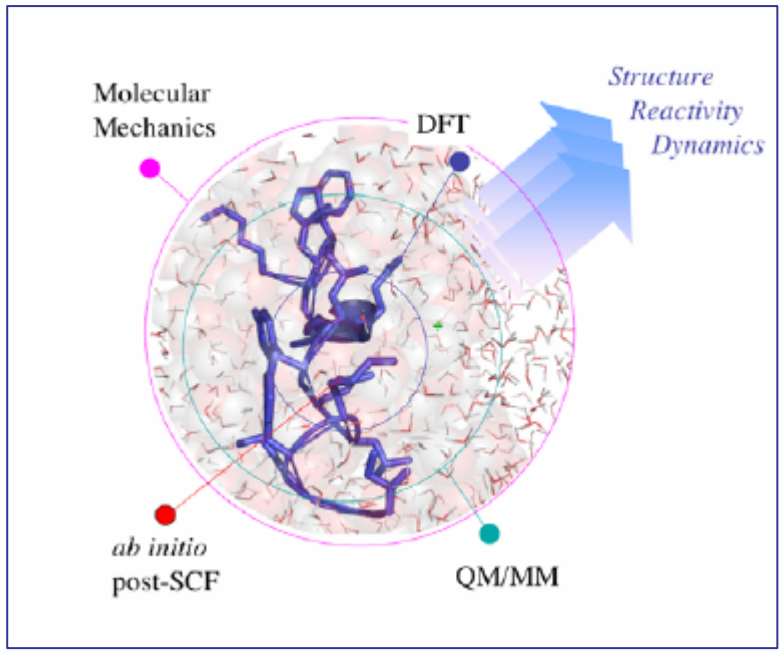
Societa' Chimica Italiana – Gruppo Interdivisionale di Chimica Computazionale

Scuola di Chimica Computazionale

Introduzione, per Esercizi, all'Uso del Calcolatore in Chimica Organica e Biologica



Modellistica Molecolare: un'introduzione alla meccanica molecolare e all'analisi conformazionale

Anna Bernardi – Universita' di Milano
Dip. Di Chimica Organica e Industriale

Siena, Aula Informatica del Complesso Didattico di San Miniato
25-29 Settembre 2006

Some of the material presented in this lecture was rielaborated from

http://bioinfo.tau.ac.il/~hanochs/cerius2/2004-2005/cerius2_winter_2004.html

a course on Molecular Modeling by Hanoch Senderowitz at Tel Aviv University

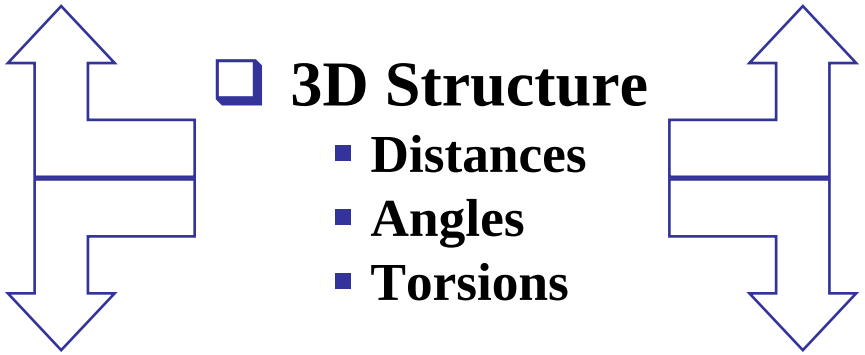
Molecular Modeling

- **We call "Molecular Modeling" the ensemble of methods used to for mimick the behavior of molecules or molecular systems**

- **Modeling is an efficient way to:**
  - **Study molecular properties**
  - **Rationalize and interpret experimental results**
  - **Make predictions for yet unstudied systems**
  - **Study hypothetical systems**
  - **Design new molecules**
  - **Understand and predict reactiviy**

Molecular Structure and Molecular Properties

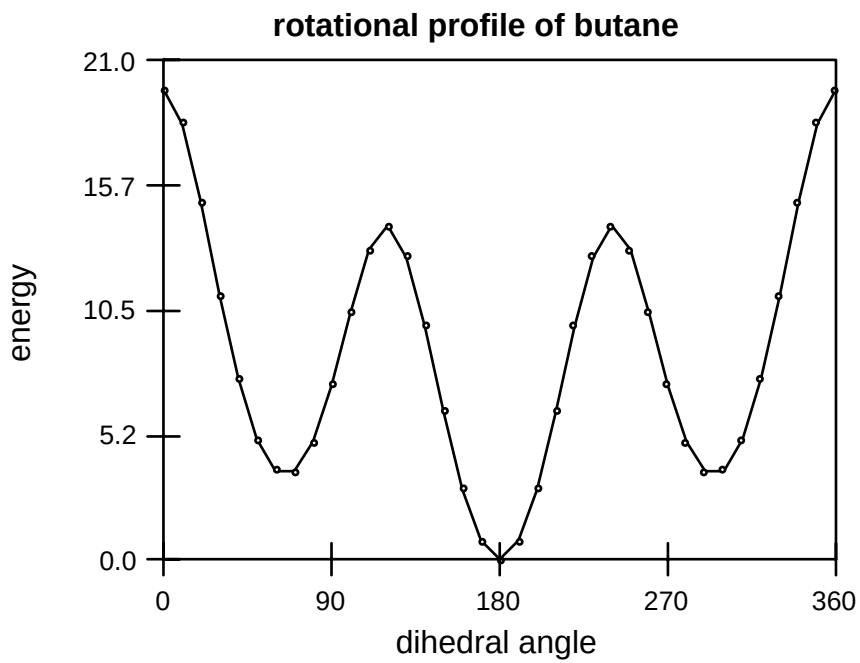
- **Energy**
  - Free energy (ΔG)
  - Enthalpy (ΔH)
  - Entropy (ΔS)
  - Steric energy (ΔG)
- **Molecular Spectroscopy**
  - NMR
  - IR
  - UV
  - MW



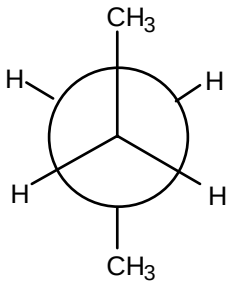
- **3D Structure**
  - Distances
  - Angles
  - Torsions
- **Kinetics**
  - Reaction Mechanisms
  - Rate constants
- **Electronic Properties**
  - Molecular orbitals
  - Charge Distributions
  - Dipole Moments

Molecular Structure

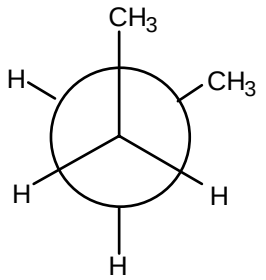
- The 3D structure of a molecule consisting of *N* atoms is defined by *3N* cartesian coordinates or *3N-6* internal coordinates.
- Each molecule may have a large number of 3D structures, only a few of which are "reasonable".
- The probability of a structure is a function of the structure's energy.

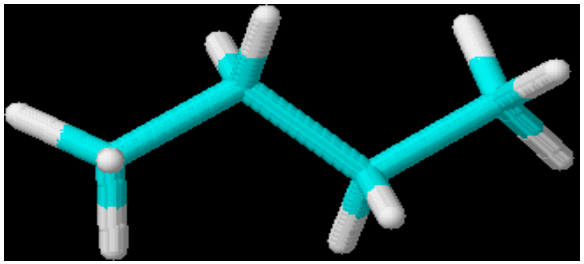


anti butane



gauche butane





From structure to properties

- Any molecular property is calculated by averaging the properties of all its 3D structures (conformations). The averaging weights are given by the relative energies of the conformations.

To describe a molecule we need to find all the low-energy conformations, and to determine their relative energy with some accuracy

How can we model molecular structures ?

The techniques of choice will vary both with **the questions we are trying to answer** and with **the size of the molecule involved :**

- **Macromolecules** (proteins): we are rarely interested in the whole conformational space, rather we want to get information on the "native" state of the macromolecule. This is often obtained by including experimental information (e.g. NOE contacts) in the initial steps of the model generation. It works because the macromolecules in question present essentially a single target conformation featuring a small subset of disordered loops

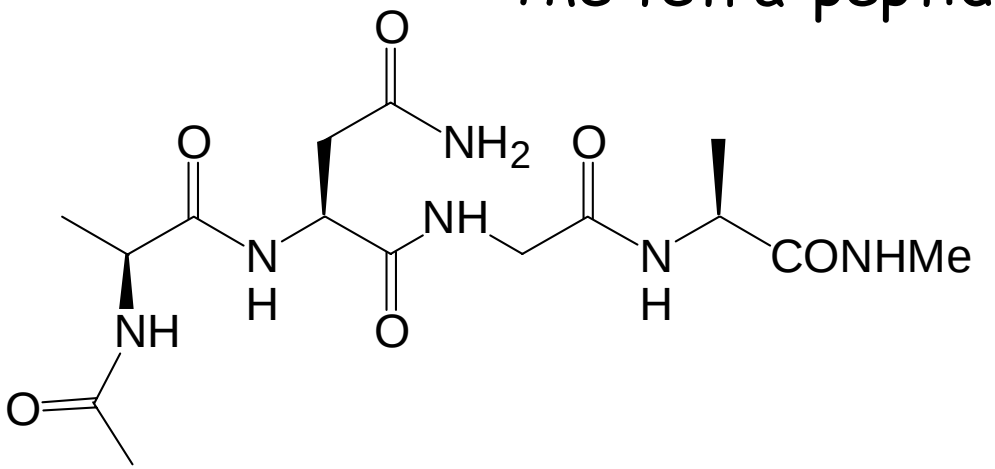
- **Flexible molecules** (small and medium-size peptides, oligosaccharides, most organic molecules): usually characterized by families of conformations which equilibrate rapidly on the NMR time scale and produce an averaged spectrum. Trying to fit simultaneously all the experimentally observed "constraints" will result in so-called *virtual conformations*.

The latter group includes the majority of naturally occurring biomolecular effectors.

Case study

To obtain a 3D model of a medium size, flexible molecule

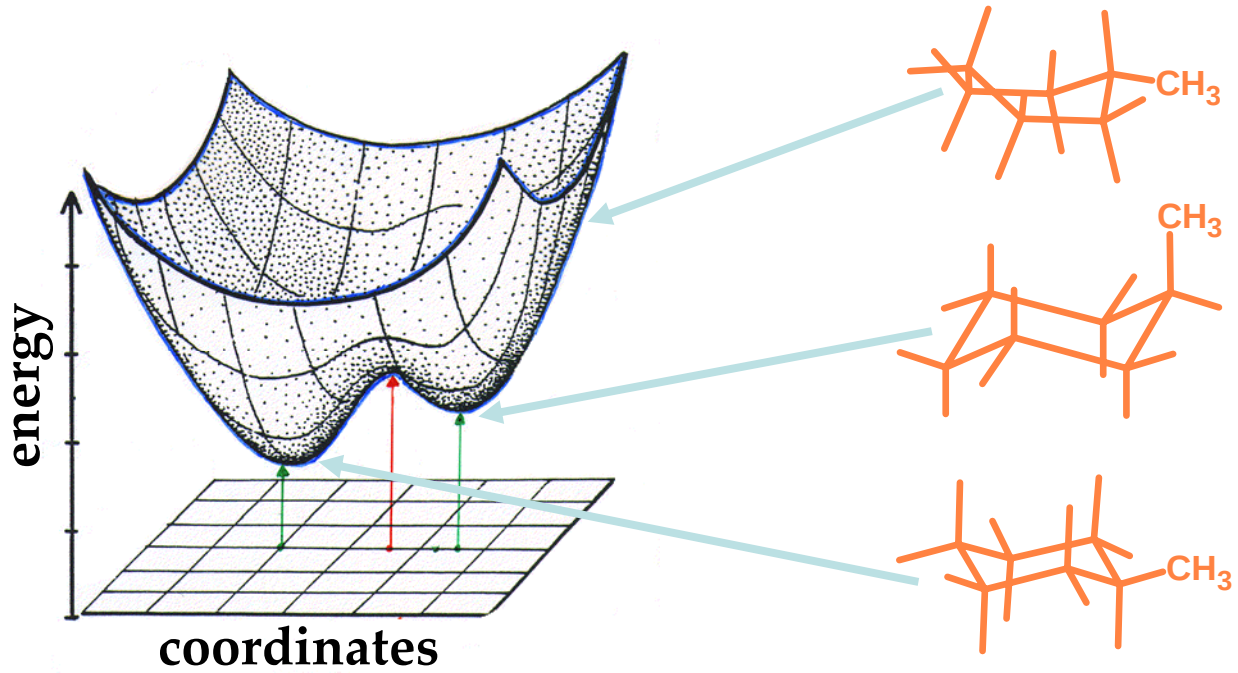
The tetra-peptide Ala-Asn-Gly-Ala (ANGA)



Or, an introduction to force field – based conformational analysis

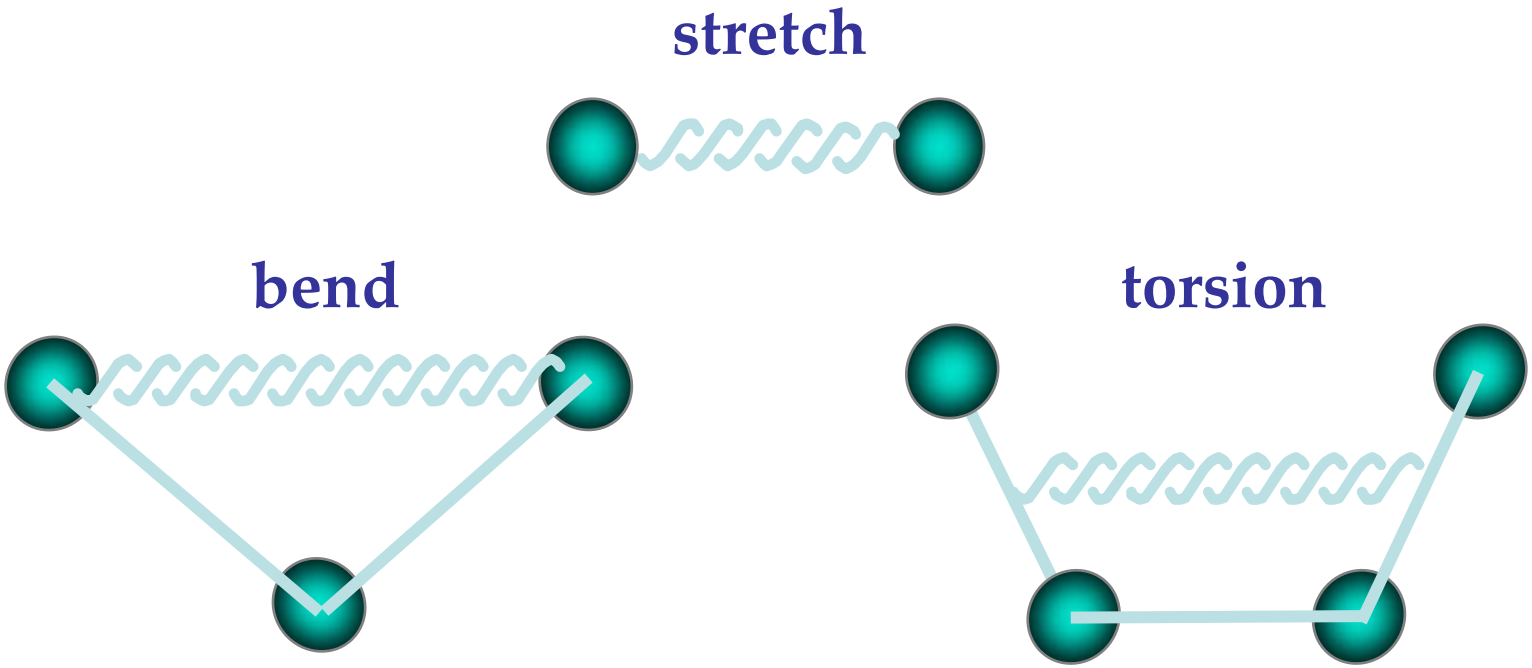
Force Field and Potential Energy Surface

- A force field is a method to describe a molecule as a collection of atoms held together by forces .
- A force field defines for each molecule a unique PES.
- Each point on the PES represents a molecular conformation characterized by its structure and energy.
- Each conformation is characterized by an energy value. This value is used to optimize the geometry of the 3D structure.
- The optimized structure is used to calculate various molecular properties.



A Molecule is a Collection of Atoms Held Together by Forces

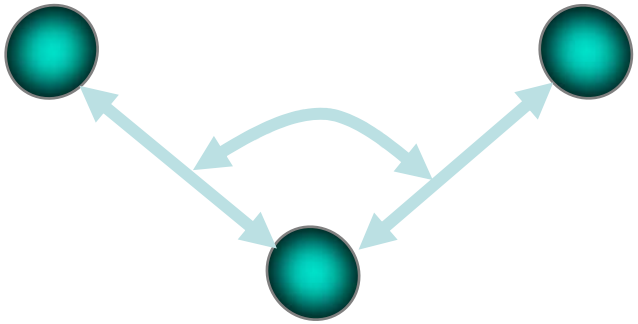
- The simplest forces act between bonded atoms
- They work to return structural parameters to the equilibrium values



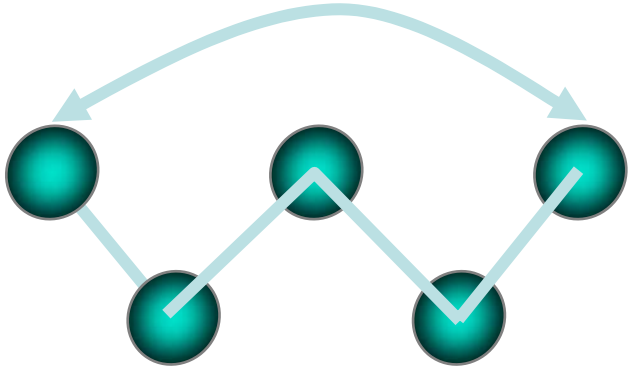
Other Forces…

- Forces also act between non-bonded atoms
- Cross terms couple the different types of interactions

Stretch-bend



Non-bonded

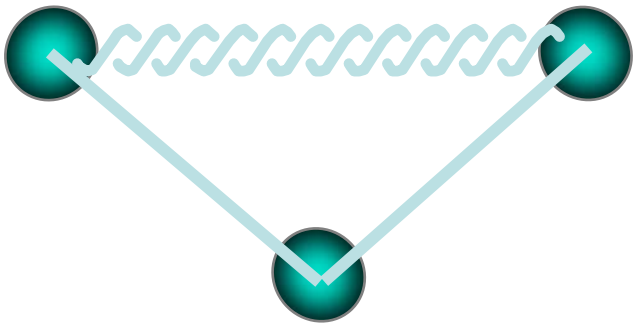


The Forces are Described by Potential Energy Functions

stretch

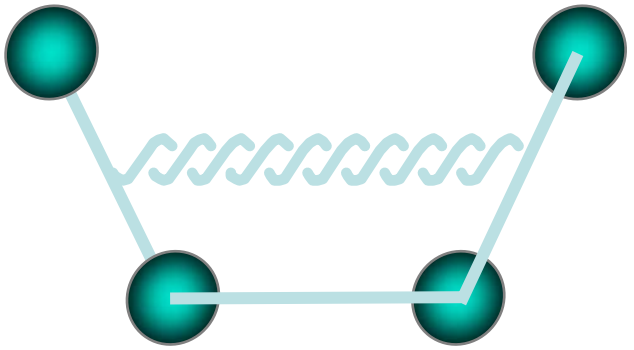
$$v(l) = \frac{k}{2}(l - l_0)^2$$

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



bend

$$v(\omega) = \sum_{n=0}^{N} \frac{V_n}{2}(1 + \cos(n\omega - \gamma)]$$



torsion

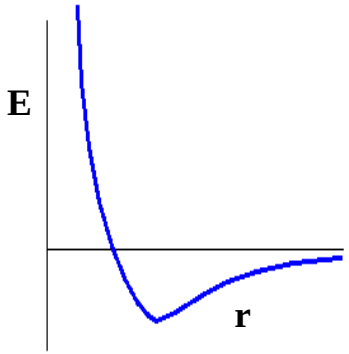
Bond Stretching

- **Morse potential**

$$v(l) = De\{1 - \exp[-a(l - l_0)]\}^2$$

$$a = \omega\sqrt{\mu / 2De}$$

$$\omega = \sqrt{k/\mu}$$

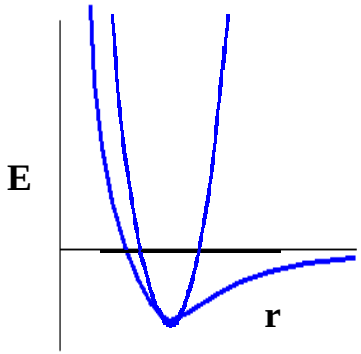


D_e = Depth of the potential energy minimum
l_0 = Equilibrium bond length (v(l)=0)
ω = Frequency of the bond vibration
μ = Reduced mass
k = Stretch constant

- Accurate.
- Computationally inefficient (form, parameters).
- Catastrophic bond elongation.

- **Harmonic potential (AMBER)**

$$v(l) = \frac{k}{2}(l - l_0)^2$$



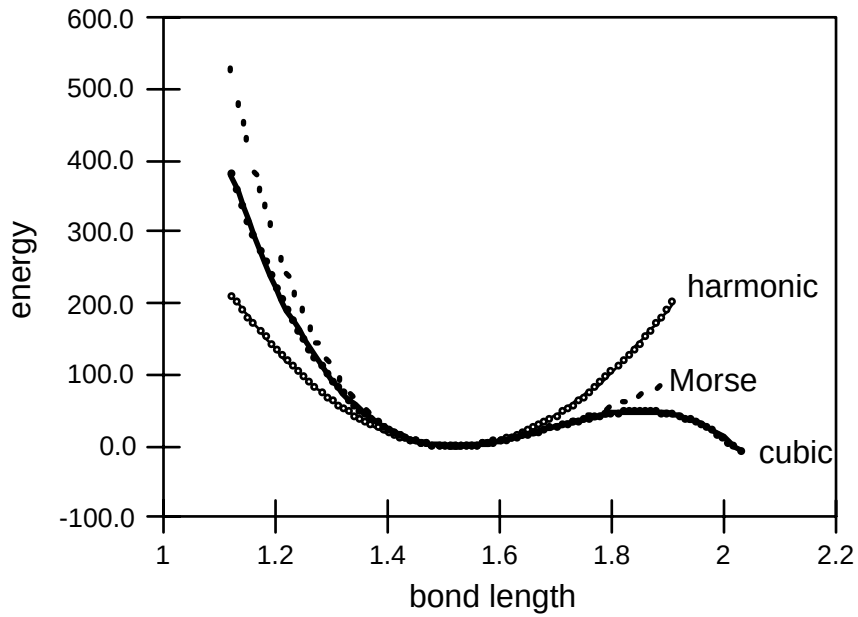
- Innacurate
- Coincides with the Morse potential at the bottom of the well.
- Computationally efficient.

Bond Stretching

- Cubic (MM2) and quartic (MM3) potentials

$$v(l) = \frac{k1}{2}(l - l_0)^2 - \frac{k2}{2}(l - l_0)^3$$

$$v(l) = \frac{k1}{2}(l - l_0)^2 - \frac{k2}{2}(l - l_0)^3 + \frac{k3}{2}(l - l_0)^4$$



Force Field: Parameters

- Types of parameters
  - Stretch: natural bond length (l_0) and force constants (k).
  - Bend: natural bond angles (θ_0) and force constants (k).
  - Torsions: V_i's.
  - VdW: (e, VdW radii).
  - Electrostatic: Partial atomic charges.
  - Cross-terms: Cross term parameters.

- **The values for the parameters in the potential function equations are chosen to best reproduce experimental data.**

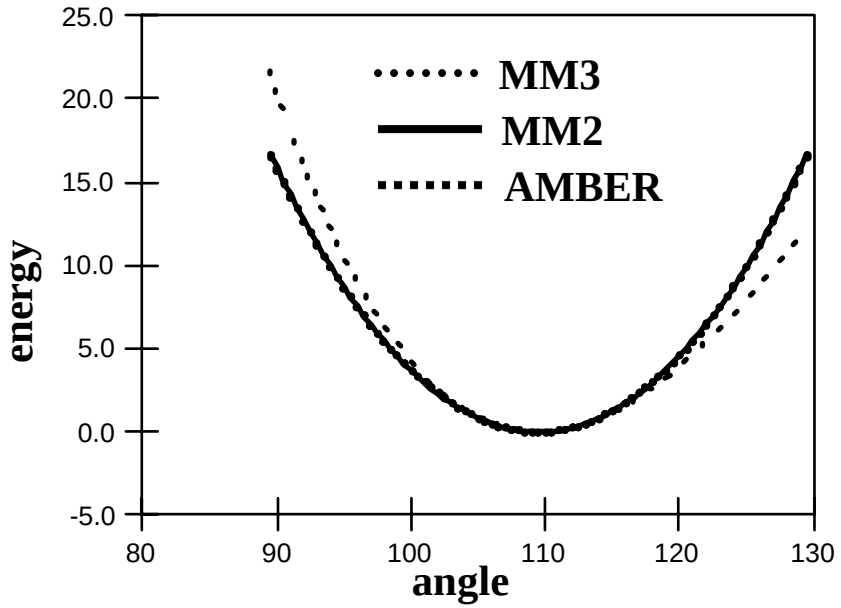
Bond Stretching Parameters (MM2)

Bond	l_0 (A)	k (kcal mol^{-1} A^{-2})
Csp^3-Csp^3	1,523	317
Csp^3-Csp^2	1,497	317
Csp^2=Csp^2	1,337	690
Csp^2=O	1,208	777
Csp^3-Nsp^3	1,438	367
C-N (amide)	1,345	719

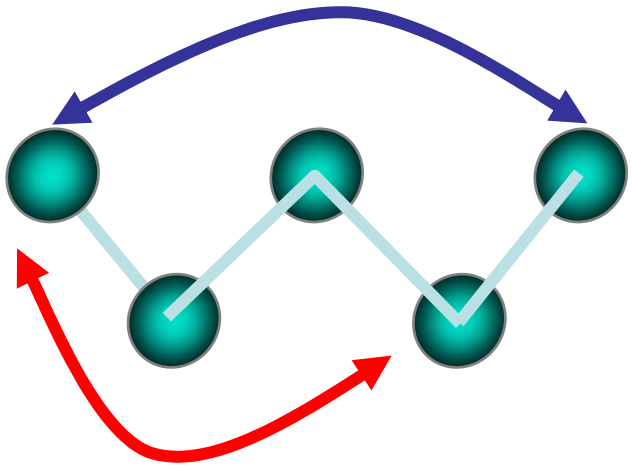
- Bond types correlate with l_0 and k values.
- A 0.2Å deviation from l_0 when k=300 leads to an energy increase of 12 kcal/mol.

Angle Bending

- **AMBER:** $v(\theta) = \frac{k}{2}(\theta - \theta_0)^2$
- **MM2:** $v(\theta) = \frac{k1}{2}(\theta - \theta_0)^2 + \frac{k2}{2}(\theta - \theta_0)^6$
- **MM3:** $v(\theta) = \frac{k1}{2}(\theta - \theta_0)^2 + \frac{k2}{2}(\theta - \theta_0)^3 + \frac{k3}{2}(\theta - \theta_0)^4 + \frac{k4}{2}(\theta - \theta_0)^5 + \frac{k5}{2}(\theta - \theta_0)^6$



Non-bonded terms



$$v = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right)$$

Calculated for each pair of atoms separated by at least 3 bonds

Non-Bonded Interactions

- Operate within molecules and between molecules.
- Through space interactions.
- Modeled as a function of an inverse power of the distance, up to a cut-off point.
- Divided into:
  - Electrostatic interactions
  - VdW interactions

Electrostatic Interactions

$$V_{intermoleclar} = \sum_{i=1}^{A}\sum_{j=1}^{B}\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad V_{intramoleclar} = \sum_{i=1}^{A}\sum_{j=i+1}^{A}\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

- q_i, q_j are *point charges*.
- Charge-charge interactions are effective also at long range (decay as r^{-1}).
- When q_i, q_j are centered on the nuclei they are called *partial atomic charges*. They can be determined from:
  - Fitting to known electric moments (*e.g.,* dipole, quadrupole etc.)
  - Fitting to thermodynamic properties.
  - *Ab initio* Calculations
    - by fitting to electrostatic potential (ESP)

      (at a given point, the ESP is the force acting on a positive charge placed at that point. It's an observable quantity, that can be determined from a wavefunction)

Van Der Waals Interactions (VdW)

- Electrostatic interactions can't account for all non-bonded interactions within a system (*e.g.,* rare gases).

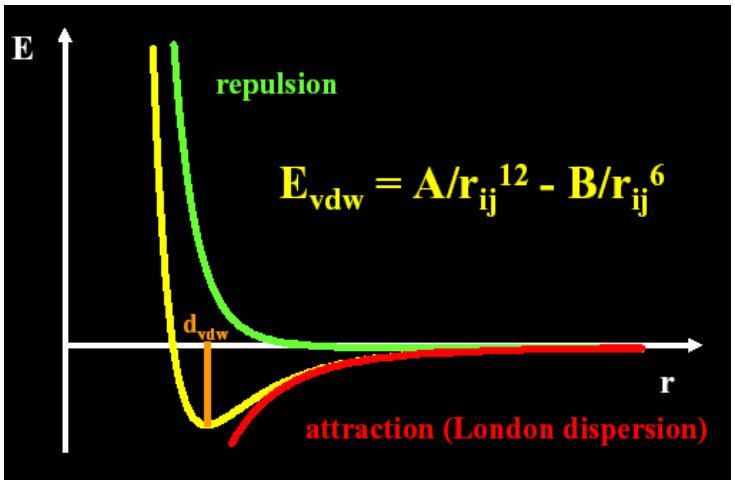
- VdW interactions:

  - Attractive (dispersive) contribution (London forces)
    - Instantaneous dipoles due to electron cloud fluctuations.
    - Decays as r^6

  - Repulsive contribution
    - Nuclei repulsion.
    - At short distances ($r < 1$) rises as $1/r$.
    - At large distances decays as $exp(-2r/a_0)$; a_0 the Bohr radius.

Van Der Waals (VdW) Interactions

- The observed VdW potential results from a balance between attractive and repulsive forces.

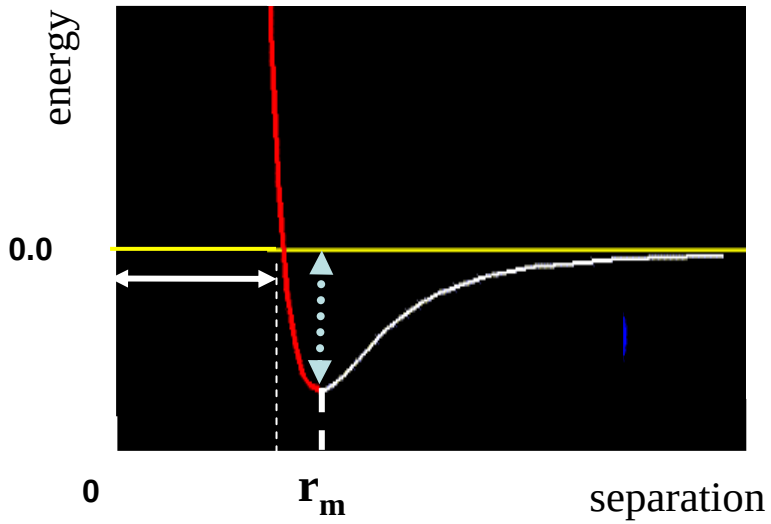


- It is often modeled with a Lennard-Jones Potential

$$v(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

σ: separation for which E=0
ε: well depth

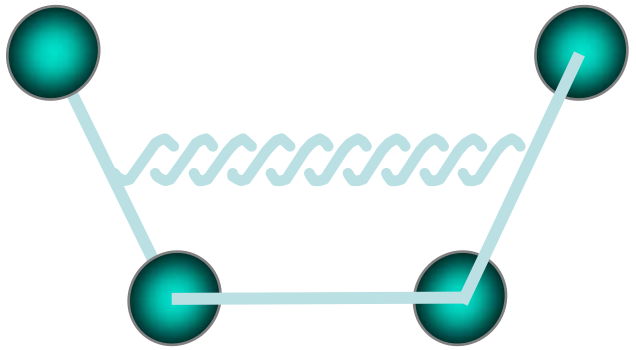
- Rapid to calculate
- Attractive part theoretically sound
- Repulsive part easy to calculate but too steep



Torsional (Dihedral) Terms

- Reflect the existence of barriers to rotation around chemical bonds.
- Used to set the relative energies of the rotational minima and maxima. (Can also be used to set the rotational barriers)
- Together with the non-bonded terms are responsible for most of the structural and energetic changes
- Usually **parameterized last.**

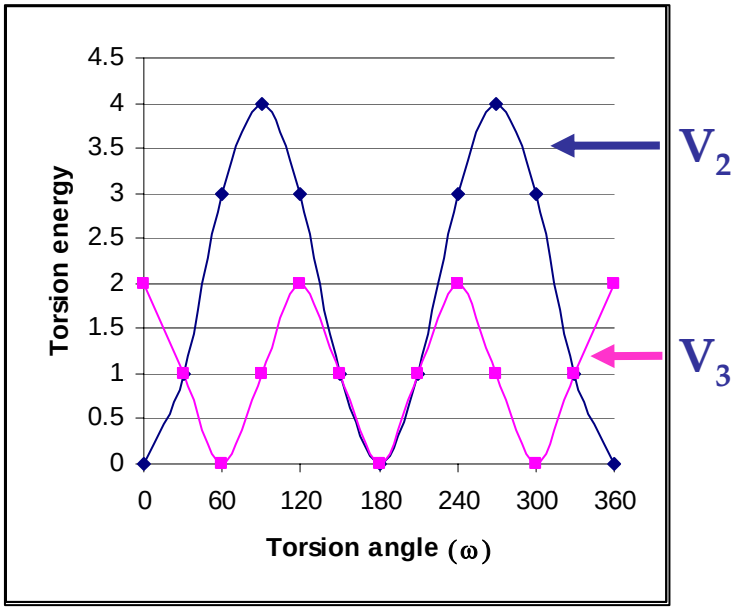
$$v(\omega) = \sum_{n=0}^{N} \frac{V_n}{2}(1 + \cos(n\omega - \gamma)]$$



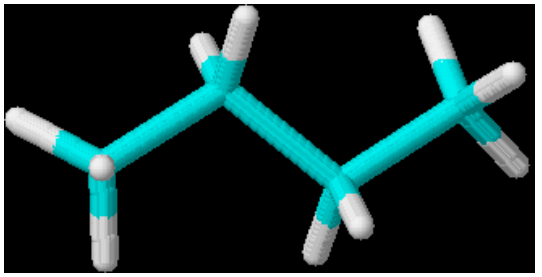
Torsional Term: Functional Form

Amber $v(\omega) = \sum_{n=0}^{N} \frac{V_n}{2}(1+\cos(n\omega-\gamma)]$

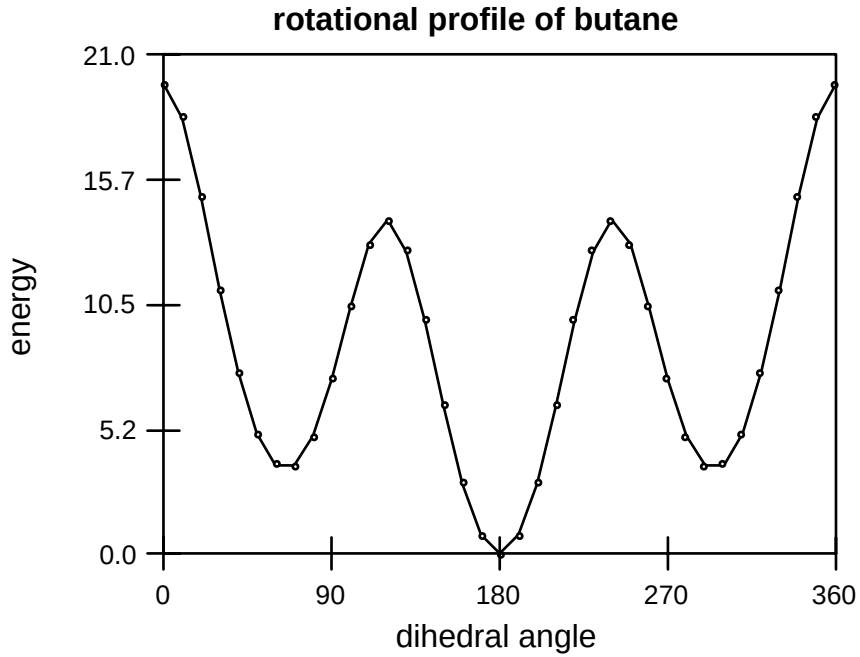
MM2 $E_t = V_1/2\ (1 + \cos\omega) + V_2/2\ (1 - \cos2\omega) + V_3/2\ (1 + \cos3\omega)$



Example: Butane

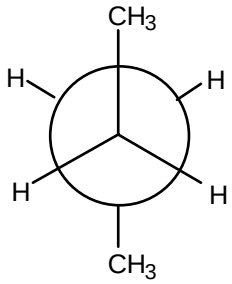


- The barrier to rotation around the C-C bond in butane is ~20kJ/mol.
- All 9 torsional interactions around the central C-C bond should be considered for an appropriate reproduction of the torsional barrier.

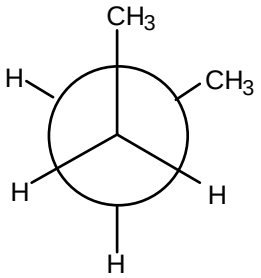


#	Type	V1	V2	V3
1	C-C-C–C	0.200	0.270	0.093
4	C-C-C–H	0.000	0.000	0.267
4	H-C-C–H	0.000	0.000	0.237

anti butane



gauche butane



A Force Field is a Collection of Potential Energy Functions

- **It is defined by the functional forms of the energy functions and by the values of their parameters.**

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

Steric Energy

Steric Energy = $E_{stretch} + E_{bend} + E_{torsion} + E_{VdW} + E_{electrostatic} + E_{stretch\text{-}bend} + E_{torsion\text{-}stretch} + \ldots$

- $E_{stretch}$ Stretch energy (over all bonds)
- E_{bend} Bending energy (over all angles)
- $E_{torsion}$ Torsional (dihedral) energy (over all dihedral angles)
- E_{VdW} Van Der Waals energy (over all atom pairs > 1,3)
- $E_{electrostatic}$ Electrostatic energy (over all charged atom pairs >1,3)
- $E_{stretch\text{-}bend}$ Stretch-bend energy
- $E_{torsion\text{-}stretch}$ Torsion-stretch energy

$E_{VdW} + E_{electrostatic}$ are often called non-bonded energies

Steric Energy is Calculated from the Force Field

- **A molecular mechanics program will return an energy value for every conformation of the system.**
- **Steric energy is the energy of the system relative to a reference point, which depends on the bonded interactions and is both *force field dependent* and *molecule dependent*.**
- **Steric energy can only be used to *compare the relative stabilities of different conformations of the same molecule* and can not be used to compare the relative stabilities of different molecules.**
- **All conformational energies must be calculated with the same force field.**

Force Fields: General Features

- **Force field definition**
  - Functional form (usually a compromise between accuracy and ease of calculation.
  - Parameters (transferability assumed).
- **Force fields are empirical**
  - There is no "correct" form of a force field.
  - Force fields are evaluated based solely on their performance.
- **Force field are parameterized for specific properties**
  - Structural properties
  - Energy
  - Spectra

Parameters

- **Types of parameters**
  - **Stretch: natural bond length (l_0) and force constants (k).**
  - **Bend: natural bond angles (q_0) and force constants (k).**
  - **Torsions: V_i's.**
  - **VdW: (e, VdW radii).**
  - **Electrostatic: Partial atomic charges.**
  - **Cross-terms: Cross term parameters.**
- **Parametrization techniques**
  - **Trial and error**
  - **Least square methods**
- **The values of the parameters are chosen to reproduce experimental data.**

Parameters : Source

- A force field parameterized according to data from one source (*e.g.*, experimental gas phase, experimental solid phase, *ab initio*) will fit data from other sources only qualitatively.
- Experiment: geometries and non-bonded parameters
  - X-Ray crystallography
  - Electron diffraction
  - Microwave spectroscopy
  - Lattice energies
- Advantages
  - Real
- Disadvantages
  - Hard to obtain
  - Non-uniform
  - Limited availability

Parameters - Source

- **High level molecular orbital calculations**
  - Conformational preferences, rotational barriers and partial charges (HF/6-31G*).
  - Corrections for electron correlation effects are often important: MP2, MP3 etc.
  - Charges obtained by fitting electrostatic potentials.
- **Advantages**
  - Relatively easy to obtain.
  - Uniform.
  - Unlimited availability.
  - Complete potential energy surfaces are available.
- **Disadvantages**
  - “Unreal”.
  - Model dependent.
  - Computationally expensive.

Parameters - Quality

- **"Unusual" functional groups or combinations of functionality will probably not be well described by existing force fields.**
- **The success of modeling with molecular mechanics depends on the use of parameters which are "high quality" *e.g.* not generalized.**
- **Each standard force field has been designed with a particular (or few) molecular property and target group of compounds in mind.**

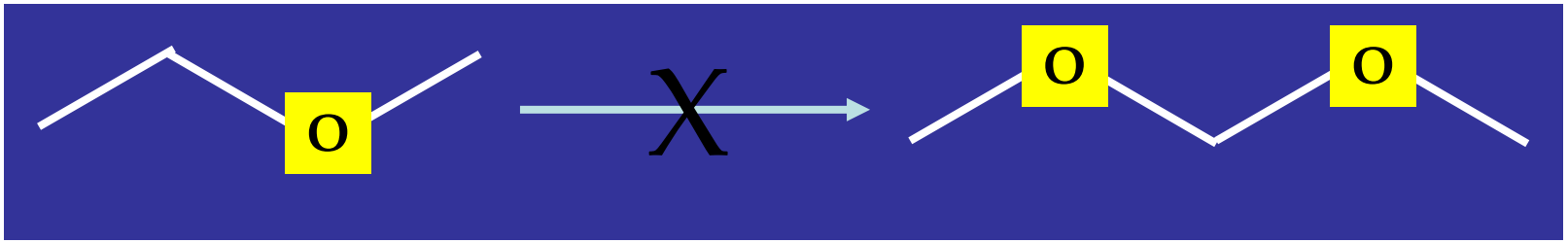
Know your parameters!!!!!

Transferability of Parameters

- **FF work under the assumption that parameters are transferable, i.e. that the same set of parameters can be used to model a related series of compounds**

  Missing parameters can be added by
  - Educated guess
  - Development (and optimization) of new parameters

- **This assumption breaks down for close interacting functional groups:**



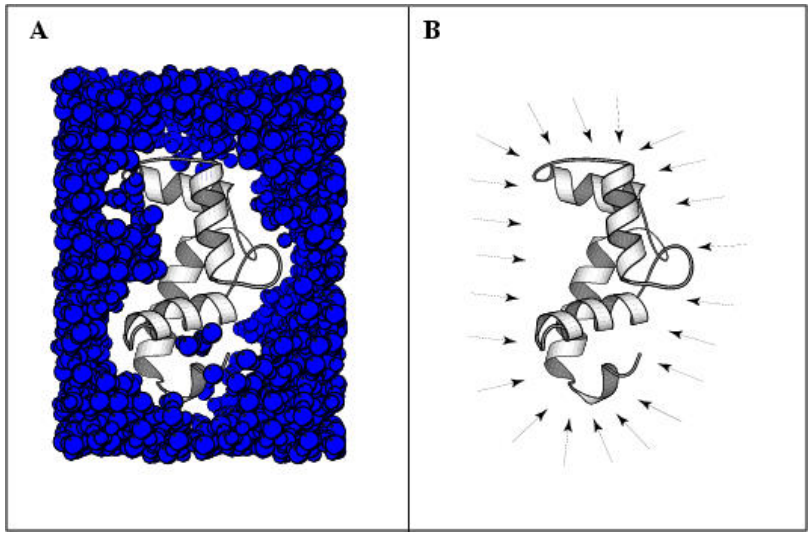
Solvent Dielectric Models

$$V = \sum_{i=1}^{A} \sum_{j=i+1}^{A} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

- ε_0 = Dielectric constant of vacuum (1).
- For a given set of charges and distance, ε_0 determines the strength of the electrostatic interactions.
- Solvent effect dampen the electrostatic interactions and so can be modeled by varying ε_0:
  - ✓ $\varepsilon_{eff} = \varepsilon_0 \varepsilon_r$
  - ✓ ε_r(protein interior) = 2-4
  - ✓ ε_r(water) = 80

Solvent Models

- **Explicit models** use hundreds or thousands of discrete solvent molecules



- **Implicit models** treat the solvent as a continuous medium surrounding the solute beginning at the van der Waals surface.
  Many models have been described. Among these, the generalized Born (GB), Surface Area (SA) model (an empirical model) has become very popular

The GB/SA solvation model

$$\Delta G_{sol} = \Delta G_{elec} + \Delta G_{VDW} + \Delta G_{cav}$$

- **The solvent-solvent cavity term (*Gcav*) and the solute-solvent van der Waals term (*GvdW*) are computed together as a linear function of the solvent-accessible surface areas**

$$\Delta G_{VDW} + \Delta G_{cav} = \sum s_i A_i$$

- **The electrostatic term ΔG_{elec} is represented by the generalised Born equation**

$$G_{pol} = -166.0\left(1 - \frac{1}{\epsilon}\right)\sum_{i=1}^{N}\sum_{j=1}^{N}\frac{q_i q_j}{\sqrt{r_{ij}^2 + b_i b_j \exp\left(-\frac{r_{ij}^2}{4 b_i b_j}\right)}}$$

- q_i and q_j partial charges of atoms i and j
- r_{ij} interatomic distance
- ε dielectric constant of the medium
- b_i and b_j Born radius of atoms i and j,

Implicit Solvent Models

Advantages:

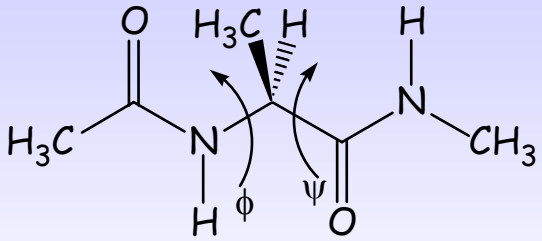
- Cost effective
- No convergence problems for minimizations and conformational search
- Good reproduction of some experimentally known data or systems

Disadvantages:

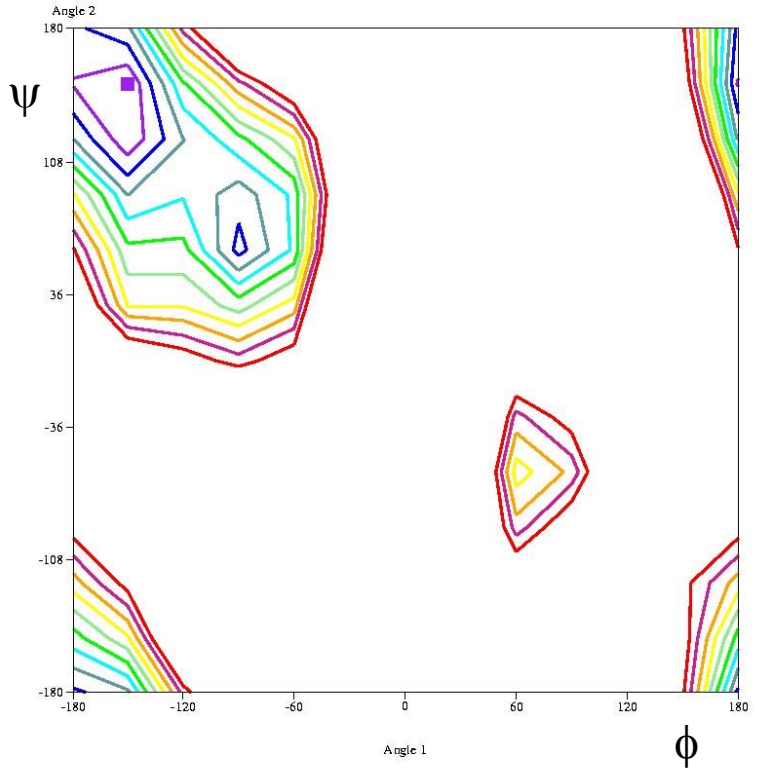
- Problems dealing with charged groups and molecules
- Blind to areas seen as non accessible (molecular complexes)
- Misses effects which are due to the presence of individual water molecules in the vicinity of the solute (CAVEAT !)

Ac-Ala-NHMe

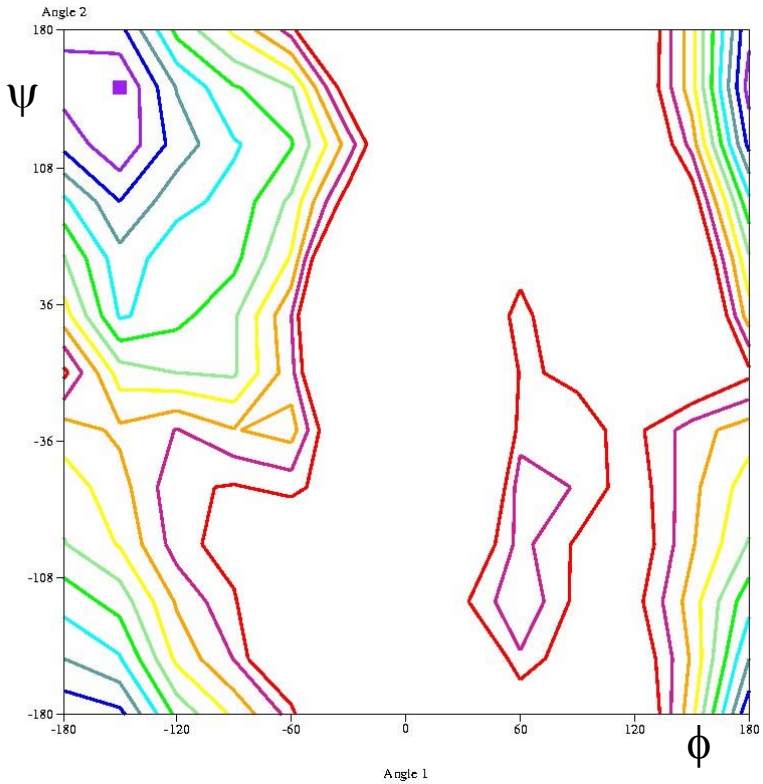
ϕ (C-N-Cα-C)
Ψ (N-Cα-C-N)



Amber*



Amber*, GB/SA water solvation



Existing Force Fields

- **AMBER** (http://amber.scripps.edu)
  - Parameterized specifically for proteins and nucleic acids
  - United Atoms or All Atoms parameters (including charges).
  - Uses only 5 bonding and non-bonding terms
  - No cross terms are included.
  - Results can be very good for proteins and nucleic acids, less so for other systems.
- **CHARMM** (http://www.charmm.org)
  - Originally devised for proteins and nucleic acids.
  - Now used for a range of macromolecules, molecular dynamics, solvation, crystal packing, vibrational analysis and QM/MM studies.
  - Uses 5 valence terms, one of which is electrostatic term.
  - Basis for other force fields (*e.g.,* MOIL).

Existing Force Fields

- **GROMOS (http://www.igc.ethz.ch/gromos)**
  - **Popular for predicting the dynamical motion of molecules and bulk liquids.**
  - **Also used for modeling biomolecules.**
  - **Uses 5 valence terms, one of which is an electrostatic term.**
- **MM1, 2, 3, 4 (http://europa.chem.uga.edu)**
  - **General purpose force fields for (mono-functional) organic molecules.**
  - **MM2 was parameterized for a lot of functional groups.**
  - **MM3 is probably one of the most accurate ways of modeling hydrocarbons.**
  - **MM4 is very new and little is known about its performance.**
  - **The force fields use 5 to 6 valence terms, one of which is an electrostatic term and one to nine cross terms.**

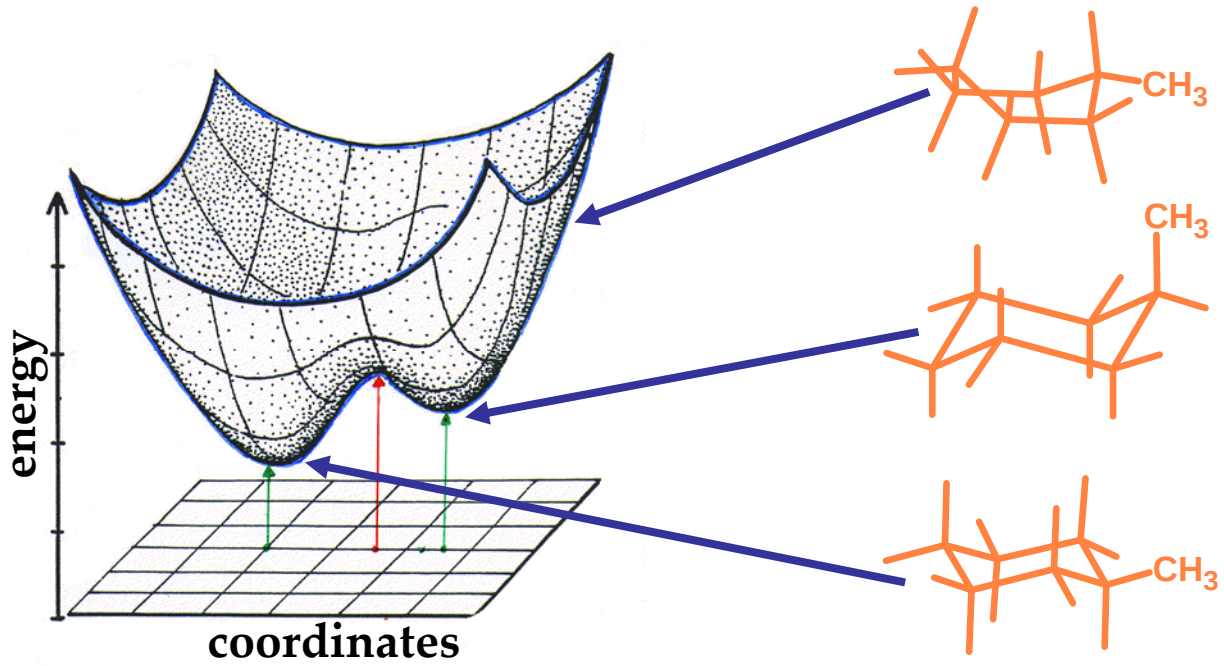
Existing Force Fields

- **MMFF (Merck Molecular Force Field)**
  (http://www.psc.edu/general/software/packages/charmm/tutorial/mackerell/MMFF_00.pdf
  - **General purpose force fields mainly for organic molecules.**
  - **MMFF94 was originally designed for molecular dynamics simulations but is also widely used for geometry optimization.**
  - **Uses 5 valence terms, one of which is an electrostatic term and one cross term.**
  - **MMFF was parameterized based on high level *ab initio* calculations.**

- **OPLS (Optimized Potential for Liquid Simulations)**
  - **Designed for modeling bulk liquids.**
  - **Has been extensively used for modeling the molecular dynamics of biomolecules.**
  - **Uses 5 valence terms, one of which is an electrostatic term but no cross terms.**

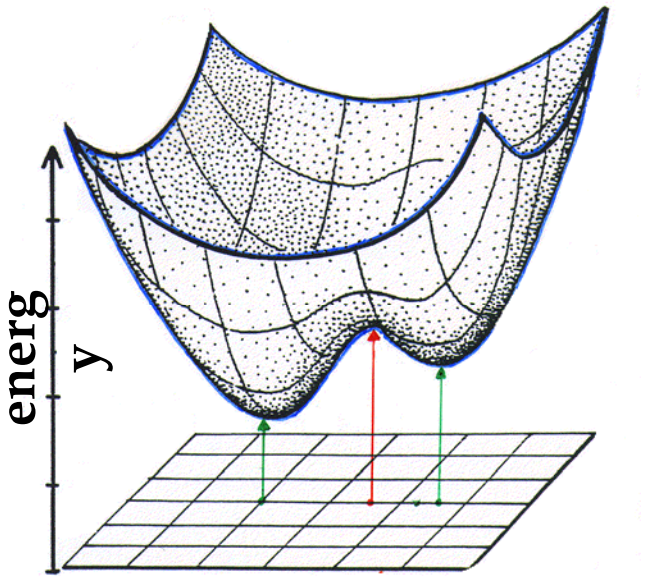
Force Field and Potential Energy Surface

- A force field defines for each molecule a unique PES.
- Each point on the PES represents a molecular conformation characterized by its structure and energy.
- Energy is a function of the coordinates.
- Coordinates are function of the energy.



Sampling the PES

- **Energy minimization**
  - Single minimum
- **Conformational search**
  - Multiple minima

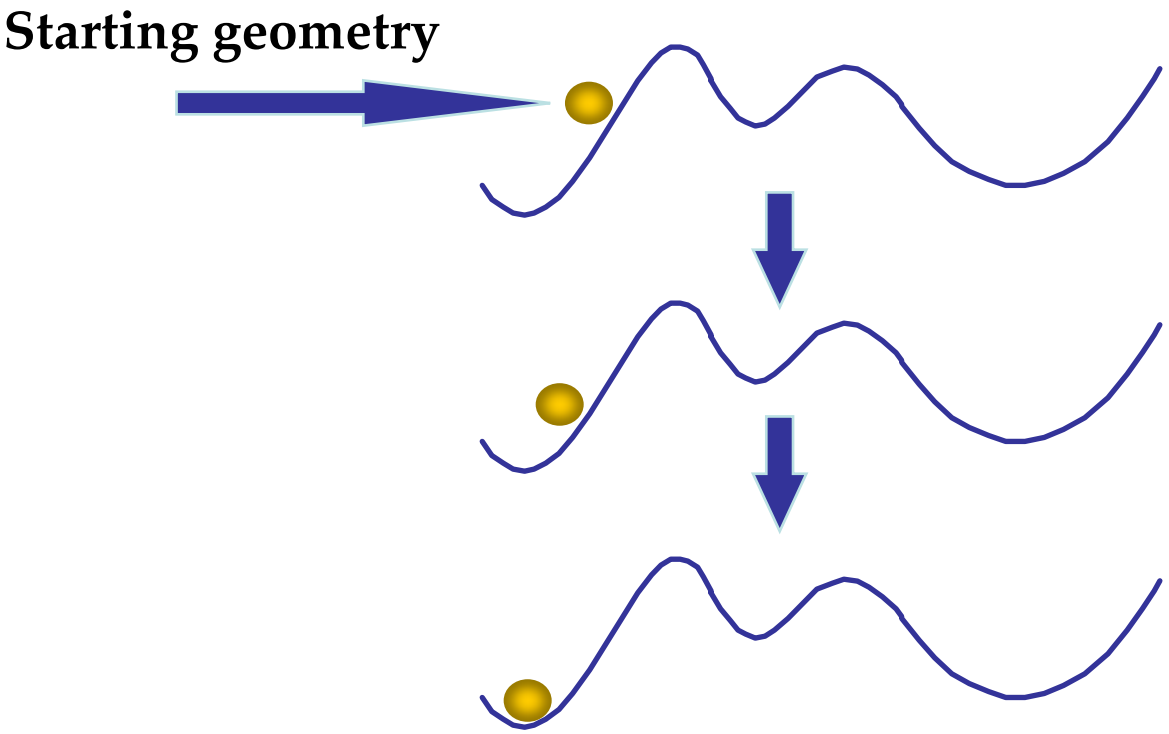


Minimization Definitions

- Given a function:
- Find values for the variables for which f is a minimum:

- **Functions**
  - Quantum mechanics energy
  - Molecular mechanics energy
- **Variables**
  - Cartesian
  - Internal
- **Minimization algorithms**
  - Derivatives-based
  - Non derivatives-based

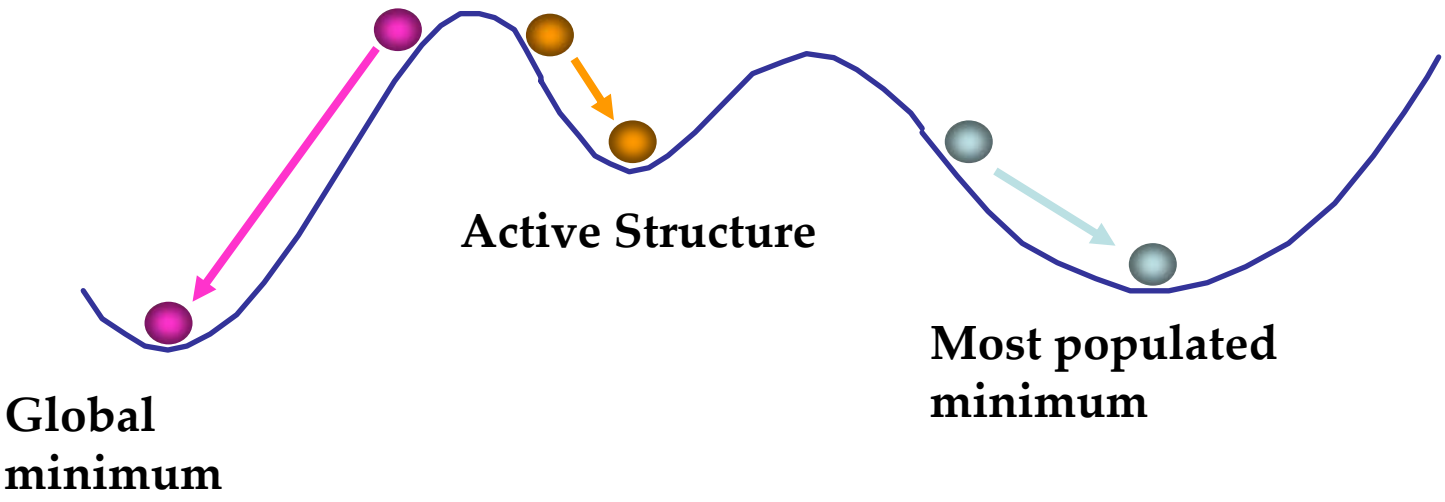
A Schematic Representation



⇑ **Easy to implement; useful for well defined structures**

⇓ **Depends strongly on starting geometry**

Population of Minima

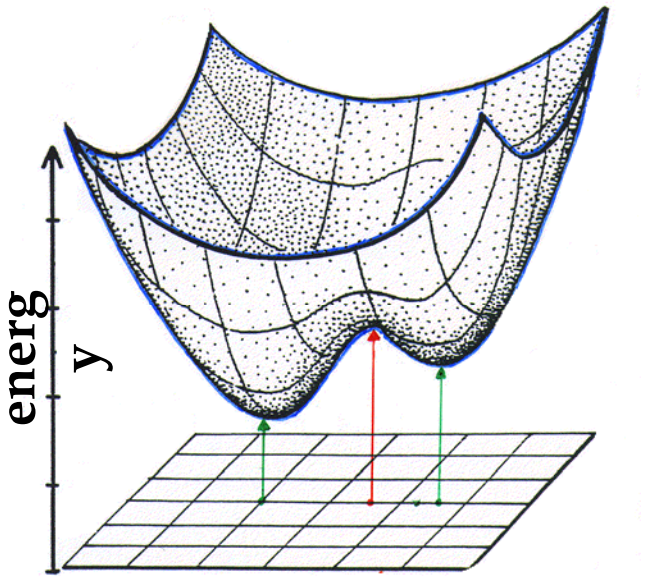


- Most minimization method can only go downhill and so locate the closest (downhill sense) minimum.
- No minimization method can guarantee the location of the global energy minimum.
- No method has proven the best for all problems.

Sampling the PES

- **Energy minimization**
  - Single minimum

- **Conformational search**
  - Multiple minima



From structure to properties

To describe a molecule we need to find all the low-energy conformations, and to determine their relative energy with some accuracy

- **The probability of a conformation is a function of the conformation's (relative) energy**
- **Any molecular property is calculated by averaging the properties of all its conformations. The averaging weights are obtained from the relative energies of the conformations.**

Boltzmann Averaged Properties

- **Fraction of conformation *i* in an equilibrium mixture:**

$$x_i = \frac{\exp\left(\frac{-E_i}{RT}\right)}{\sum_{j=1}^{n} \exp\left(\frac{-E_i}{RT}\right)} \qquad R = 0.00198\, kcal/mol - {}^{\circ}K$$

Partition function

- **Equilibrium molecular properties are obtained by Boltzmann averaging the properties of the individual conformations:**

$$p = \sum_{i=1}^{n} x_i p_i$$

Boltzmann Averaging: Example

Conformer	Relative energy (kJ/mol)	%	^{3}J (Hertz)
1	0	41	1,6
2	1	28	8,7
3	2	19	3,8
4	3	12	1,9

observed $^{3}J = 1.6 \times 0.41 + 8.7 \times 0.28 + 3.8 \times 0.19 + 1.9 \times 0.12 = 4.0$ Hz