

Introduction à la Modélisation Moléculaire

Georges Czaplicki, UPS / IPBS-CNRS
cgeorge@ipbs.fr, tél. : 05 61 17 54 04

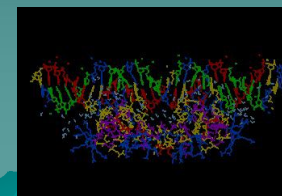
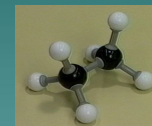
Qu'est-ce que c'est la MM ?

La **modélisation moléculaire** (MM) permet d'étudier la structure et la fonction de molécules grâce à la construction de modèles et au calcul de leur propriétés.

Les **modèles** peuvent être simple, comme une représentation "ball and stick", ou une image animée sur l'écran d'ordinateur.

Le **calcul** couvre les sujets divers :

- mécanique quantique *ab initio* et semi-empirique
- mécanique moléculaire empirique
- dynamique moléculaire
- méthodes statistiques (Monte Carlo)
- calcul d'énergie libre
- relation structure-activité (SAR)
- information chimique/biochimique
- bases de données
- utilisation de données expérimentales (RMN, RX) pour la détermination de structures moléculaires



Quels sont les domaines d'application de la MM ?

La MM cherche les réponses aux questions, concernant la nature et le fonctionnement de divers systèmes biologiques. Par exemple, elle peut faciliter la compréhension de :

- structures d'équilibre des sous-unités de biopolymères,
- aspects énergétiques de la formation de ponts d'hydrogène dans les protéines et les acides nucléiques,
- cinétique du repliement de protéines,
- le fonctionnement des agrégats supramoléculaires.

Les approches basées sur la modélisation sont nécessaires pour mieux comprendre le fonctionnement de systèmes biologiques et pour concevoir les théories, qui fourniront les prédictions (à tester expérimentalement !).

Quels sont les avantages de la MM ?

Pour comprendre le fonctionnement de molécules biologiques, il faut d'abord déterminer leurs structures tridimensionnelles. Les techniques expérimentales, comme la RMN ou les rayons X, nécessitent beaucoup de temps (et d'argent). Elles ont aussi leurs problèmes, qui limitent les champs de leurs applications.

Par conséquent, les approches basées sur la modélisation et le calcul deviennent les options extrêmement attractives.

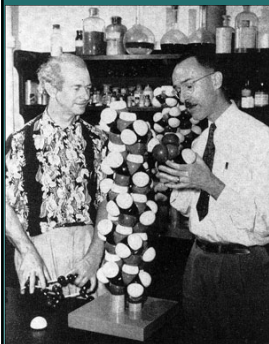
Un exemple : le développement de nouveaux médicaments.

La recherche de ligands de haute affinité pour les récepteurs sélectionnés passe par le criblage haut débit, implique la synthèse chimique et le filtrage de molécules, et finalement nécessite plusieurs années des essais cliniques.

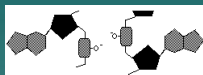
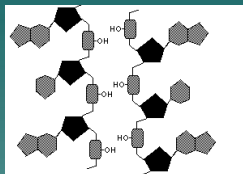
La totalité de la procédure peut durer jusqu'à 15 ans et coûter ~1 milliard d'euros. L'approche basée sur la MM peut raccourcir la procédure à quelques années seulement et permet d'économiser les sommes importantes.

Quels sont les dangers de la MM ?

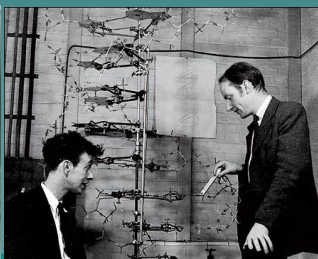
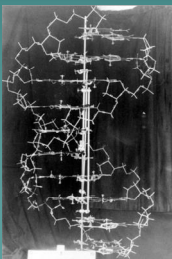
La MM est basée sur les multiples approximations. Leur connaissance détaillée est indispensable, surtout dans le contexte scientifique pertinent.



Pauling et Corey avec leur modèle de l'ADN à 3 hélices



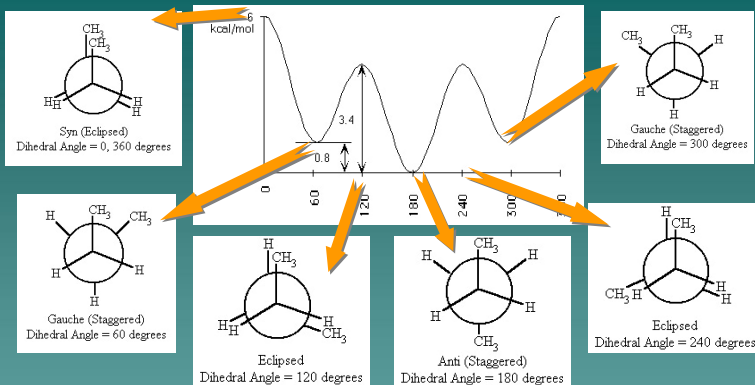
Watson et Crick avec le modèle de l'ADN à 2 hélices



Références générales

- T. Schlick, *Molecular Modeling and Simulation*. Springer-Verlag New York, Inc., 2002. [Interdisciplinary guide to biomolecular modeling]
- A. R. Leach, *Molecular Modelling. Principles and Applications*. Addison Wesley Longman, Essex, England, 1996. [Broad introduction to many aspects of molecular modeling and computational chemistry techniques]
- M.P. Allen and D.J. Tildesley, *Computer Simulation of Liquids*. Oxford University Press, New York, 1987. [Good advanced reference book for molecular simulations]
- G.E. Schulz and R.H. Schirmer, *Principles of Protein Structure*. Springer Advanced Texts in Chemistry, Springer-Verlag, New York, 1990. [Advanced text on the rapidly-changing field of protein folding]

Paysage énergétique de n-butane



In the example above, the g^* and g' minima are local, while the a minimum is global. The three structures corresponding to these minima are **conformers**. All other structures are **conformations** which may not be stable. The ratio of energy of conformers in all minima, combined with the Boltzmann distribution, will give us the ratio of populations of these conformers. Transitions between them are possible if the thermal energy is sufficient to overcome potential energy barriers separating the minima. If not, we observe oscillations and vibrations of molecules near the bottom of each potential well.

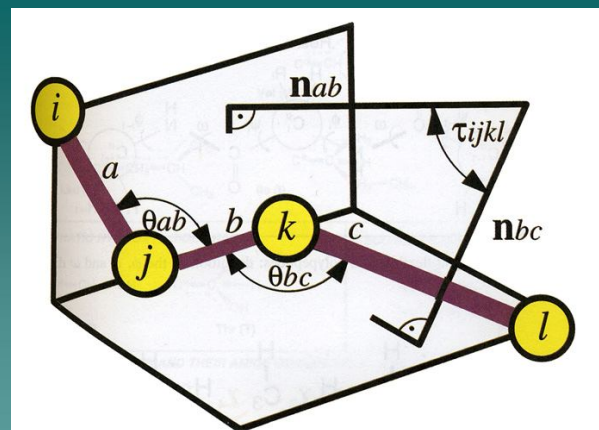


Figure 3.14. Definition of a dihedral angle $\tau_{ijkl} = \cos^{-1}(\mathbf{n}_{ab} \cdot \mathbf{n}_{bc})$, the angle between the two normals spanned by atoms i, j, k and j, k, l .

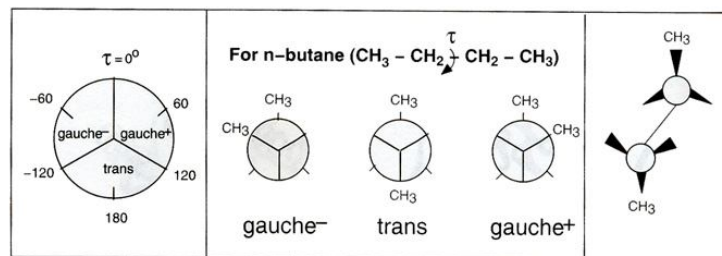
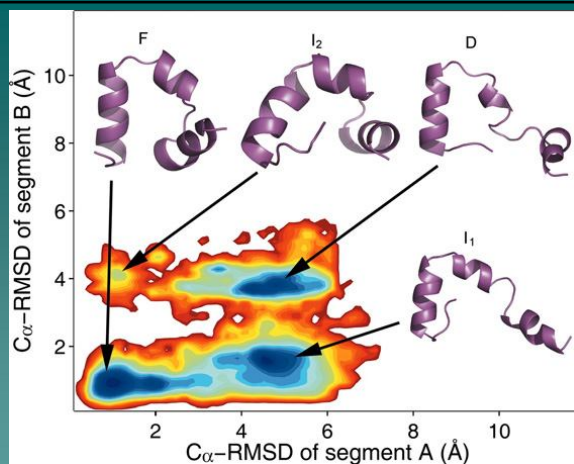
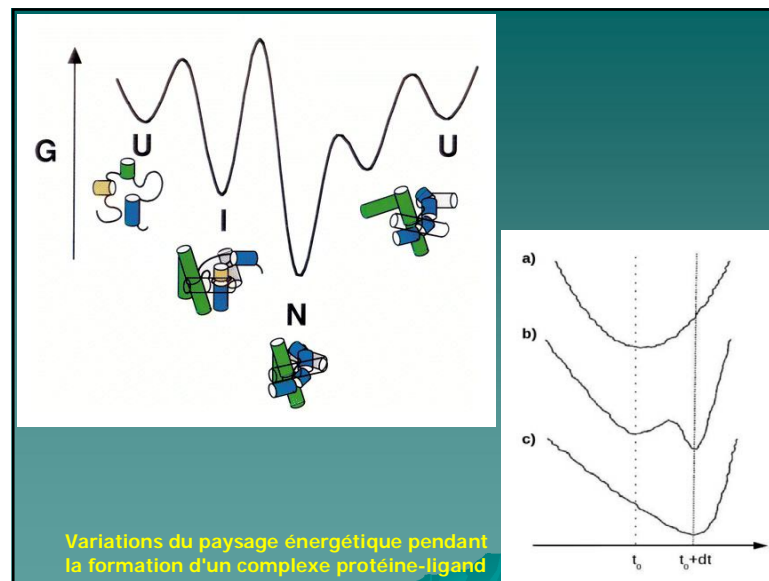
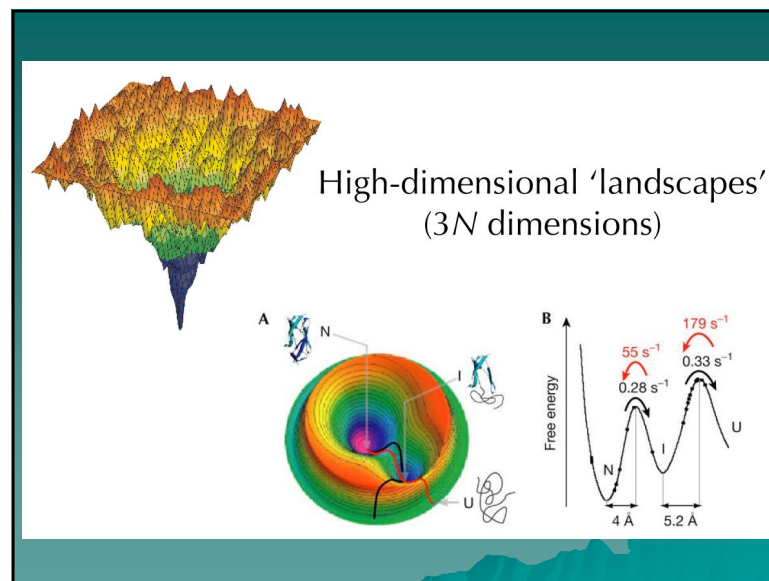


Figure 3.13. *Gauche* (g) and *trans* (t) dihedral-angle orientations for *n*-butane: (left) classification wheel; (middle) simple Newman projections that illustrate the three favored orientations of the two end methyl groups about the central C–C bond (perpendicular to the plane of the paper); and (right) the *trans* conformation, which has the least steric clashes.



The folding free-energy landscape of HP35 at 300 K from REMD. Representative structures of the four states are shown on the figure, including the folded state (F), denatured state (D), major intermediate state (I₁), and minor intermediate state (I₂).



Molecular Geometry and Energy

Molecular Mechanics (MM) focuses on determining parameters such as:

- molecular structures and energies,
- transition states,
- relative conformer populations and energies,
- harmonic vibrational frequencies.

Quantum mechanics (QM) should be able to provide tools for calculating molecular structures and energies. But even in the non-relativistic approach an accurate description of a multi-atomic system is too complex. In the Schrödinger equation:

$$H\psi(r, R) = E\psi(r, R)$$

the wave function ψ depends on nuclear coordinates R as well as on electronic coordinates r . In 1927 Born and Oppenheimer simplified the analysis by separating the movements of nuclei and electrons. With this approximation, two equations result:

1. potential energy surface

$$H\psi(r; R) = E\psi(r; R)$$

where the wave function ψ describes electrons and depends parametrically on nuclear coordinates R . This is the definition of the *potential energy surface*. There are two major approaches to solve this problem:

a) **ab initio calculations** (Gaussian, Cadpac, Hondo, GAMESS, Dmol, ...)

IF N is the number of atoms in a molecule, the computation time is roughly proportional to N^4 , which limits the applicability of this approach.

Molecular Geometry and Energy

b) **semi-empirical calculations** (CNDO, MINDO, MOPAC, AMPAC, ...)

Since some of the overlap integrals take time to calculate, a considerable gain is expected when integrals of the type:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi_1(1) \psi_2(1) \frac{1}{r_{12}} \psi_3(2) \psi_4(2) dx_1 dy_1 dz_1 dx_2 dy_2 dz_2$$

can be replaced by some empirical factors, carefully chosen to reproduce the *ab initio* results. In the above, the wave function subscripts refer to orbitals and (1) and (2) are electron indices. These integrals can be grouped in families and a decision has to be made which group to omit in the calculations. This justifies the existence of numerous semi-empirical methods.

2. quantum dynamics

The second equation gives us the structure and time evolution of a molecule:

$$H\Phi(R) = E\Phi(R)$$

The above describes the motion of nuclei on the electronic potential energy surface. But we need the empirical fit to the potential surface! Since nuclei move slowly, we can replace the Hamiltonian above by the Newton's equations of motion:

$$m \frac{d^2 R}{dt^2} = - \frac{dV}{dR}$$

Its solution with empirical fit to the potential surface V is called *molecular dynamics*.

Champs de force

1. Objectif: interpolation & extrapolation

2. Paramétrisation: ajustement multiparamétrique et vérification du modèle

3. Forme générale :

$$V = \sum V_{stretch} + \sum V_{bend} + \sum V_{torsion} + \sum V_{vdW} + \sum V_{Coulomb} + \dots$$

- Déviations par rapport aux paramètres "idéaux"

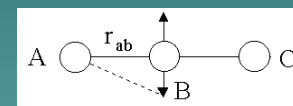
- Valide pour une classe particulière de molécules

Champs de force

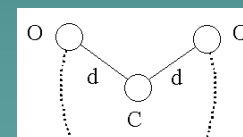
4. Sources de paramètres "idéaux"

- rayons X
- diffraction des électrons
- spectroscopie de micro-ondes
- diffraction de neutrons

- mouvements internes :



- "shrinkage" :



5. Limitations

Représentations des interactions

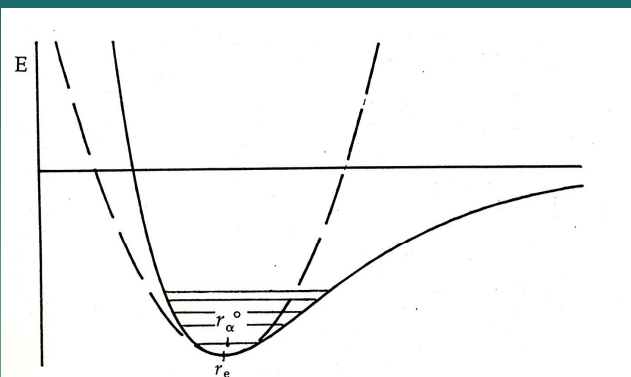
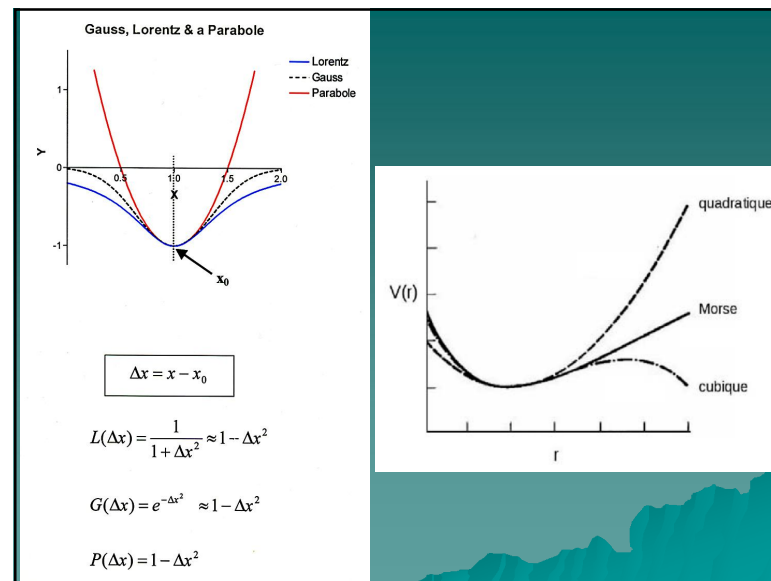


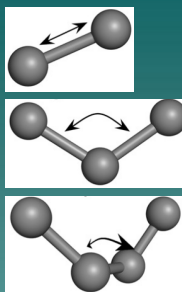
Figure 1.1. The energy–distance relationship between two atoms bonded together. Solid curve, Morse function; dashed curve, harmonic approximation.



Types des interactions

I. Interactions covalentes :

- "Bond stretching"
- "Angle stretching"
- Torsions
- interactions hors plan
- Couplage de coordonnés internes



II. Interactions non-liantes:

- interactions Van der Waals
- interactions électrostatiques

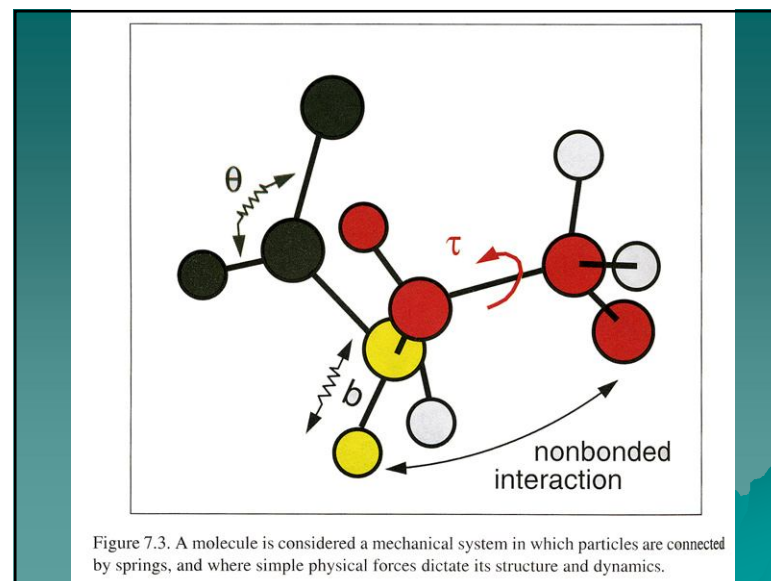
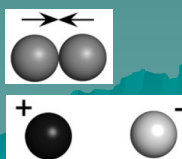
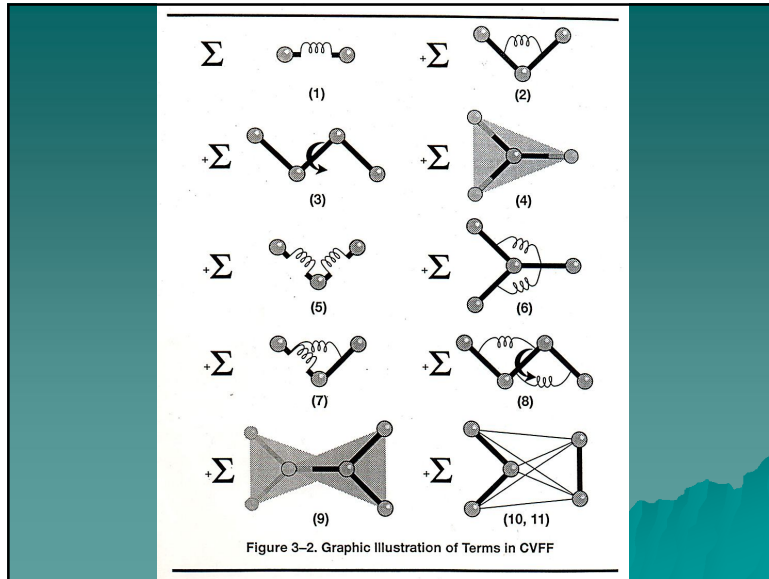
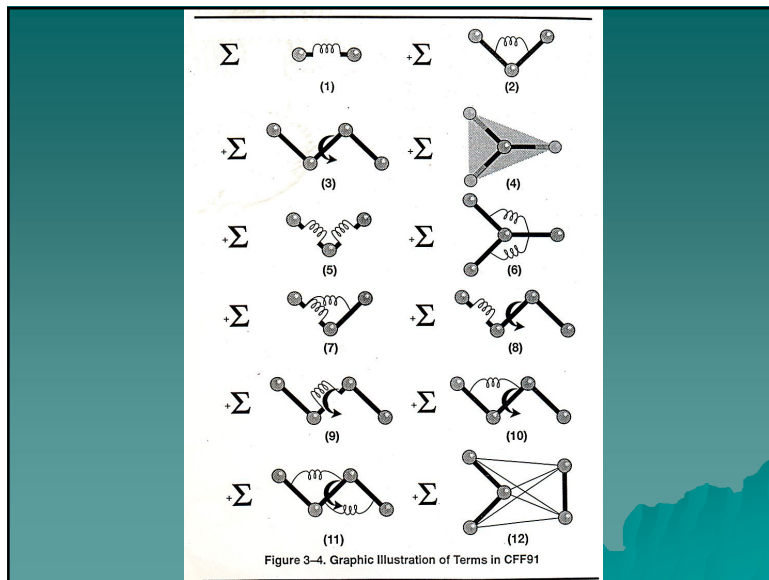


Figure 7.3. A molecule is considered a mechanical system in which particles are connected by springs, and where simple physical forces dictate its structure and dynamics.

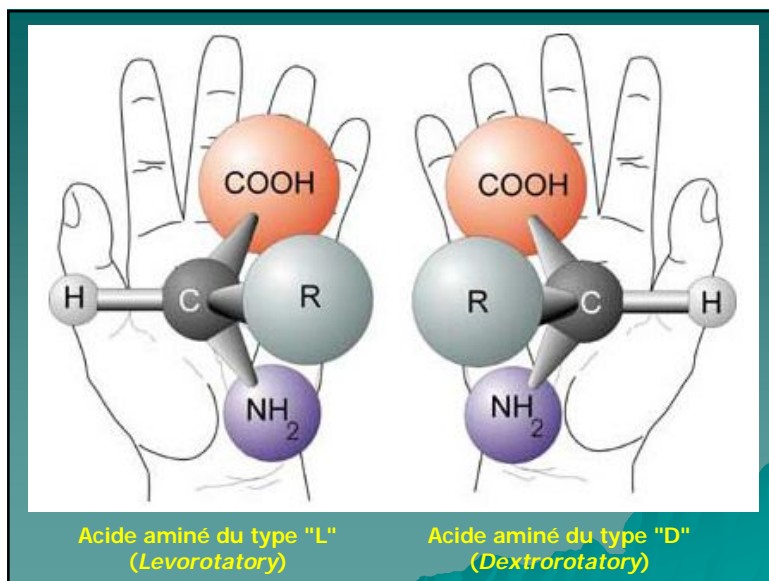
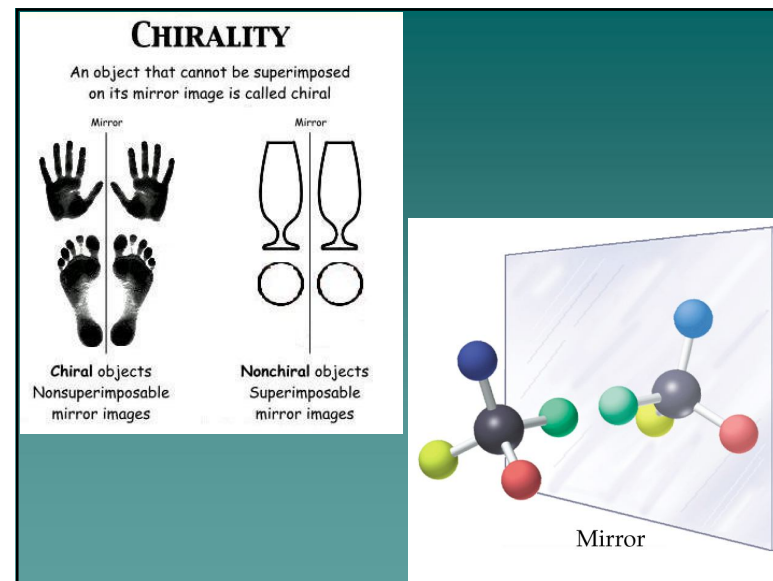
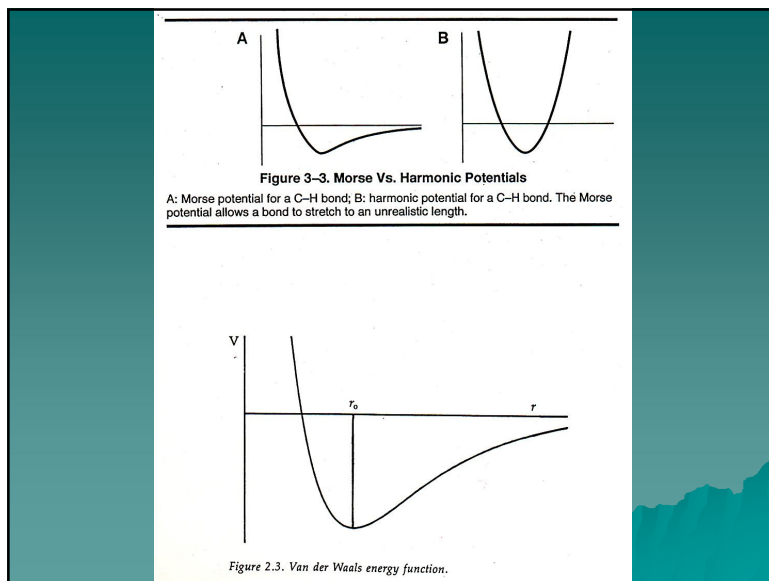


$$\begin{aligned}
 E_{pot} = & \sum_b [K_2 (b - b_0)^2 + K_3 (b - b_0)^3 + K_4 (b - b_0)^4] \quad (1) \\
 & + \sum_\theta H_2 (\theta - \theta_0)^2 + H_3 (\theta - \theta_0)^3 + H_4 (\theta - \theta_0)^4 \quad (2) \\
 & + \sum_\phi [V_1 [1 - \cos(\phi - \phi_0^0)] + V_2 [1 - \cos(2\phi - \phi_2^0)] + V_3 [1 - \cos(3\phi - \phi_3^0)]] \quad (3) \\
 & + \sum_\chi K_\chi \chi^2 + \sum_b \sum_{b'} F_{bb'} (b - b_0) (b' - b'_0) + \sum_\theta \sum_{\theta'} F_{\theta\theta'} (\theta - \theta_0) (\theta' - \theta'_0) \quad (4) \quad (5) \quad (6) \\
 & + \sum_b \sum_\theta F_{b\theta} (b - b_0) (\theta - \theta_0) + \sum_b \sum_\phi (b - b_0) [V_1 \cos \phi + V_2 \cos 2\phi + V_3 \cos 3\phi] \quad (7) \quad (8) \\
 & + \sum_{b'} \sum_\phi (b' - b'_0) [V_1 \cos \phi + V_2 \cos 2\phi + V_3 \cos 3\phi] \quad (9) \\
 & + \sum_\theta \sum_\phi (\theta - \theta_0) [V_1 \cos \phi + V_2 \cos 2\phi + V_3 \cos 3\phi] \quad (10) \\
 & + \sum_\phi \sum_\theta \sum_{\theta'} K_{\phi\theta\theta'} \cos \phi (\theta - \theta_0) (\theta' - \theta'_0) + \sum_{i>j} \frac{q_i q_j}{\epsilon r_{ij}} + \sum_{i>j} \left[\frac{A_{ij}}{r_{ij}^9} - \frac{B_{ij}}{r_{ij}^6} \right] \quad (11) \quad (12) \quad (13)
 \end{aligned}$$



$$\begin{aligned}
 E_{pot} = & \sum_b D_b [1 - e^{-\alpha(b - b_0)}]^2 + \sum_\theta H_\theta (\theta - \theta_0)^2 + \sum_\phi H_\phi [1 + s \cos(n\phi)] \quad (1) \quad (2) \quad (3) \\
 & + \sum_\chi H_\chi \chi^2 + \sum_b \sum_{b'} F_{bb'} (b - b_0) (b' - b'_0) + \sum_\theta \sum_{\theta'} F_{\theta\theta'} (\theta - \theta_0) (\theta' - \theta'_0) \quad (4) \quad (5) \quad (6) \\
 & + \sum_b \sum_\theta F_{b\theta} (b - b_0) (\theta - \theta_0) + \sum_\phi F_{\phi\theta\theta'} \cos \phi (\theta - \theta_0) (\theta' - \theta'_0) + \sum_\chi \sum_{\chi'} F_{\chi\chi'} \chi\chi' \quad (7) \quad (8) \quad (9) \\
 & + \sum \epsilon [(r^*/r)^{12} - 2(r^*/r)^6] + \sum q_i q_j / \epsilon r_{ij} \quad (10) \quad (11)
 \end{aligned}$$

CVFF forcefield



Planarity and chirality of atoms

The volume of a tetrahedron whose vertices are the four atoms :

$$V = \frac{1}{6} \det \begin{bmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & x_3 & x_4 \\ y_1 & y_2 & y_3 & y_4 \\ z_1 & z_2 & z_3 & z_4 \end{bmatrix}$$

The above is known as the **oriented volume**.

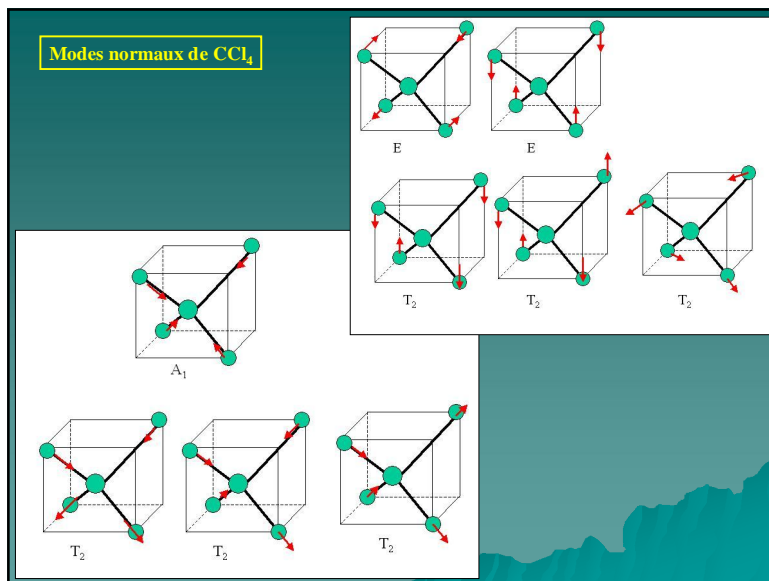
When all atoms are planar, volume $V = 0$.

Whenever the tetrahedron is reflected in a plane (x,y, or z coordinates multiplied by -1), the sign of oriented volume changes.

Hence, the chirality of the quadruple of atoms bonded to an asymmetric carbon atom, defined as

$$\chi_{1234} = \text{sgn}(V)$$

is capable of distinguishing between a molecule and its mirror image.



Vibrational frequencies

Normal modes of vibrational frequencies are useful for evaluating FFs (comparison with experiment) and studying conformational changes. The potential function may be expanded in a Taylor's series:

$$V = V_0 + \sum_{i=1}^{3N} \left(\frac{\partial V}{\partial x_i} \right)_0 \Delta x_i + \frac{1}{2} \sum_{i,j=1}^{3N} \left(\frac{\partial^2 V}{\partial x_i \partial x_j} \right)_0 \Delta x_i \Delta x_j + \frac{1}{6} \sum_{i,j,k=1}^{3N} \left(\frac{\partial^3 V}{\partial x_i \partial x_j \partial x_k} \right)_0 \Delta x_i \Delta x_j \Delta x_k + \dots$$

The first term (V_0) represents an arbitrary reference point and can be set to 0. At minimum all gradients are equal to 0, so the second term vanishes. The third term, representing second derivatives, is at the heart of the harmonic approximation of vibrations. Higher terms can be neglected in this approximation. Thus, we are left with:

$$V = \frac{1}{2} \sum_{i,j=1}^{3N} f_{ij} \Delta x_i \Delta x_j$$

$$f_{ij} = \left(\frac{\partial^2 V}{\partial x_i \partial x_j} \right)_0$$

where f_{ij} are force constants, arranged in a matrix with diagonal terms ($i=j$) and cross terms ($i \neq j$).

We try to solve the Newton's equations of motion:

$$m_i \frac{d^2 \Delta x_i}{dt^2} = - \sum_{j=1}^{3N} f_{ij} \Delta x_j \quad (i = 1, 2, \dots, 3N)$$

in mass-weighted coordinates $q_i = \sqrt{m_i} \cdot \Delta x_i$, which gives:

$$\frac{d^2 q_i}{dt^2} + \sum_{j=1}^{3N} f_{ij} q_j = 0$$

$$f_{ij} = \left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right)_0$$

We search for the solution in the form:

$$q_i = A_i \cos(\sqrt{\lambda} t + \delta)$$

Substitution yields the following eigenvalue problem:

$$\sum_{i=1}^{3N} \left[\left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right)_0 - \delta_{ij} \lambda \right] A_i = 0 \quad (j = 1, 2, \dots, 3N)$$

Vibrational frequencies are found from

$$\nu_i = \frac{\sqrt{\lambda}}{2\pi}$$

OF 3N coordinates used to calculate frequencies, 6 correspond to translations and rotations of a molecule as a whole (5 for linear systems). These modes have no restoring power, hence their frequencies are 0. In practice, deviations from 0 are a good indication of the quality of the calculation (it only works for well minimized structures; gradients < 0.001 kcal/mol/Å).

