

**INITIATION
A LA
MODELISATION MOLECULAIRE**

Objectifs du cours :

- Développement de la notion de modèle
 - Application à la Modélisation Moléculaire
 - Initiation aux calculs de Mécanique Moléculaire
-

Plan du cours :

- I) Philosophie de la modélisation moléculaire
 - 1) La modélisation moléculaire : pour qui ? pour quoi ?
 - 2) Représentation des molécules : mythe ou réalité ?
 - 3) La modélisation moléculaire : comment ?
 - 4) Approches de la structure moléculaire
- II) La Mécanique Moléculaire (MM)
 - 1) Le champ de force
 - 2) Les calculs
- III) Le logiciel de modélisation moléculaire
 - 1) Construction, visualisation, manipulation
 - 2) Calculs
 - 3) Sauvegarde des structures
 - 4) Etude des propriétés moléculaires

I - Philosophie de la modélisation moléculaire.

1) La modélisation moléculaire : Pour qui ? Pour quoi ?

La modélisation moléculaire est un outil destiné aux chercheurs préoccupés par **la structure et la réactivité des molécules**

La connaissance de la structure des édifices moléculaires permet de **comprendre** ce qui est réalisé dans une transformation physique, chimique ou biologique. Elle peut permettre aussi de **prévoir** de telles transformations. La compréhension comme la prévision sont considérablement facilitées lorsque l'on peut **visualiser** les structures.

Une molécule est correctement décrite par sa géométrie et ses propriétés thermodynamiques. La visualisation doit rendre compte de l'ensemble de ces caractéristiques.

La question essentielle est **de représenter une molécule sur l'écran de la façon la plus proche possible de la "réalité"**.

L'utilisation de l'informatique a permis de mettre au point un outil performant :

la modélisation moléculaire.

2) Représentation des molécules mythe ou réalité ?

a) Quelques définitions :

Mythe = Construction de l'esprit qui ne repose pas sur un fond de réalité.

Réalité = Caractère de ce qui existe effectivement.

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

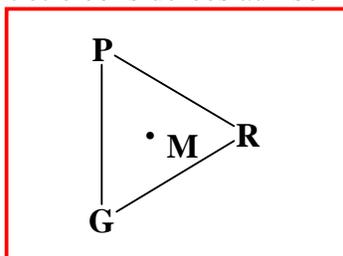
La modélisation a donc inévitablement un effet réducteur.

Modèle = Ensemble des paramètres et des fonctions mathématiques permettant une "représentation simplifiée" de la réalité

Le modèle (M) idéal serait :

Précis (P) , **Général (G)** et **Réel (R)**.

Ces trois propriétés doivent être considérées aux sommets d'un triangle :



Un modèle ne peut donc être à la fois précis, général et réel
il est donc toujours faux !

Il faut le considérer comme un moyen de "**scruter le réel**".

Modélisation moléculaire = Elaboration et application d'un modèle mathématique qui permet de représenter les molécules à l'échelle microscopique.

b) Conclusion

La représentation des molécules n'est pas un mythe... il y a un fond de vérité. Reste à en sonder la profondeur !

3) La modélisation moléculaire : Comment ?

a) Principe

Modéliser une molécule consiste à préciser, à **partir de calculs**, la position des atomes qui la constituent, dans l'espace et de calculer l'énergie de la structure ainsi engendrée. Une représentation "la plus proche possible de la réalité" correspondra à une structure de plus basse énergie

b) Calculs

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.

Les méthodes de calculs utilisées répondent plus ou moins bien à ces deux types de préoccupation :

- **la Mécanique Moléculaire (MM)** basée sur les calculs de mécanique classique qui permet de calculer l'énergie stérique du système. Elle utilise comme outil le "**champ de force**"

- **la Mécanique Quantique (MQ)** basée sur la résolution d'une équation différentielle fonction des seules coordonnées électroniques du système (équation de

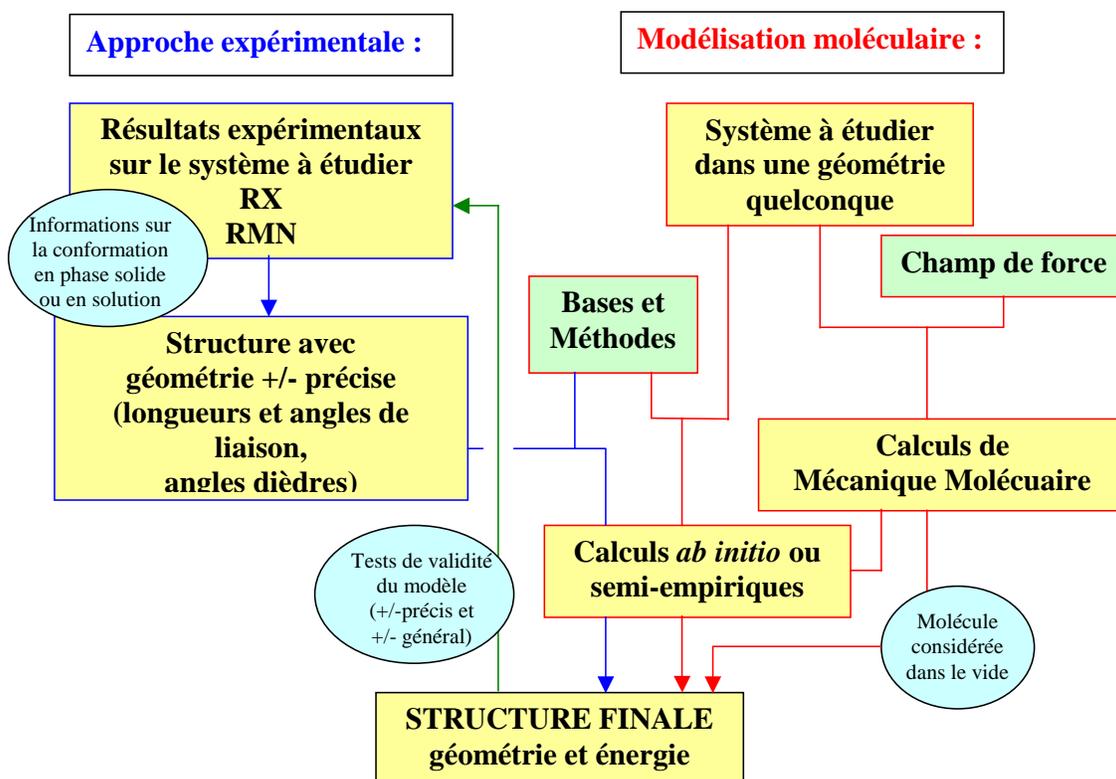
Schrödinger). Le principe de ces calculs est d'exprimer les orbitales moléculaires comme combinaisons d'orbitales atomiques ou "**bases**". La méthode de Hückel et la méthode de Hartree-Fock (calculs semi-empiriques et *ab initio*) mettent en jeu différentes approximations correspondant à différentes "**méthodes**". 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.

Le choix du type de calcul dépendra donc du problème étudié (degré de liberté du système et précision souhaitée du calcul) et évidemment des ressources de calculs (puissance de l'ordinateur) :

Limitation du nombre d'atomes par le potentiel de calcul de la machine

Coût en temps (et donc en argent) : pour n orbitales, il y a n^4 calculs *ab initio* et n^2 ou n^3 calculs semi-empiriques suivant les approximations admises; le nombre de calcul varie *linéairement* avec le nombre d'atomes en mécanique moléculaire.

4) Approches de la structure moléculaire



a) Approche expérimentale :

Il y a deux méthodes physiques qui fournissent les éléments indispensables à la connaissance de la géométrie moléculaire :

- **la structure aux RX** (lorsqu'elle est accessible!) fournit les paramètres de base (positions atomiques : longueurs et angles de liaison, angles dièdres) correspondant à la conformation **en milieu solide**.

- **les spectres de RMN**, par les constantes de couplage, les techniques bidimensionnelles et les NOE, permettent de reconstituer une structure tridimensionnelle correspondant à la conformation **en solution**.

Construite à partir de ces données, la structure peut être affinée par une minimisation par calculs de Mécanique Quantique.

b) Approche par Modélisation Moléculaire

A partir d'une structure quelconque du système étudié, **le calcul de l'énergie** est réalisé par **mécanique moléculaire** ou par **mécanique quantique**, les deux types de calcul pouvant être couplés. La **minimisation** de l'énergie permet une représentation probable. Celle-ci est obtenue indépendamment de toute interaction extérieure au système donc considérée **dans le vide**. Il est cependant possible par des techniques plus ou moins sophistiquées de tenir compte du milieu extérieur (constante diélectrique du milieu, interactions avec les molécules de solvant, ...).

c) Tests de validité du modèle

Il est important de valider le résultat des calculs par une comparaison des données structurales du modèle obtenu (angles et longueurs de liaison) avec les données expérimentales RX et RMN propres au système lorsque l'on dispose de ces données.

Si le modèle est validé, on peut admettre une bonne adaptation de la méthode de calculs utilisée au problème étudié et appliquer celle-ci à l'étude de structures analogues (hypothétiques ou dont on ne possède pas de données expérimentales).

II) La Mécanique Moléculaire (MM)

1) Le champ de force

L'outil de base de la mécanique moléculaire est :

Le champ de force :

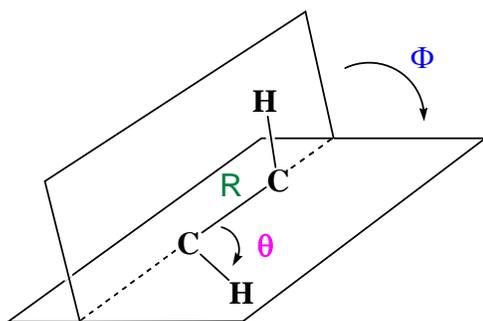
modèle mathématique représentant
aussi bien que possible les variations de **l'énergie potentielle**
avec la géométrie moléculaire.

a) Expression de l'énergie potentielle

On considère une molécule (en général dans le vide) comme un ensemble d'atomes sur lesquels s'exercent des forces élastiques et harmoniques; on ne s'intéresse qu'aux noyaux, les électrons sont ignorés. Chacune de ces forces est décrite par une fonction d'énergie potentielle. La combinaison de toutes ces fonctions est appelée "**champ de force**".

$$E = E_s + E_b + E_t + E_q + E_v + E_{sb} + \dots$$

Chaque terme représente la variation de l'énergie avec une variable du système :



- E_s** : énergie d'élongation
(longueur de liaison R)
- E_b** : énergie de liaison
(angle de valence θ)
- E_t** : énergie de torsion
(angle dièdre ϕ)
- E_q** : énergie électrostatique
(charge ou dipôle)
- E_v** : énergie de Van der Waals
(atomes non liés)
- E_{sb}** : terme croisé élongation - liaison

Exemple dans le champ de force d'Allinger (MM2):

$$E_s = C_s K_s (R - R_o)^2 (1 + C_b(R - R_o))$$

constantes (valables pour tous les atomes) : **C_s** , **C_b**

paramètres (spécifiques à la variable considérée) :

K_s : constante de rappel

R_o : distance standard entre les atomes liés

variable **R** : longueur de la liaison

b) La paramétrisation

- Elle est faite à partir des données expérimentales (RMN, RX, calculs ab initio) sur un ensemble le plus grand possible de molécules. Le choix des données expérimentales est important et le modèle obtenu en dépend étroitement. Les constantes sont ajustées pour rendre l'expression de E la plus générale possible. C'est le travail du concepteur du logiciel (informaticien-chimiste).

- Plusieurs types de champ de force existent et sont plus ou moins adaptés aux différents problèmes étudiés.

Exemples :

MM2, MM3, MM4, champs de force d'Allinger purement stérique spécialisés pour les petites molécules.

utilisés (typiques ou modifiés) par les logiciels :

Macromodel, PC-Modèle, MacMimic, Chem-3D, Hyperchem

Tripes, plus adapté aux molécules plus grosses

utilisé par les logiciels Alchemy et SYBYL

Amber, pour les macromolécules, adapté à la dynamique moléculaire

utilisé par les logiciels, Hyperchem, Macromodel.

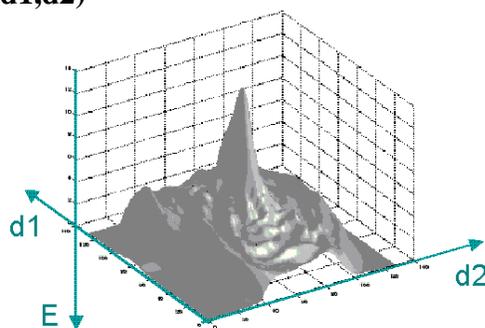
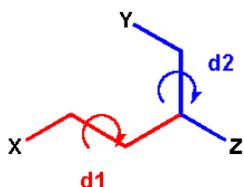
UFF (united force field), adapté aux molécules organométalliques

utilisé par le logiciel Cerius

c) Surface de potentiel

La variation de l'énergie d'une structure avec les n grandeurs définissant sa géométrie (longueurs de liaison, angles de valence, angles dièdres, ...) engendre une surface de potentiel à n dimensions. Si l'étude est faite à partir de deux variables indépendantes (par exemple : $d1$ et $d2$, deux angles dièdres) on peut représenter cette surface en 3 dimensions :

$$E = f(d1, d2)$$



2) Calculs de mécanique moléculaire (M.M.)

L'utilisation du champ de force pour le calcul de l'énergie E d'une structure est appelée : "**Mécanique Moléculaire**" (M.M.)

Ces calculs permettent de rechercher l'état de plus basse énergie par minimisation de E en ajustant les variables.

a) Signification de l'énergie E :

L'énergie E est appelée **énergie stérique**. Elle est en fait :

la différence d'énergie entre la molécule "modèle" et une molécule hypothétique dans laquelle toutes les variables auraient leur valeur "naturelle" ou "de référence".

Cette énergie n'a donc pas une signification physique bien définie.

Empiriquement, les auteurs des champs de force admettent que E représente l'énergie potentielle de la molécule immobile à 0°K (associée aux interactions entre atomes) plus l'énergie de

vibration $E_{\text{vibration}}$ puisque le champ de force est établi à partir de données expérimentales moyennes à T non nulle.

$$E = E_{0K} + E_{\text{vibration}}$$

b) Utilisation de l'énergie :

- Stabilité relative des molécules

On peut rapprocher E de la chaleur de formation ΔH°_f et de l'énergie interne U :

$$\Delta H^\circ_f = E_{0K} + E_{\text{liaison}} + E_{\text{vibration}} + E_{\text{rotation}} + E_{\text{translation}}$$

Donc :
$$\Delta H^\circ_f = E + E_{\text{liaison}} + E_{\text{rotation}} + E_{\text{translation}}$$

$$\text{Avec } E_{\text{rotation}} + E_{\text{translation}} = 3kT$$

Or :
$$\Delta H^\circ_f = U + PV \quad \text{avec à volume constant } PV = kT$$

D'où
$$U = E + E_{\text{liaison}} + 2kT$$

La variation d'énergie stérique E entre deux états d'un système peut donc être considérée comme la somme de la variation de l'énergie interne et de la variation de l'énergie de liaison du système:

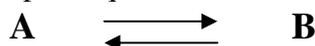
$$\Delta U = \Delta E + \Delta E_{\text{liaison}}$$

L'énergie de liaison n'étant pas quantifiable, la variation de l'énergie stérique ΔE pourra être utilisée comme critère de stabilité relative de deux structure si et seulement si $\Delta E_{\text{liaison}}$ est nul c'est à dire que ne pourront être comparées que des molécules isologues (ayant le même enchaînement d'atomes : conformères, stéréoisomères).

$$\Delta E_{\text{liaison}} = 0 \quad \Rightarrow \quad \Delta U = \Delta E$$

- Conformères ou diastéréoisomères en équilibre.

La proportion de deux structures isologues **A** et **B** en équilibre à une température T peut être calculer par l'équation ci-dessous si on admet que leur différence d'entropie est négligeable.



$$\Delta G = \Delta H - T\Delta S$$

$$[B]$$

$$K = \frac{[B]}{[A]} = e^{-\Delta G/RT}$$

$$\Delta H = \Delta(\Delta H^\circ_f) \text{ et } \Delta S \neq 0$$

$$\Delta G \neq \Delta H = \Delta E \text{ (kcal.mole}^{-1}\text{)}$$

$$R = 2.10^{-3} \text{ kcal.mole}^{-1} \text{ et } T = 293K$$

$$\text{Pour } \Delta E = -2 \text{ kcal.mole}^{-1} \quad [B] = 97\%$$

Compte tenu de l'erreur sur l'évaluation de E, on admet que pour deux structures présentant une différence d'énergie ΔE :

$$\Delta E > 2 \text{ kcal.mole}^{-1}$$

une seule structure existe

$$1 \text{ kcal.mole}^{-1} < \Delta E < 2 \text{ kcal.mole}^{-1},$$

les deux structures existent dans les proportions 70-80/30-20

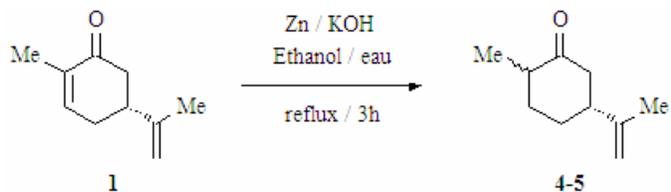
$$\Delta E < 1 \text{ kcal.mole}^{-1}$$

les deux structures sont en quantité égale

Exemple de problème structural

En Travaux Pratiques de Synthèse Organique, on étudie la sélectivité de deux réactions de réduction de la carvone 1 réalisées dans les conditions décrites ci-dessous. Dans les deux cas, les données spectrales permettent d'établir la structure plane des produits obtenus et les chromatogrammes indiquent qu'ils sont composés d'un mélange de deux diastéréoisomères.

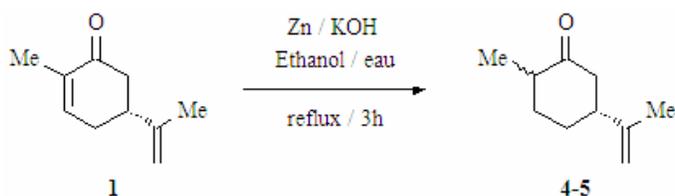
Réduction 1 :



IR : $\nu_{\text{CO}} = 1715\text{cm}^{-1}$
Masse : $M^+ = 152$
CPG : deux composés
R time 6:55 83%
7:00 17%

→ On obtient un mélange de *cis*- et *trans*-carvéols

Réduction 2 :

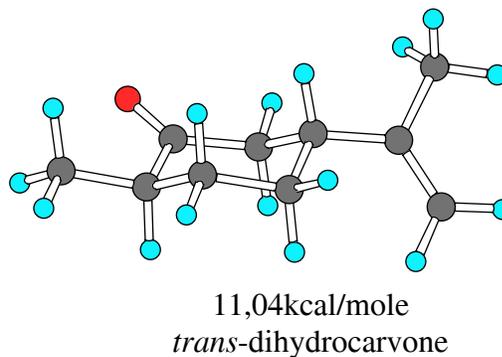
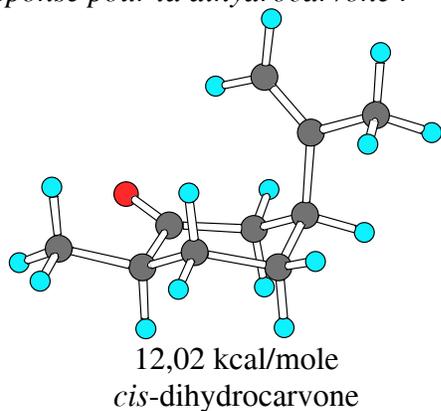


IR : $\nu_{\text{CO}} = 1715\text{cm}^{-1}$
Masse : $M^+ = 152$
CPG : deux composés
R time 6:55 83%
7:00 17%

→ On obtient un mélange de *cis*- et *trans*-dihydrocarvones

Dans chaque cas, les données expérimentales ne permettent pas de déterminer le diastéréoisomère majoritaire. La modélisation moléculaire peut-elle apporter une réponse à cette interrogation ?

Réponse pour la dihydrocarvone :



Les calculs, réalisés à l'aide du logiciel Chem3D utilisant le champ de force MM2 montrent que la trans-dihydrocarvone est le diastéréoisomère le plus stable avec une différence d'énergie d'environ 1 kcal/mole.

Si les deux diastéréoisomères sont en équilibre - c'est-à-dire si la réaction de réduction est **sous contrôle thermodynamique** - les calculs indiquent donc que la **trans-dihydrocarvone est le produit majoritaire** dans un mélange de proportion environ 80/20.

Ce résultat est en accord avec l'expérience :

- les conditions expérimentales (Zn/KOH - EtOH/eau - reflux/3h) correspondent à des conditions thermodynamiques
- le mélange obtenu est donc : **4/5 = 83/17**

III) Le logiciel de Modélisation Moléculaire

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

- 1 - Construction, visualisation et manipulation des molécules.
- 2 - Calculs
- 3 - Sauvegarde des structures et gestion des fichiers
- 4 - Etude des propriétés moléculaires.

1) Construction, visualisation, manipulation

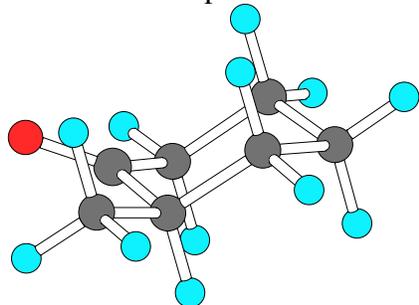
a) **La construction** des molécules peut se faire selon plusieurs méthodes :

- **à partir de fragments préexistants** (Chem 3D). Ceux-ci sont associés par substitution, la structure est alors élaborée par des modifications successives du squelette initial (type de liaison, type d'atome, stéréochimie). Les paramètres sont ajustés au fur et à mesure des substitutions et modifications.

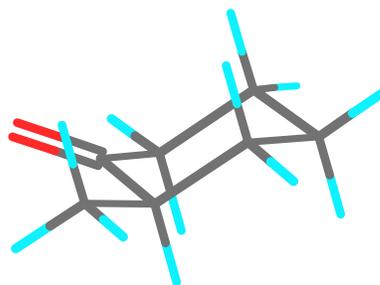
- **à partir d'une structure dessinée en deux dimensions** (Hyperchem). La transformation en structure en trois dimensions est ensuite obtenue par ajustement global des paramètres. La stéréochimie aléatoire peut alors être modifiée.

- **à partir de données géométriques** (coordonnées des atomes provenant d'une structure de RX) ou **à partir de bases de données courantes** (ex : Cambridge Structural Data Base, Brookhaven Protein Bank).

b) **La visualisation** des molécules peut être faite dans une grande variété de modes graphiques permettant dans chaque cas de mettre en valeur les informations demandées. Ci-dessous trois représentations de la 2-méthylcyclohexanone :

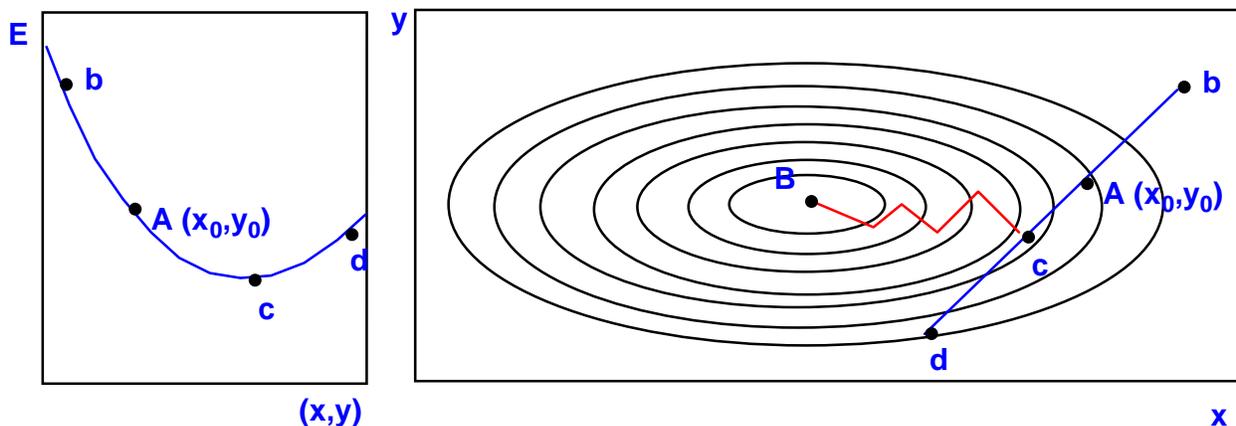


"Boules et bâtons"

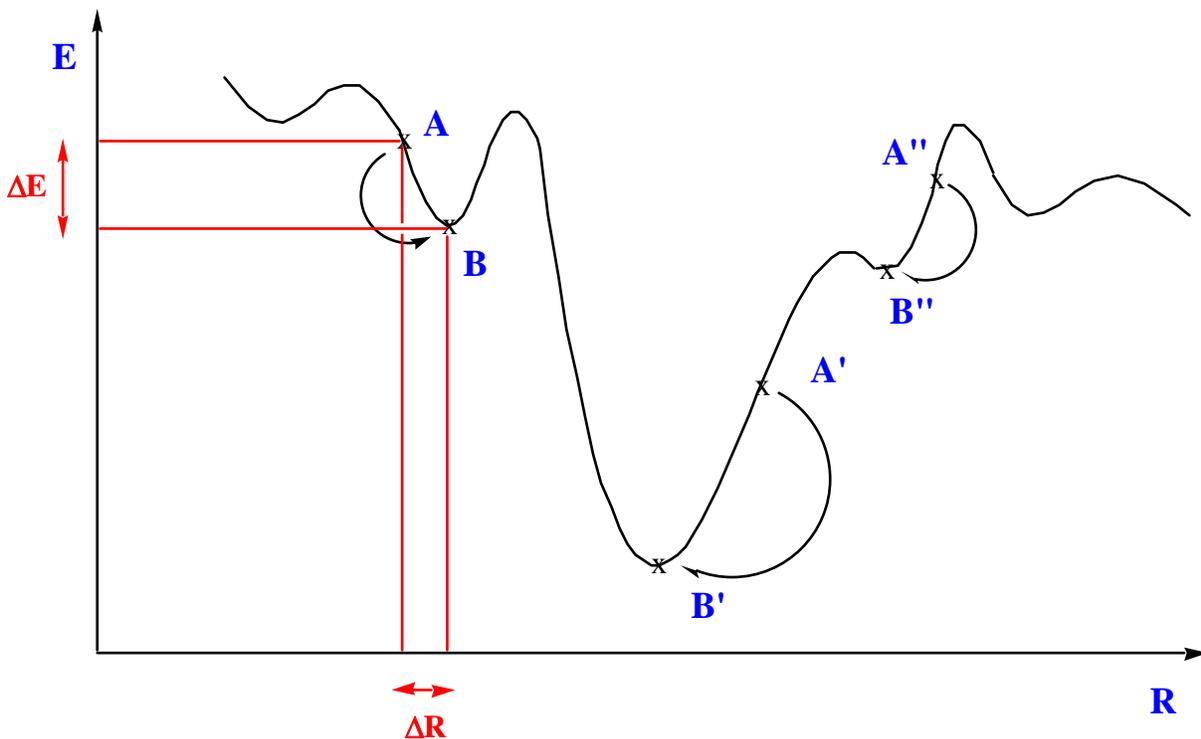


"Bâtons"

Surface d'énergie harmonique :



Le puits de potentiel **B**, résultat de la minimisation apparaît comme un "**minimum local**" qui dépend de la structure de départ **A** (x_0, y_0). Il est d'une part judicieux pour un gain de temps dans les calculs, de partir d'une structure fortement probable. D'autre part, par une simple minimisation, il est impossible de trouver la structure de plus basse énergie correspondant au "**minimum absolu**" comme le montre la figure suivante :



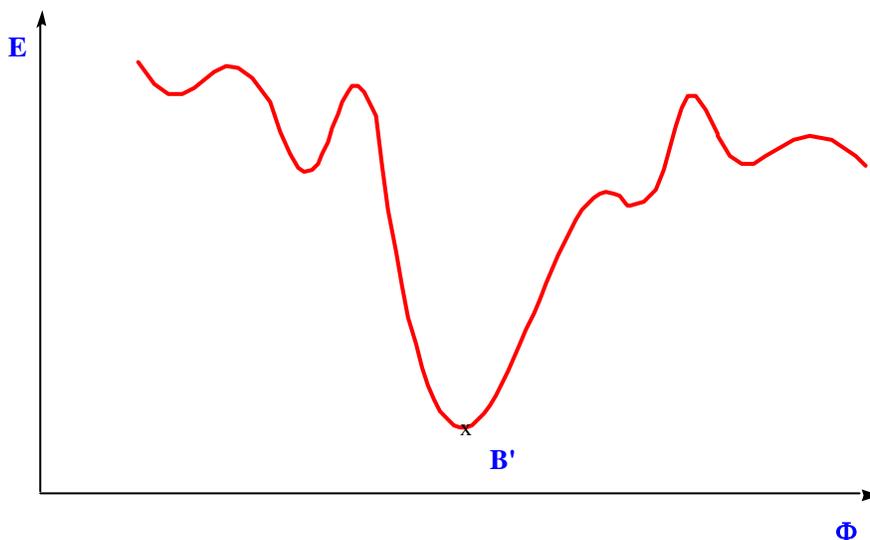
b) La recherche de conformères

La recherche du "**minimum absolu**" est le véritable problème de la Modélisation Moléculaire. L'exploration de tout l'espace conformationnel est limité par le potentiel de calcul de la machine et le temps. Cette recherche sera donc plus facile en Mécanique Moléculaire.

Toutes les variables doivent être considérées; mais de façon générale, l'angle dièdre est la variable la plus "molle" : sa déformation coûte moins cher en énergie que celle d'un angle de valence ou d'une longueur de liaison. La recherche de conformères s'effectue donc en ne faisant varier que les angles dièdres. Plusieurs méthodes sont utilisées :

Recherche "manuelle" ou systématique

Pour chaque dièdre de la structure, on décrit toute la courbe de potentiel. Ce qui permet d'en trouver **LE** minimum. Avec un grand nombre de dièdres à examiner, le temps de calcul devient rapidement prohibitif! La méthode n'est valable que si un nombre restreint de dièdres est responsable des problèmes stériques de la molécule (le terme "systématique" devenant alors abusif).

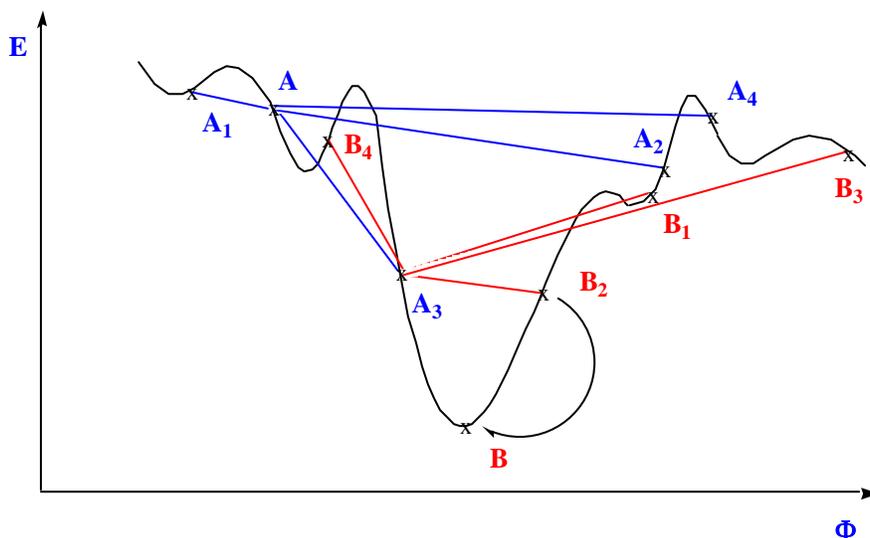


Recherche "automatique" basée sur l'algorithme Métropolis -Monte Carlo

Une série de grandes variations de tous les dièdres, effectuées de manière aléatoire (algorithme de Monte-Carlo) à partir d'un point **A** permet d'explorer toute la surface de potentiel en traversant les barrières d'énergie (**A**, **A₁**, **A₂**, **A₃**, **A₄**).

La structure de plus basse énergie **A₃** obtenue est alors choisie comme point de départ d'une nouvelle recherche (**B₁**, **B₂**, **B₃**, **B₄**). De proche en proche, le calcul permet de trouver une structure de basse énergie (**B₂**). Une minimisation conduit finalement au minimum recherché **B**.

Comme tout calcul aléatoire, plus le nombre de calculs est grand, plus le résultat est fiable.



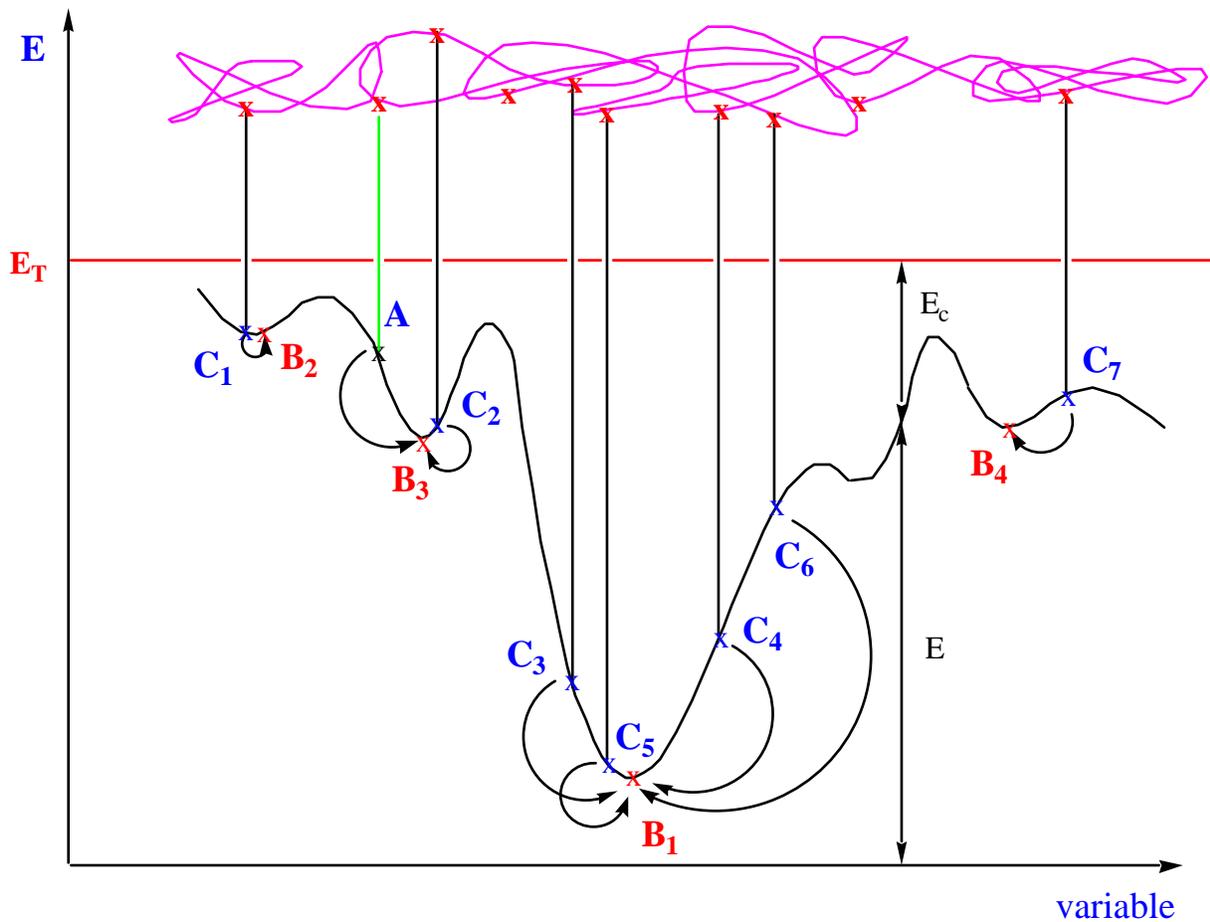
Dynamique moléculaire (D.M.)

On simule une **augmentation de la température** de la molécule (A). Les atomes entrent en mouvement (énergie cinétique E_C). Lorsque le système a atteint une certaine énergie totale E_T que l'on maintient constante, on laisse l'équilibre s'établir (l'équilibration peut aussi se faire à température constante - $E_C = cte$).

$$E_T = E + E_C$$

Le système va alors se **déformer et explorer** tout l'espace conformationnel. E_T doit être ajustée de telle sorte que l'équilibre puisse s'établir au-dessus de toutes les barrières de potentiel.

L'énergie cinétique devenant faible au voisinage de ces barrières, celles-ci deviennent difficilement franchissables et statistiquement, le système finit par évoluer essentiellement au-dessus des plus grands puits.



La simulation de ces deux phases (**augmentation de la température** et **équilibration**) est effectuées par calcul de mécanique classique (dynamique) :

$$\begin{aligned}
 x_A &= \text{coordonnée de l'atome A} \\
 v_A &= \text{vitesse de l'atome A} & x_A &= x_A^0 + v_A \Delta t \\
 a_A &= dv_A / dt = (m_A)^{-1} dE / dx_A
 \end{aligned}$$

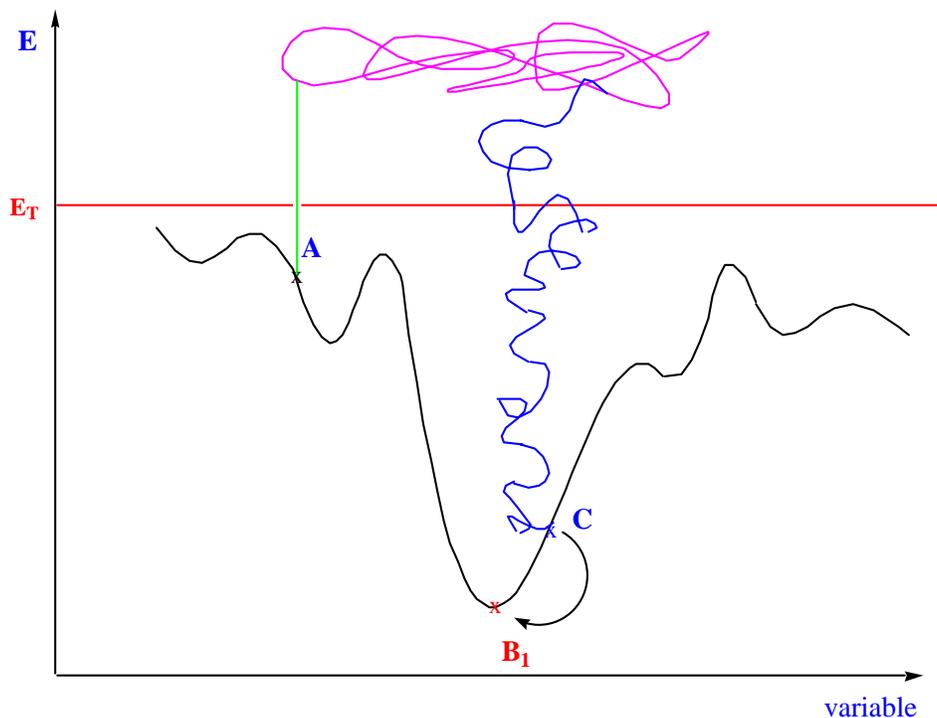
Il faut définir un pas de variation de temps (Δt_1) qui permet de recalculer les coordonnées des atomes au cours du temps. Δt_1 doit être choisi en fonction des constantes de forces et des masses atomiques qui définissent la fréquence des mouvements des atomes ($\approx 0,5$

femtoseconde)

Pour étudier les propriétés dynamiques de la molécule, on fait une collecte de données au cours de la phase d'équilibration. On définit un nouveau pas de variation de temps ($\Delta t_2 = n \Delta t_1$) qui fixe la fréquence de récupération des formes successives de la molécule. L'ensemble de ces données peut ensuite être étudié pour retrouver les différents conformères possibles de la molécule (C_1, C_2, C_3). Par minimisation, chaque conformère conduit une structure correspondant à un minimum d'énergie. Dans l'ensemble des structures obtenues ($B_1, B_2, B_3...$) doit se trouver le minimum absolu (B_1)

Recuit simulé

En utilisant le principe de la Dynamique Moléculaire, on peut rechercher directement le minimum d'énergie E correspondant au conformère le plus stable (B_1). La phase de **chauffage** et la phase d'**équilibration** sont plus rapides et au bout d'un temps défini, on simule un **refroidissement** (recuit simulé). Le système est alors piégé dans le puits de potentiel le plus profond. Une minimisation peut ensuite permettre d'atteindre le minimum (B_1).



Cette méthode met en oeuvre tout un arsenal de calculs nécessitant la mise au point de nombreux paramètres (température, temps, énergie totale, ...). Elle est très adaptée à la recherche de conformères de systèmes possédant un grand nombre de degrés de liberté (macromolécules) et permet d'étudier les propriétés dynamiques de ces molécules.

3) - La sauvegarde des structures

Les structures sont sauvegardées sous forme d'une matrice contenant les coordonnées des atomes et une table de liaisons.

- leur utilisation ultérieure dans le même logiciel
- l'utilisation de ces données dans d'autres programmes

4) - Etudes des propriétés moléculaires

A partir de calculs de Mécanique Moléculaire :

- mesures de géométrie : angles de valence, longueurs de liaison, angles dièdres,
- la caractérisation des centres asymétriques
- analyse des différentes composantes de l'énergie stérique (énergie d'élongation, de torsion, etc...)

A partir de calculs de Mécanique Quantique :

- analyse du calcul de l'énergie
- orbitales moléculaires
- potentiel électrostatique
- caractéristiques spectrales (IR, UV)

CONCLUSION

On peut en guise de conclusion citer G. SEGAL dans son article du *Technoscope de Biofutur* :

**"La modélisation moléculaire n'est pas la boîte de Pandore :
il en sort des choses passionnantes et non tous les maux de la terre.
Elle est peut-être une auberge espagnole :
ce qu'on y trouve dépend beaucoup de ce qu'on y apporte !"**

BIBLIOGRAPHIE

- G. SEGAL, "La modélisation moléculaire - Les logiciels : tendances et évolution",
Le Technoscope de Biofutur, n° 34, février 1990.
- J. S. LOMAS, "La mécanique moléculaire, une méthode non quantique pour le calcul de
la structure et de l'énergie d'entités moléculaires",
L'actualité chimique, mai 1986, 7.
- D.B. BOYD et K.B. LIPKOWITZ, "Molecular Mechanics",
Journal of Chemical Education, n°59, 1982, 269 - 274.
- J. P. BAYS, "So You Want To Do Molecular Modeling ?",
Journal of Chemical Education, n°69, 1992, 209 - 215.
- M. LOUP, "Le temps des modèles",
Le Journal du CNRS, n°50, 1994.
- K. N. KOUK, "Theory and Modeling of stereoselective Organic Reactions",
Science, vol 231, 1986, 1108.