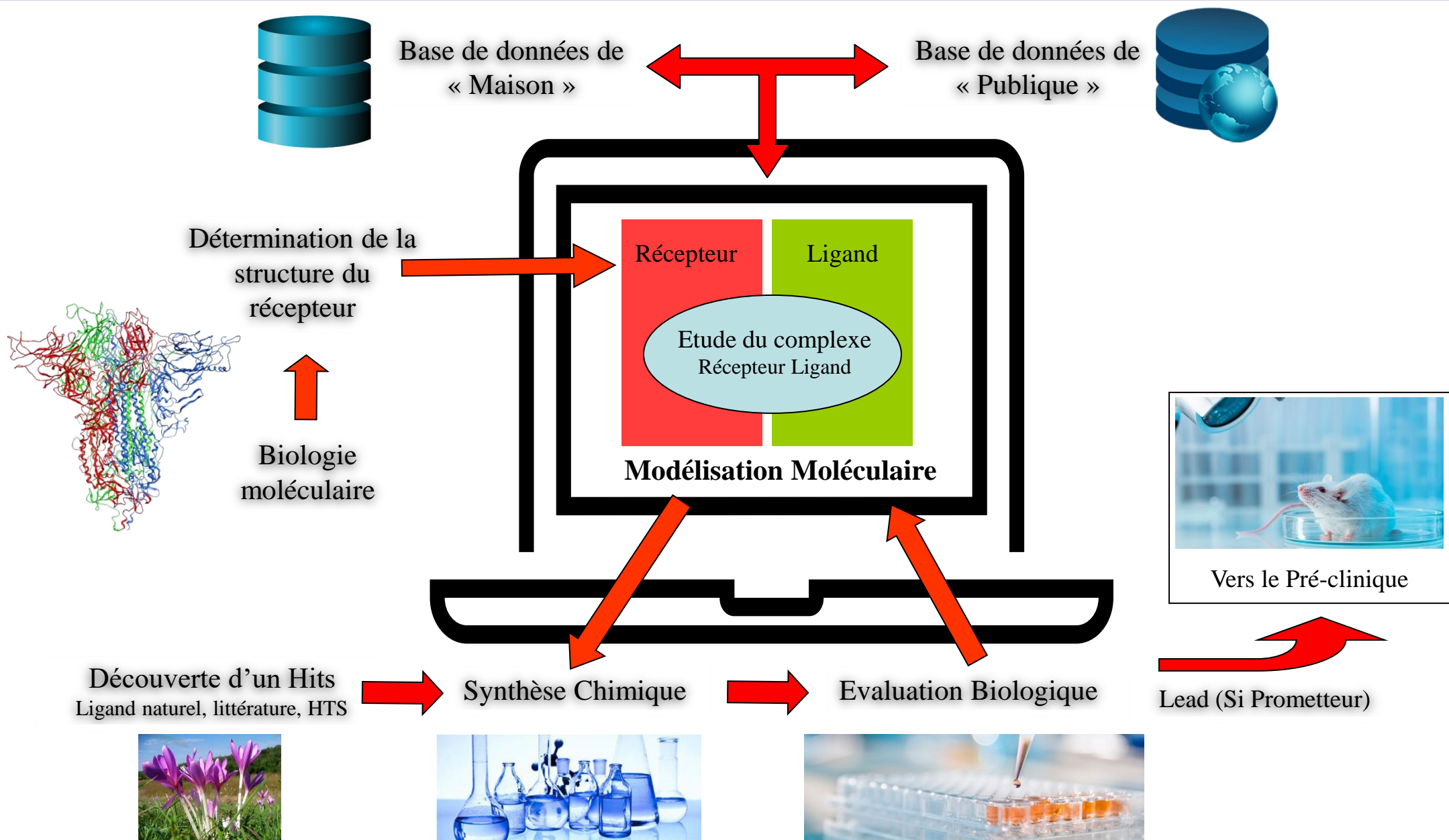


# 1.1 Conception d'une molécule active

- Industrie Pharmaceutique
  - Recherche et développement d'une molécule active longue (10 ans) et couteuse 1.000M€
  - Contrainte due à la durée de vie du Brevet: 20 ans.



# 1.1 Conception d'une molécule active



# 1.1 Conception d'une molécule active

## Identification d'un « hit »

1. Approche classique: A partir de substances naturelles :
  - Broyage / Extraction
  - Recherche d'activité
  - Synthèse et production
2. Chimie combinatoire / Criblage à haut débit (High Throughput Screening : HTS):
  - Synthèse automatique et rapide de milliers de molécules
  - Criblage expérimental à haut débit: HTS
  - Analyse des informations, tri et évaluation de la diversité moléculaire
  - Capacité +100.000 molécules / jour
3. Modélisation moléculaire



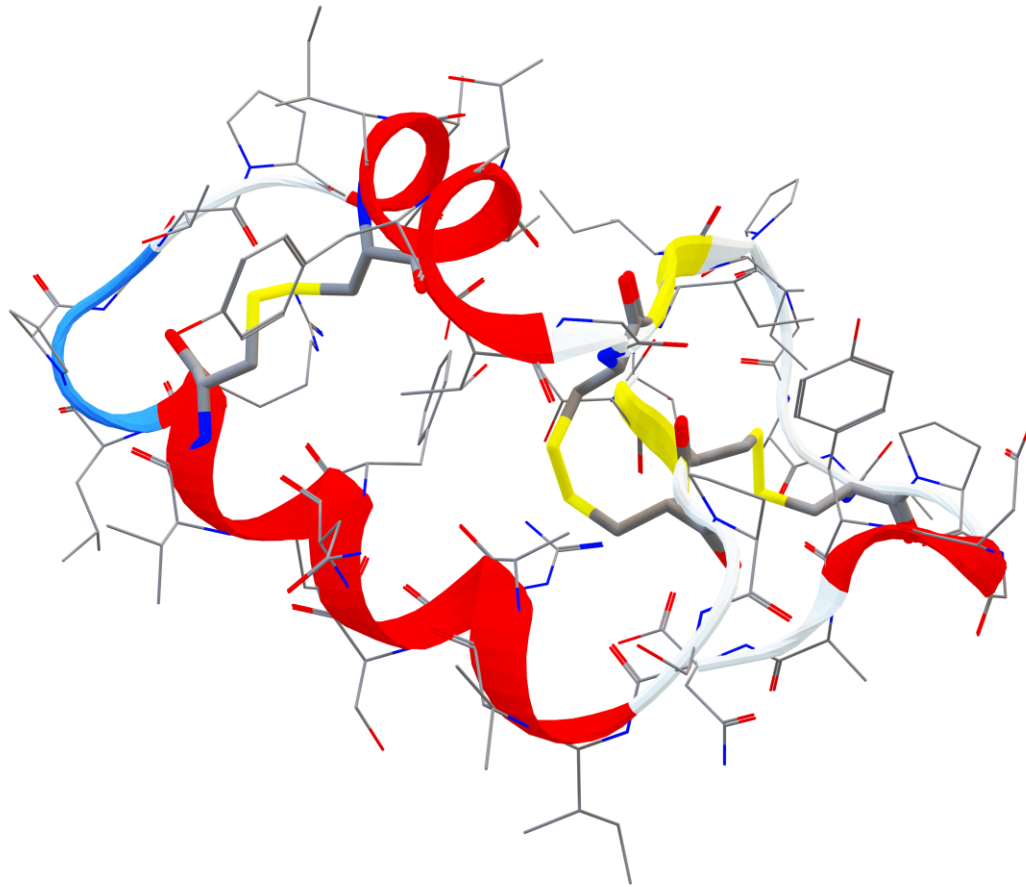
# 1.2 Modèle et Modélisation Moléculaire

- Modèle : représentation 3D simplifiée de la réalité
- Définition :  
« La modélisation moléculaire est l'investigation des **structures et des propriétés moléculaires**, utilisant la **chimie calculatoire** sur ordinateur et les techniques de **visualisation graphique** afin de donner une représentation tridimensionnelle plausible dans des circonstances définies. »
- Technique  $\approx$  50 ans
- Domaines d'étude : Chimie, Biochimie, Biologie moléculaire
  - Mise au point de molécules bio-actives
  - Interactions :
    - Petite molécule / Biopolymère (Protéine, ARN, ADN)
    - Biopolymère / Biopolymère
  - Structure des protéines, ARN, ADN, membranes
  - Simulation de réactions chimiques





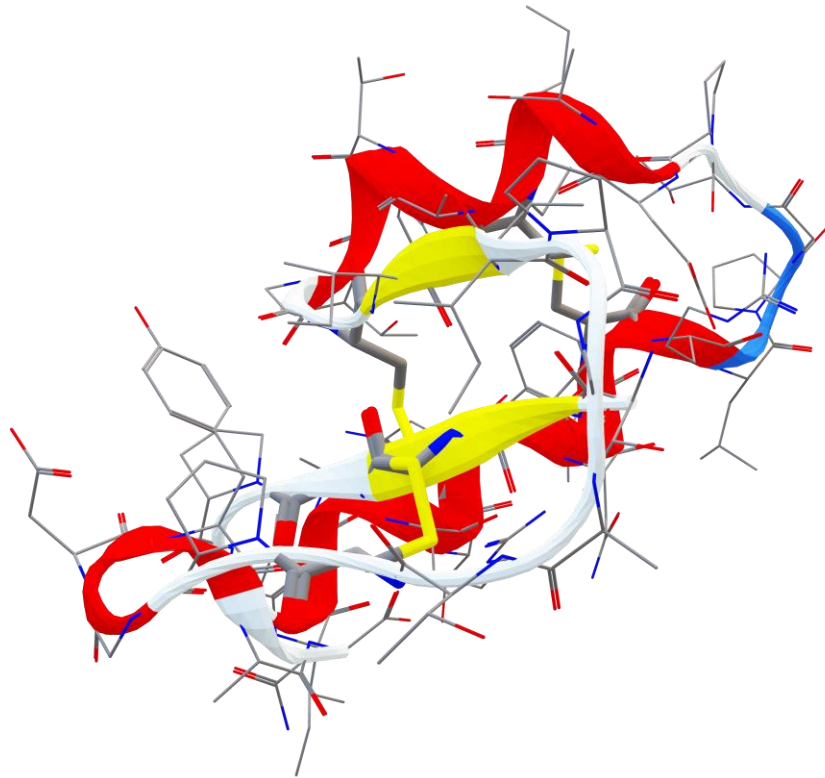
## 1.2 Modèle et Modélisation Moléculaire



- Modèle : représentation 3D simplifiée de la réalité
- Modèle de la protéine de Crambine (code PDB 1CRN)
- Le ruban représente la chaîne peptidique dont la couleur varie avec la structure secondaire
  - Hélice Alpha : Rouge
  - Boucle : Bleu
- Les 3 ponts disulfure entre 2 cystéines sont représentés sous forme bâton gris (carbone) / Jaune (souffre)



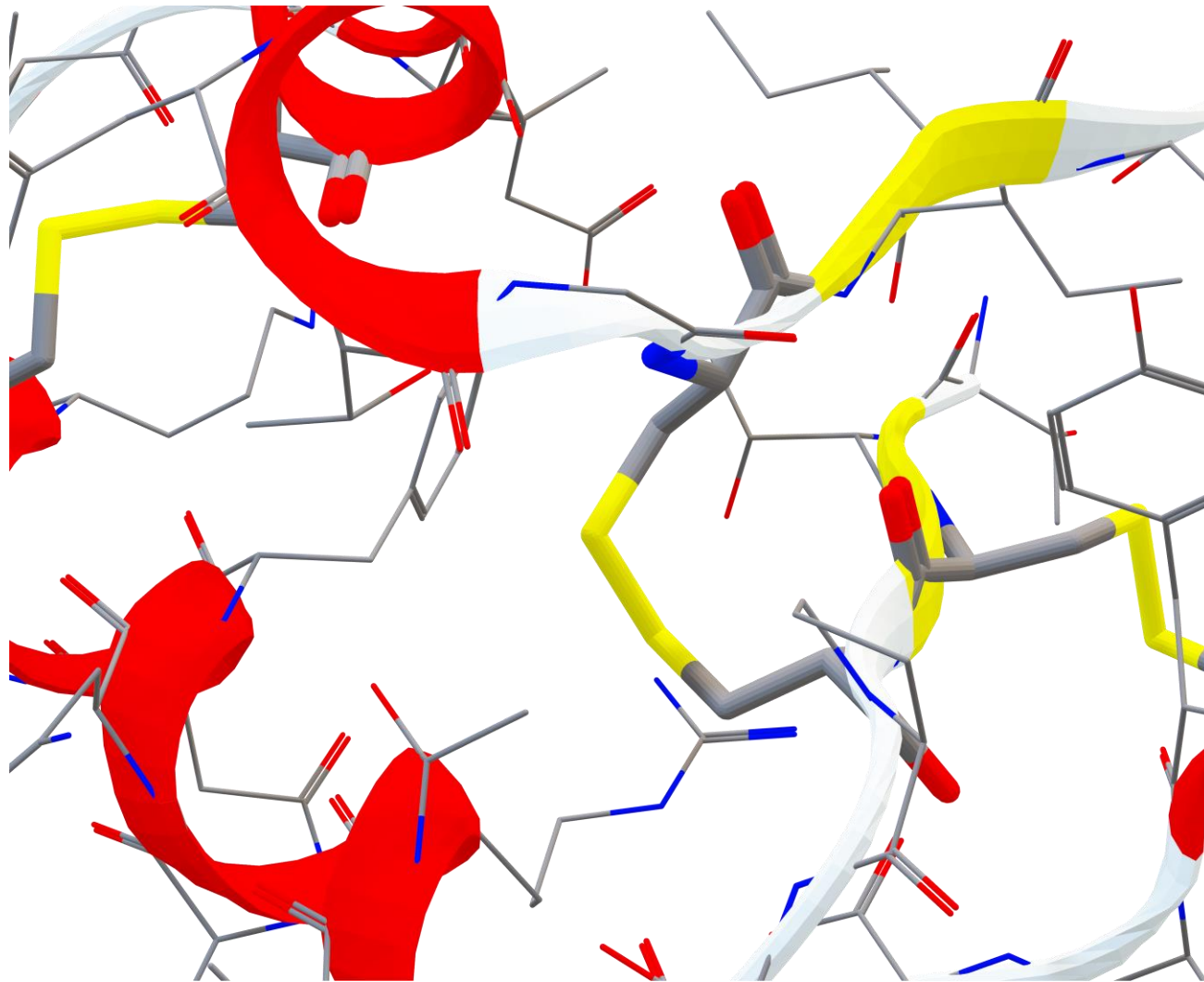
## 1.2 Modèle et Modélisation Moléculaire



- Modèle : représentation 3D simplifiée de la réalité
- Modèle de la protéine de Crambine (code PDB 1CRN)
- Le ruban représente la chaîne peptidique dont la couleur varie avec la structure secondaire
  - Hélice Alpha : Rouge
  - Boucle : Bleu
- Les 3 ponts disulfure entre 2 cystéines sont représentés sous forme bâton gris (carbone) / Jaune (souffre)



# 1.2 Modèle et Modélisation Moléculaire



- Modèle : représentation 3D simplifiée de la réalité
- Modèle de la protéine de Crambine (code PDB 1CRN)
- Le ruban représente la chaîne peptidique dont la couleur varie avec la structure secondaire
  - Hélice Alpha : Rouge
  - Boucle : Bleu
- Les 3 ponts disulfure entre 2 cystéines sont représentés sous forme bâton gris (carbone) / Jaune (souffre)



## 1.3 Méthodes de calcul

- Évaluer l'énergie interne d'une molécule
- Puis la minimiser pour avoir la conformation la plus correcte
- Plusieurs familles de méthodes, aucune n'est parfaite...
  - Réactivité chimique
    - ⇒ Mécanique Quantique
  - Interactions, structures...
    - ⇒ Mécanique Moléculaire



# 1.3.1 Mécanique quantique (QM)

- But : Déterminer les Orbitales Moléculaires qui sont une Combinaison Linéaire d'Orbitales Atomiques
- Résolution exacte de l'équation de Schrödinger possible seulement pour  $H_2^+$   
⇒ approximations
- Sous familles de méthodes
  - Semi-empirique : plus d'approximations car certains termes sont définis de façon empirique donc plus rapide mais peu précis
  - *Ab-initio* :
    - Hartree Fock (HF): très précise
    - Post Hartree Fock : ultra précise
  - DFT « Density Functional Theory » (DFT) : Théorie Fonctionnelle de la densité de gaz (d'électrons) : Méthode corrélative ultra précise



## *Applications :*

- Optimisation géométrique
- **Réactivité moléculaire**
- Calcul de propriétés chimiques
  - Énergie
  - Densité électronique
  - Orbitales atomiques et moléculaires
  - Charges atomiques
  - Moment dipolaire
- Propriétés RMN, IR
- Dichroïsme circulaire



# 1.3.2 Mécanique Moléculaire

## Définitions:

- Atome : boule indéformable chargée
- Liaison chimique : ressort

Affectation de caractéristiques (type) aux atomes et liaisons  $\Rightarrow$  Champs de force

Évaluation de l'énergie interne : somme d'énergie de déformation ( $E_{\text{pot}}$ )

$$\begin{aligned}
 E_{\text{pot}} = & \sum_b K_2 (b - b_0)^2 + \sum_{\theta} H_{\theta} (\theta - \theta_0)^2 + \sum_{\phi} \frac{V_n}{2} [1 + \cos(n\phi - \phi_0)] \\
 & + \sum \varepsilon [(r^*/r)^{12} - 2(r^*/r)^6] + \sum q_i q_j / \varepsilon_{ij} r_{ij} + \sum \left[ \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right]
 \end{aligned}$$

Exemple de Champ de force: AMBER

### Terme atomes liés

- (1) Déformation longueur de liaison
- (2) Déformation angulaire
- (3) Déformation angle dièdre

### Terme Atomes non liés

- (4) Contact entre atomes
- (5) Interaction électrostatique
- (6) Prime « liaisons hydrogène »



# 1.3.2 Mécanique Moléculaire

- Paramètres des atomes et de déformations spécifiques de chaque liaison  $\Rightarrow$  champs de force (généralistes ou spécialisés)
  - AMBER : acides nucléiques / acides aminés
  - CHARMM : acides aminés
  - COMPASS : matériaux en phase condensée
  - MMFF94 : Petites molécules

## Applications :

- Optimisation géométrique
- **PAS** de réactivité moléculaire
- Étude des mouvements moléculaires : Dynamique moléculaire
- Simulation d'arrimage de petites molécules / récepteur
- Étude des interactions





## 1.3.4 Optimisation géométrique

- Trouver une structure possible de la molécule  $\Rightarrow$  conformation d'énergie interne minimale
- Minimisation de l'énergie de déformation
  - Méthode MM (plus rapide, moins précise) ou QM
  - Procédé itératif
  - Algorithme de minimisation
    - Simplex : Procédé simple pour les premiers pas
    - Gradient conjugué : Bonne méthode
    - Newton – Raphson : Méthode recommandée pour les derniers pas...



## 1.3.5 Dynamique moléculaire

- Étudier la flexibilité d'une molécule, biopolymère...
- Utilisation de la relation de Newton:

$$\mathbf{F}_i(t) = m_i \mathbf{a}_i(t)$$

- Résolution de l'équation de mouvement par une méthode d'intégration numérique ( $t + \Delta t$ ), selon un procédé itératif
- Ajout d'énergie cinétique pour simuler l'agitation thermique

$$E_t = E_c + E_p$$

- Équilibration entre énergie cinétique (température système) et énergie potentielle (conformation)
- Plusieurs types de simulations : Pression et Température constantes (NPT), Volume et Température constants (NVT), Énergie et Volume constants (NVE), Énergie et Enthalpie constantes (NPH)
- Temps de simulation : quelques nanosecondes (ns)



# Plan général



## 1. Partie I : Concepts & Théories

1. Conception d'une molécule active
2. Modèles et Modélisation Moléculaire: définitions et objectifs
3. Méthodes de calcul

## 2. Partie II : Applications

1. Approche « récepteur »
2. Approche « sans récepteur »



## 2.1.1 Construction du récepteur

- Petite molécule : facile de créer un modèle
- Biopolymère : problème ?
  - Acide nucléique : ARN, ADN, peu de structures 3D valides mais possibilité de créer un modèle
  - Protéine : trop de degrés de liberté et plusieurs structures 3D possibles (coude, hélice, feuillet  $\beta$ ...)
- Structure d'une protéine :
  - Résolution expérimentale de la structure : RMN-2D ; Rayon X, Cryo EM...

- Récupération dans une base de données
- Modélisation par homologie

Modélisation Moléculaire



## 2.1.2 Bases de données

- Mise en commun de structures expérimentales au niveau mondial pour les chercheurs
- Principales Bases de Données :
  - Bases de données de séquences (26/09/2020)
    - UniProt: 563.082 séquences validées / 194.104.109 séquences brutes
    - <http://www.uniprot.org>
  - Bases de données de structures (26/09/2020)
    - Protein Data Bank: 169.117 structures
    - <http://www.rcsb.org/>



## 2.1.3 Modélisation par homologie

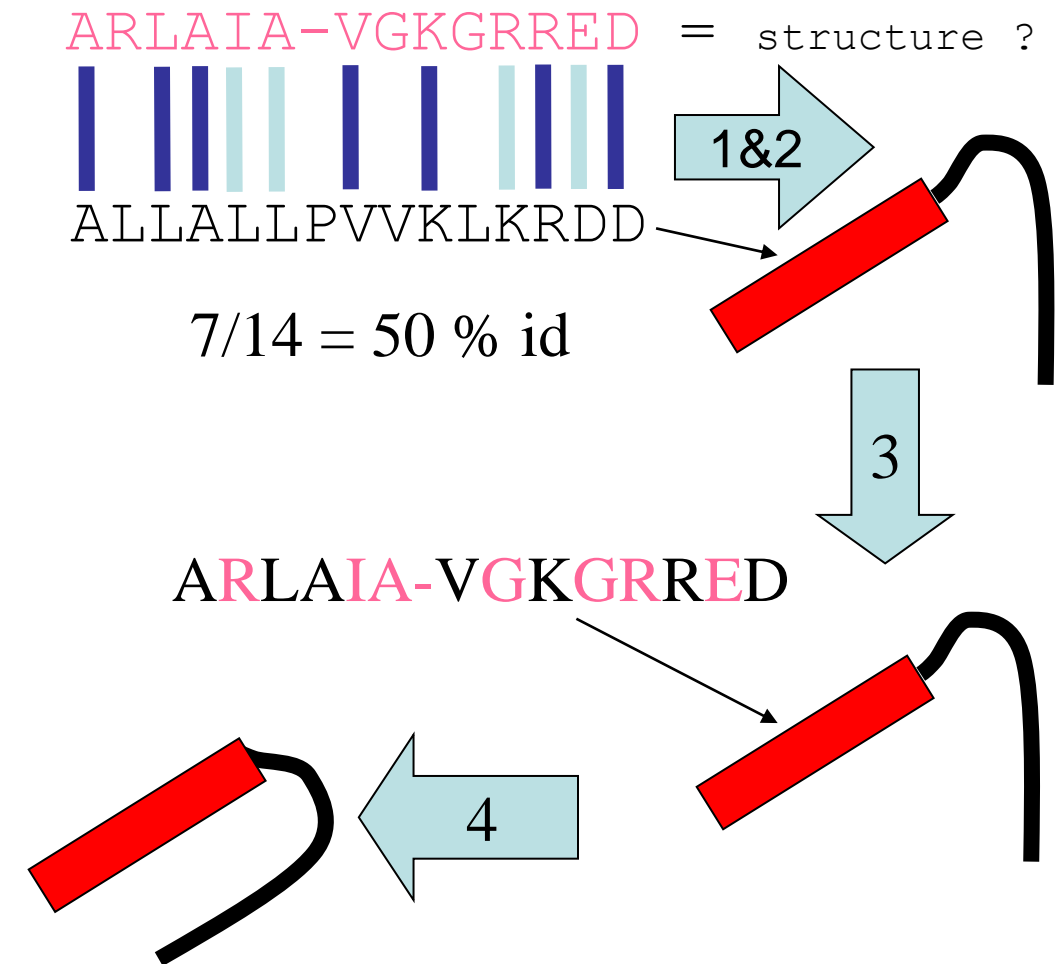
On veut construire un modèle 3D, mais nous n'avons que la séquence

1° Recherche d'une structure 3D (Fold) dans la PDB ayant la plus grande analogie de séquence (%id) (Même aa dans les deux séquences) avec votre séquence.

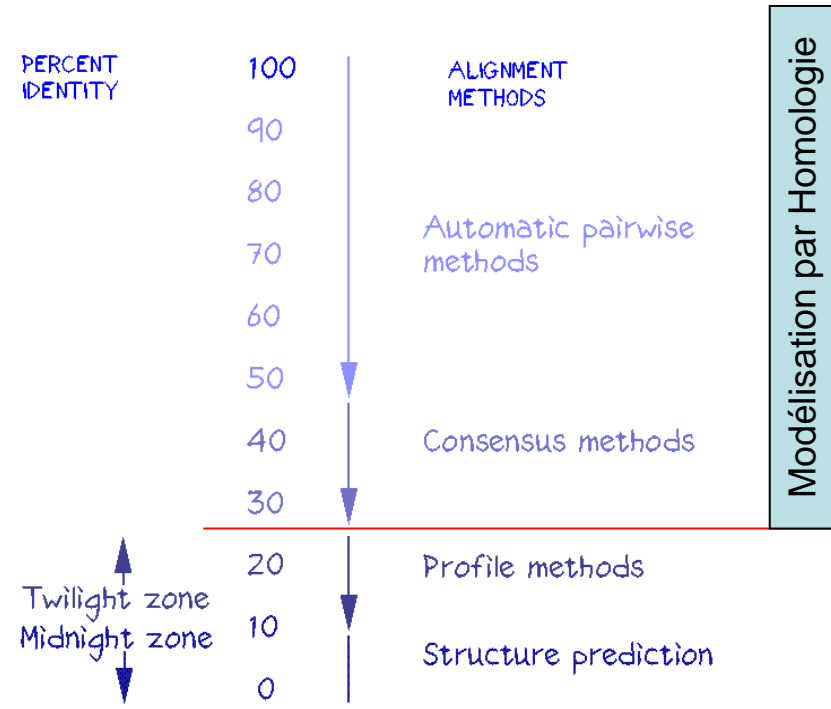
2° Alignement des deux séquences

3° Construction de votre Modèle à partir de la structure du « Fold »

4° Minimisation énergétique et validation du modèle.



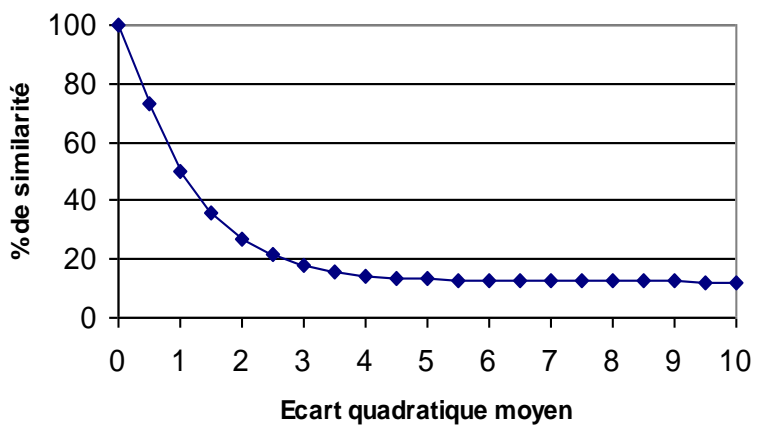
# 2.1.3 Modélisation par homologie



Modélisation par Homologie

## Limite de la méthode :

- Le %id entre la séquence de structure inconnue et la séquence de modèle du « fold » donne la qualité du modèle.
- Test à l'aveugle : La qualité du modèle est évaluée par l'écart quadratique moyen entre le modèle et la structure 3D expérimentale.
- Limite de la modélisation par homologie:
  - Forme de la protéine > 30%
  - Applications Docking > 60 ≈ 70%
- Bruits de fond d'alignement de 2 séquences ≈ 10%
- **Ne jamais utiliser cette méthode en dessous de 30% d'id.**



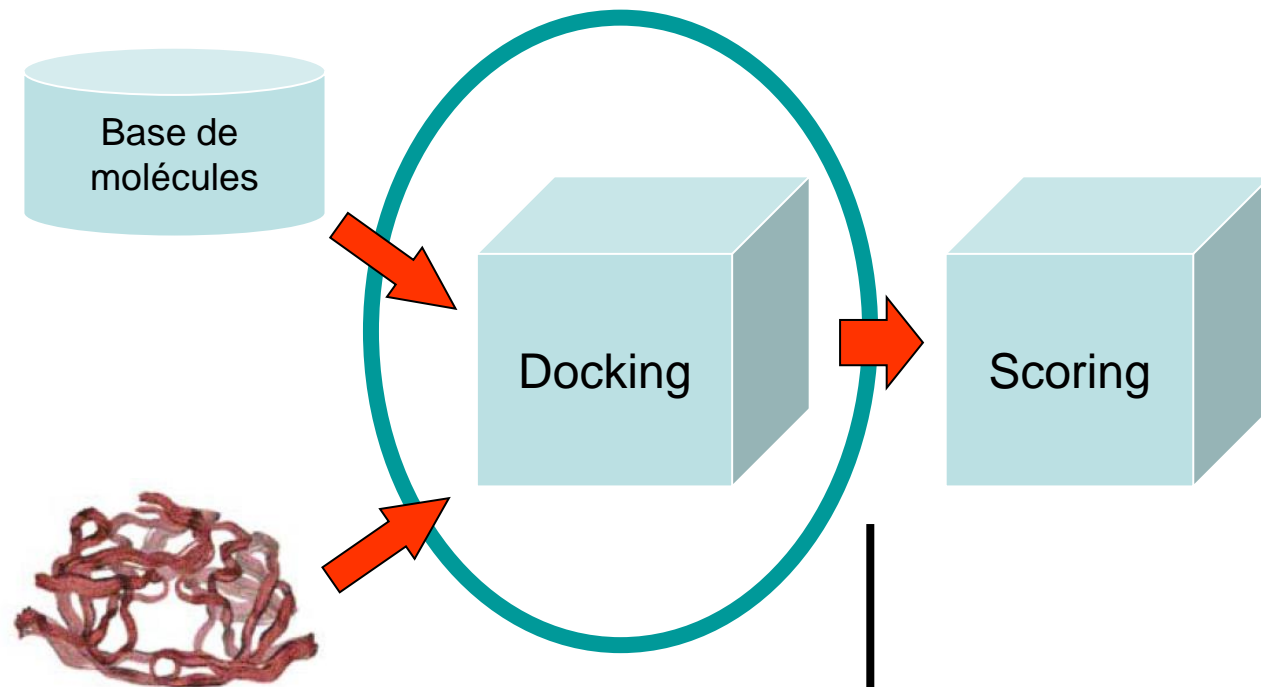
## 2.1.4 Docking

- But: Simulation de l'approche d'un ligand (Docking) et identification de la bonne conformation géométrique du ligand dans son site actif
- Parallèle : arrimage d'un bateau dans un port
- Se divise en 3 étapes :
  - Caractérisation du site actif
  - Positionnement du ligand dans le site actif
  - Évaluation des interactions entre le ligand et la protéine (définition d'un score)





## 2.1.4 Docking : étapes



1. Historique : Tout rigide
2. Molécule flexible
3. Molécule + chaînes latérales flexibles
4. Tout flexible

1. Historique : stérique + électrostatique
2. Fonctions avancées
3.  $\Delta\Delta G$

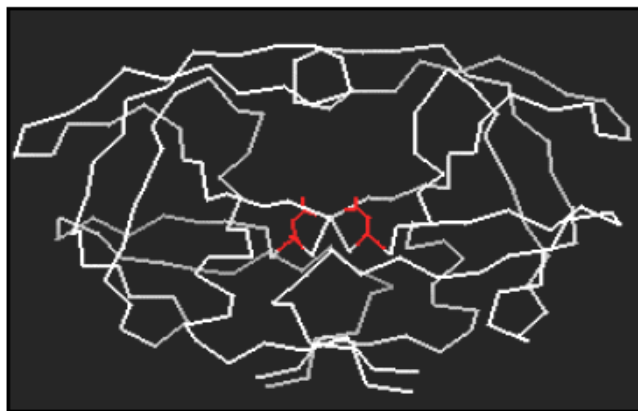


## 2.1.4 Le site actif

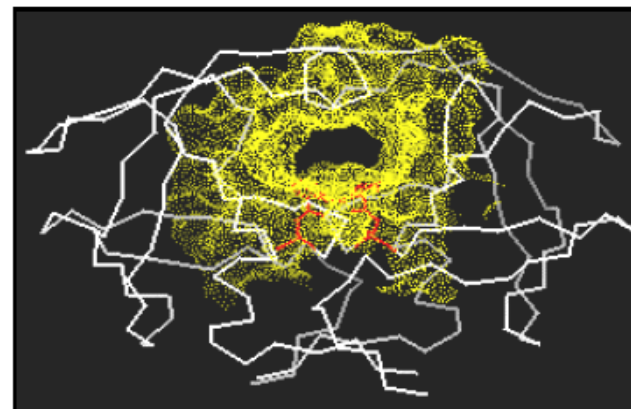
- Différentes méthodes :
  - Les cavités de la protéine (récepteur) sont utilisées pour définir une image négative du site actif se composant d'un jeu de sphères se superposant (DOCK)
  - Définition de descripteurs ensuite recherchés à la surface de la protéine: descripteurs chimiques (grpt méthyle, aromatique...) ou physico-chimiques (hydrophobicité, potentiel électrostatique)



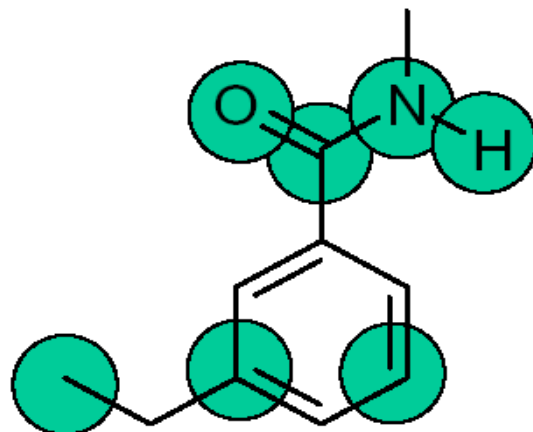
## 2.1.4 Description du site



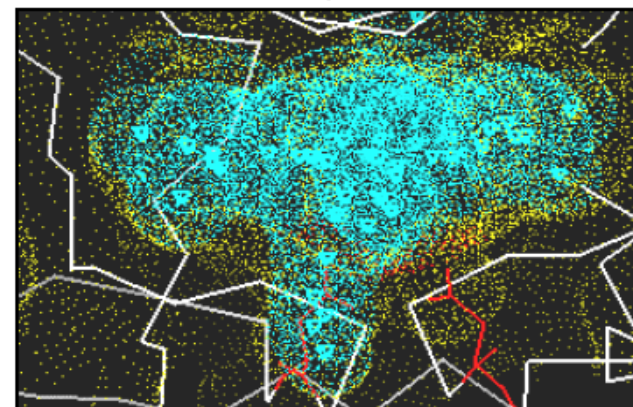
1. Structure 3D



2. Surface moléculaire



4. Superposition sphères-atomes



3. Remplissage de la surface par des sphères se chevauchant

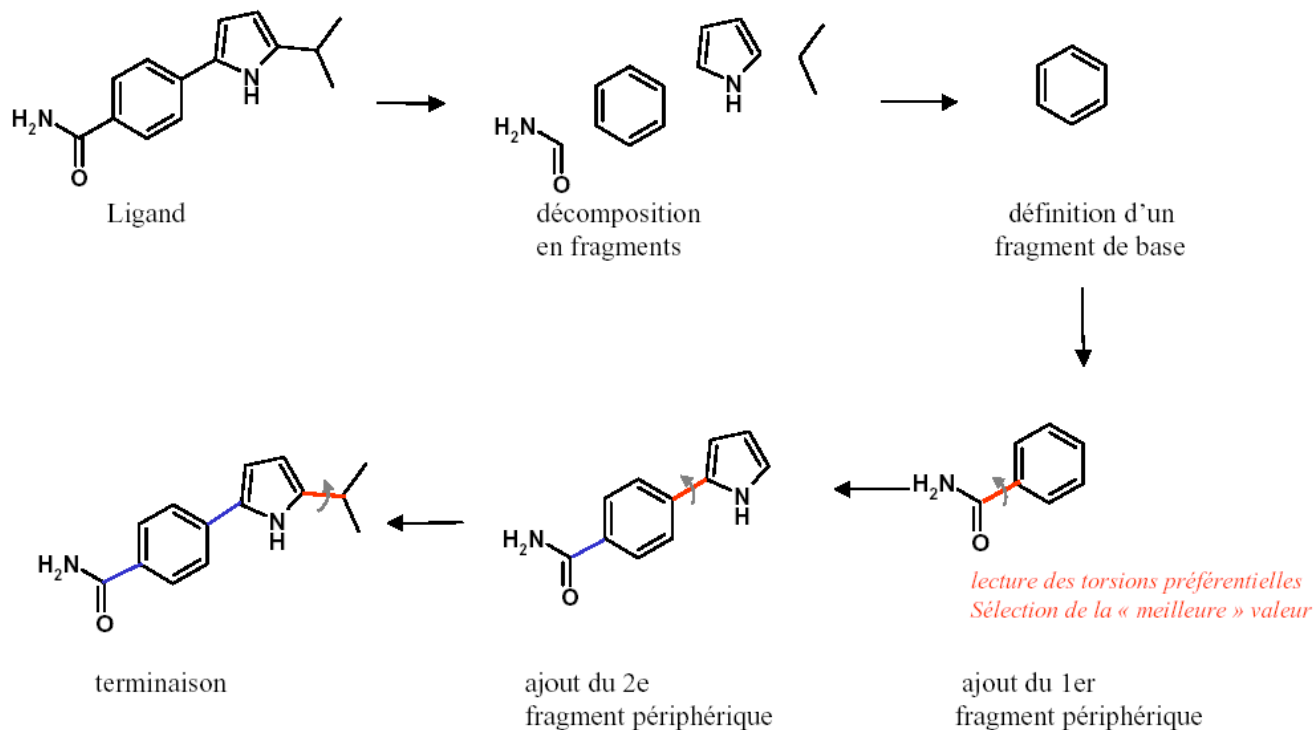
DOCK (<http://www.cmpchem.ucsf.edu/kuntz/>)



## 2.1.4 Positionnement du ligand

Il faut pouvoir orienter différentes conformations du ligand le plus vite possible

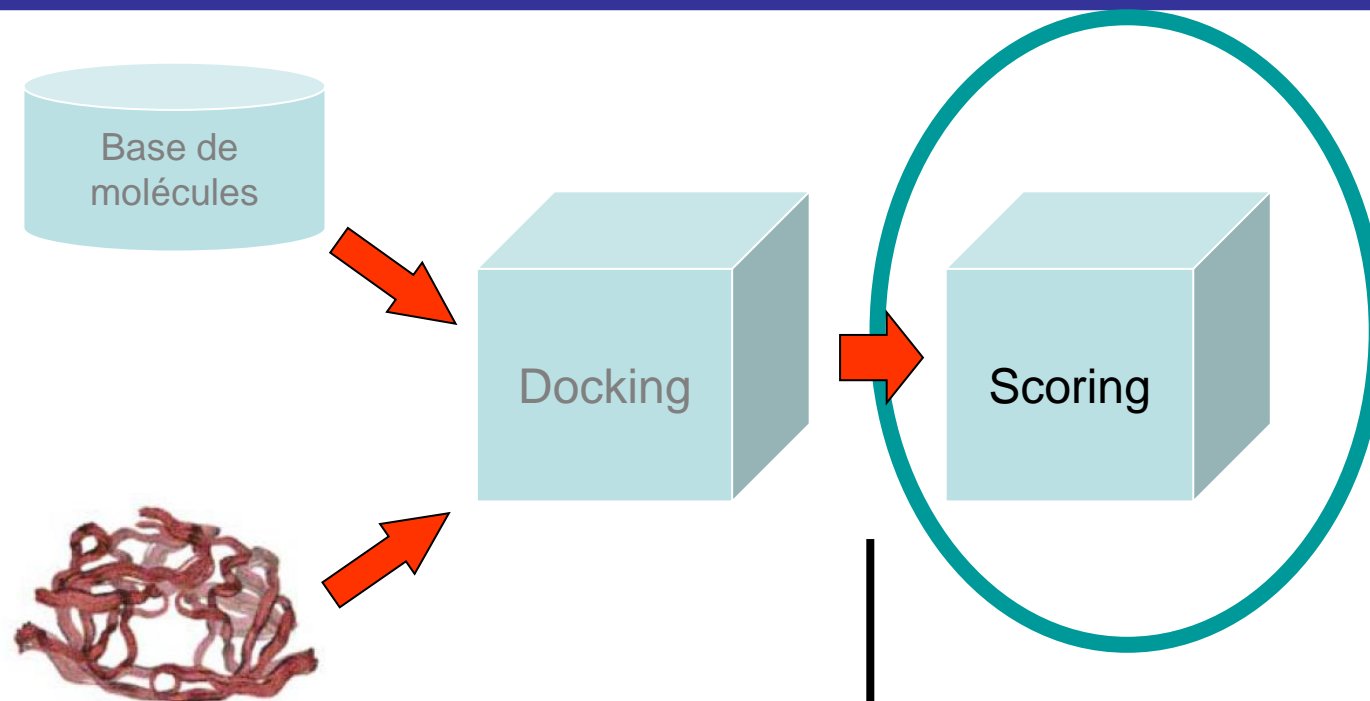
- par construction incrémentielle (e.g. FlexX, Dock)



- Identification d'un/ pl. fragments significatifs dans le ligand (ex: cycle d'un aa...)
- Placement des fragments dans le site actif en essayant de maximiser les contacts favorables
- Chaque orientation de ces fragments est le point de départ d'une étude conformationnelle du ligand entier
- Insensibilité de la conformation de départ
- Rapide (1 à 2 min par molécule)



## 2.1.4 Docking : étapes



1. Historique : Tout rigide
2. Molécule flexible
3. Molécule + chaînes latérales flexible
4. Tout flexible

1. Historique : stérique + électrostatique
2. Fonctions avancées
3.  $\Delta\Delta G$



## 2.1.4 Scoring

- Permet d'estimer la complémentarité ligand-protéine dans les complexes
- Le plus souvent = estimation du gain d'énergie libre du ligand en interaction avec le récepteur par rapport à la forme libre.
- Notation:
  - Basique: utilisant les champs de force
  - Utilisant une formule empirique
  - Plus évoluées dites « knowledge-based »

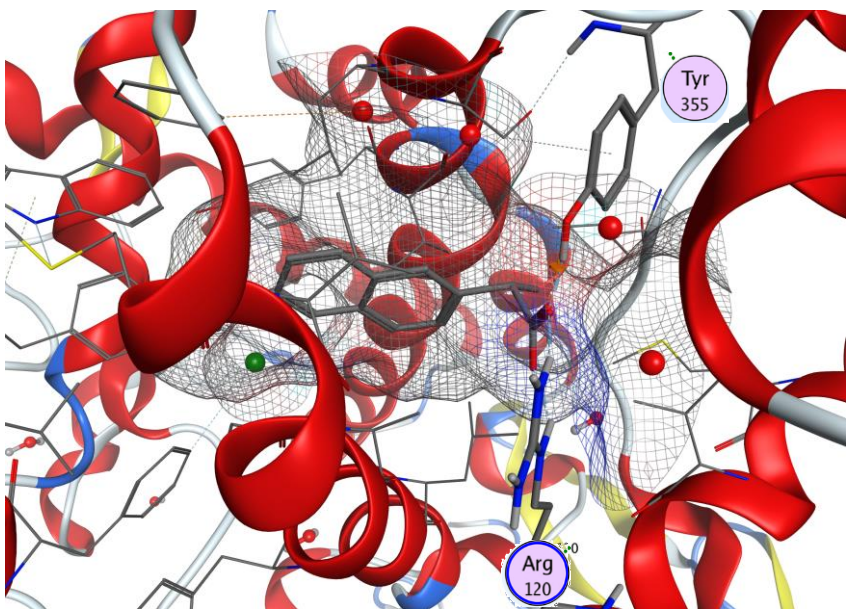


## 2.1.5 Exemple : A.I.N.S

- Les anti-inflammatoires non stéroïdiens, AINS, sont des médicaments aux propriétés:
  - analgésiques (réduisent la douleur)
  - antipyrétiques (réduisent la fièvre)
  - anti-inflammatoires (réduisent l'inflammation)
- Le terme « non stéroïdien » est utilisé pour les distinguer des glucocorticoïdes, qui ont une semblable action anti-inflammatoire.
- Les AINS sont des inhibiteurs de la cyclo-oxygénase COX. Ils agissent en inhibant la formation de prostaglandines et de thromboxane. Les COX sont aussi appelées Prostaglandin endoperoxyde H synthases (PGHSs)-1 et -2
- Les deux plus connus sont l'aspirine et l'ibuprofène.
- Problème : les inhibiteurs non sélectifs touchent COX-1, enzyme essentielle pour la protection gastrique :
  - saignements, perforation, ulcère de l'estomac
- Inhibiteurs spécifiques de COX-2 ⇒ **réduction d'effets secondaires**



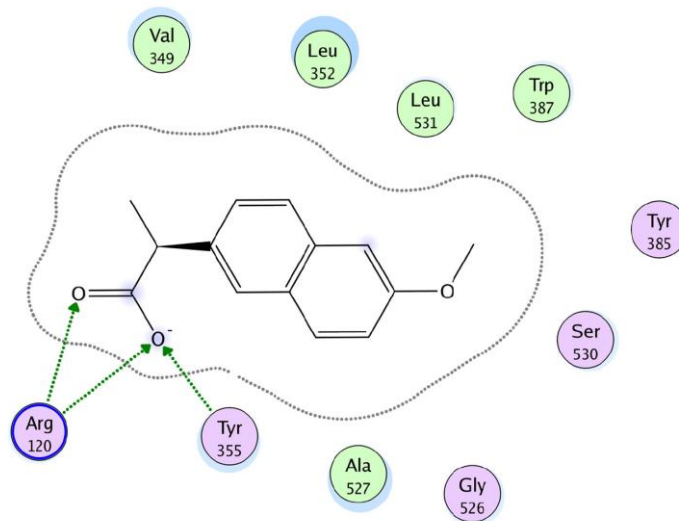
## 2.1.5 Exemple : A.I.N.S



Site en 3D de la poche (grille) du récepteur COX2, avec le naproxène (structure 3NT1)

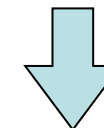
Structure cristallographique à haute résolution du complexe COX-2 avec le Naproxène (<https://www.rcsb.org/structure/3nt1>)

Obtention d'un système prédictif du  $pIC_{50}$  d'un inhibiteur en utilisant la structure du récepteur COX-2

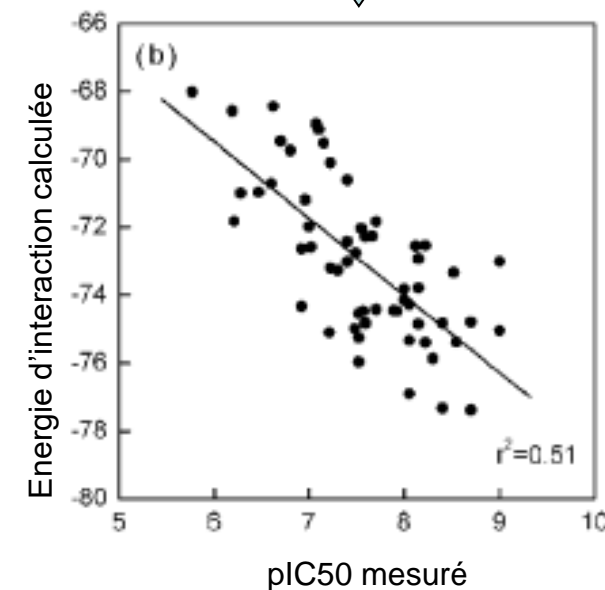
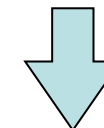


Calcul des interactions de la poche (pointillés) du récepteur COX-2, avec le Naproxène. Les liaisons hydrogène en pointillés verts sont critiques pour l'activité du Naproxène.

88 modèles d'inhibiteurs de COX-2



"DOCKING"





# Plan général



## 1. Partie I : Concepts & Théories

1. Conception d'une molécule active
2. Modèles et Modélisation Moléculaire: définitions et objectifs
3. Méthodes de calcul

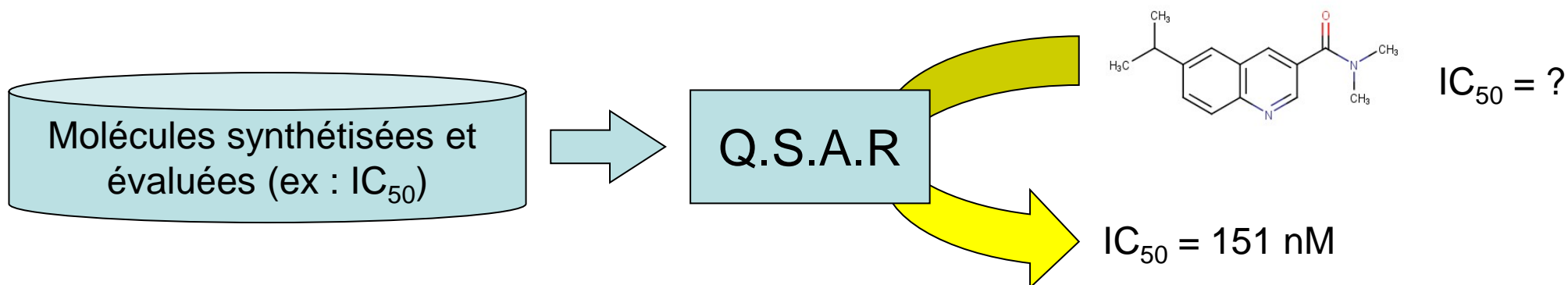
## 2. Partie II : Applications

1. Approche « récepteur »
2. Approche « sans récepteur »



## 2.2.1 Relation structure activité 2D

- But: Concevoir de nouvelles molécules virtuelles, par une relation activité/structure quantitative (QSAR), construite à partir d'un jeu de molécules synthétisées et testées expérimentalement.

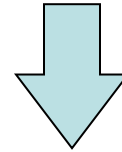


- Création du modèle de chacune des molécules et optimisation géométrique
- Description : chaque descripteur renvoie une valeur numérique à partir du modèle tridimensionnel d'une molécule. Il existe plusieurs types de descripteurs:
  - Spatial
  - Électronique
  - Topologique
  - Structurel
  - pKa
  - ...



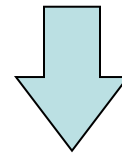
## 2.2.1 Relation structure activité 2D

Collection de molécules ayant ou non une propriété



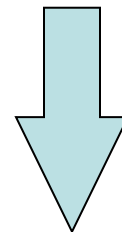
Modélisation

Collection de modèles



Description

Données numériques



Méthodes statistiques

Relation Structure Activité Quantitative



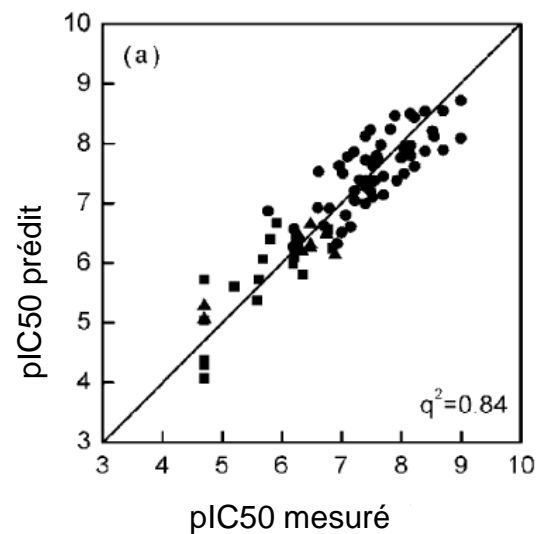
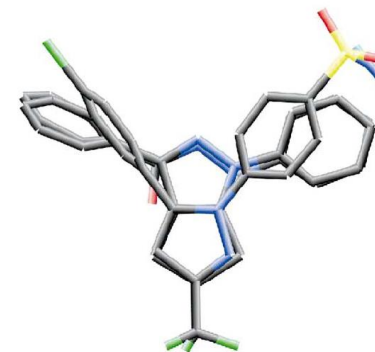
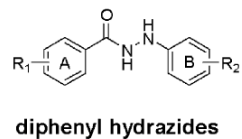
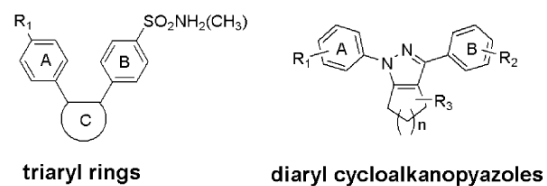
## 2.2.1 Relation structure activité 2D

- Plusieurs méthodes statistiques pour trouver la relation:
  - Simples
    - Résultats moins bons
  - PLS
    - Très bonnes méthodes pour les variables corrélées
  - Méthodes génétiques
    - GPLS ou Génétique
    - Excellentes méthodes
  - Apprentissage automatisé / Apprentissage Profond (Intelligence Artificielle)
- Deux phases :
  - **Calibration** (trouver la relation)
  - **Validation** (confirmer cette relation)



## 2.2.3 Exemple : A.I.N.S

Mesure du  $pIC_{50}$  d'une série de 88 composés spécifiques au COX-2



Q.S.A.R

Obtention d'un système prédictif du  $pIC_{50}$  d'un inhibiteur

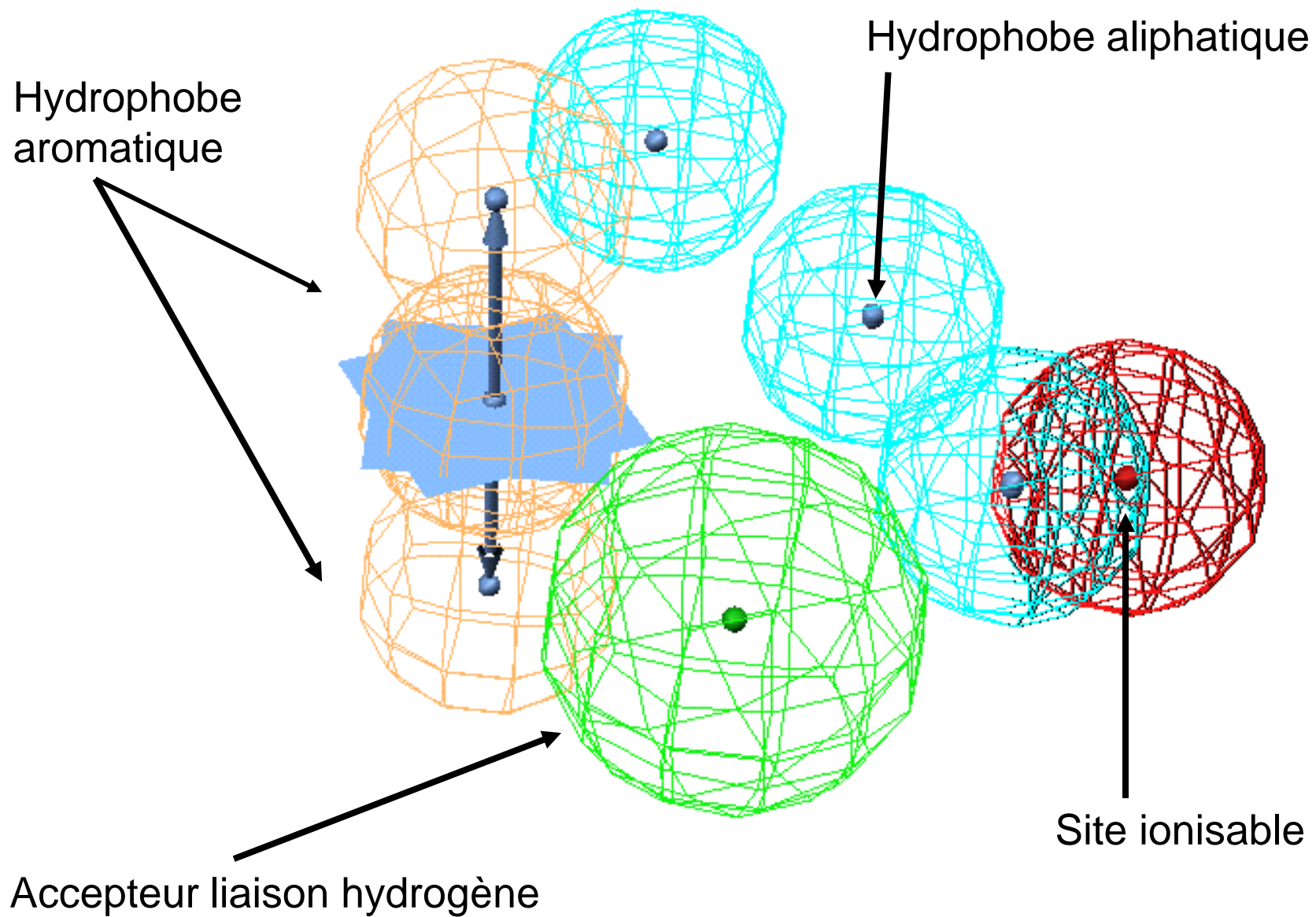


## 2.2.2 Pharmacophore

- Approche structure activité non quantitative
- Pharmacophore : Ensemble de fonctions chimiques réparties dans l'espace nécessaires à l'activité de la molécule
- Plusieurs fonctions :
  - Hydrophobe (aliphatique)
  - Hydrophobe (aromatique)
  - Charge négative
  - Charge positive
  - Groupe ionisable
  - Donneur liaison Hydrogène
  - Accepteur liaison Hydrogène
- Recherche dans une base de données de molécules ayant la propriété voulue.



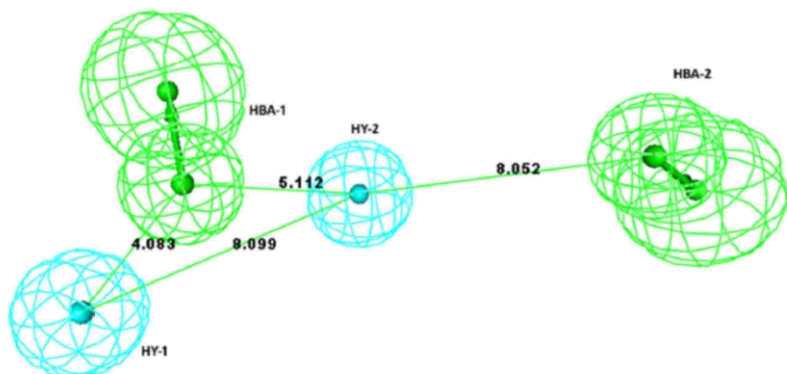
## 2.2.2 Pharmacophore



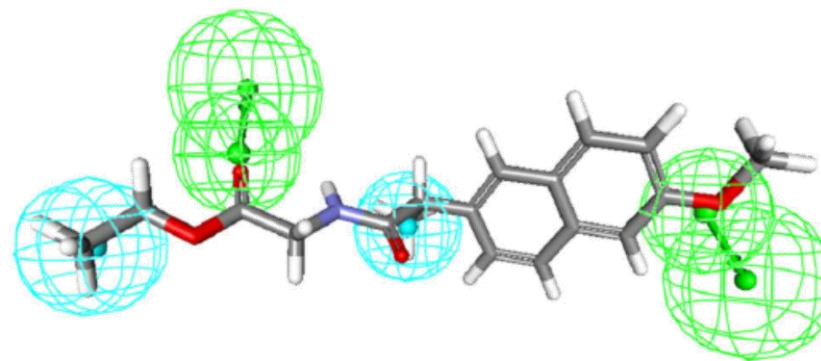
## 2.2.2 Pharmacophore

### Développement de nouvelles molécules à partir du Naproxène

- Création d'une base de données de 14 dérivés du naproxène avec leur  $IC_{50}$
- Détermination du Pharmacophore



Modèle de pharmacophore développé à partir des 14 molécules de l'étude.



Analogue du Naproxène (3a) superposé avec le modèle du pharmacophore

Compound	$IC_{50}$ ( $\mu$ M)	Fit value
Naproxen	1,150	2.1
3a	14.6	3.8
3b	123.4	2.4
3c	>250	2.3
3d	>250	2.1
3e	175.3	3.1
3f	209	2.9
3g	29.4	3.4
3h	102.1	2.5
3i	34.5	2.6
3j	74.1	2.7
3k	120.7	2.1
3l	49.3	3.9
3m	33.6	3.6

- Détermination de l'accord entre les molécules et le modèle de pharmacophore (Fit)
- Obtention d'un modèle servant à prédire l'activité ( $IC_{50}$ ) de molécules en vue de leur synthèse.

Novel non-cyclooxygenase inhibitory derivatives of naproxen for colorectal cancer chemoprevention.

T. Aboul-Fadl *et al.* Med Chem Res (2014) 23:4177–4188

