Scientific Background on the Nobel Prize in Chemistry 2013

DEVELOPMENT OF MULTISCALE MODELS FOR COMPLEX CHEMICAL SYSTEMS

KUNGL. VETENSKAPS-AKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

Martin Karplus, Harvard U., Cambridge, MA, USA
Michael Levitt, Stanford U., Stanford, CA, USA
and
Arieh Warshel, U. Southern Ca., Los Angeles, CA, USA

… there is not enough experimental information to uniquely determine the structure of the studied system. This is just one of the aspects of how computers and theoretical models have become essential tools for the experimental chemist.

Today the focus of chemical research is much more on function than on structure. Chemists asks questions like “How does this happen?” rather than “What does this look like?” . Question about function are generally difficult to answer using experimental techniques.

… This makes theoretical modelling an important tool as a complement to the experimental techniques.

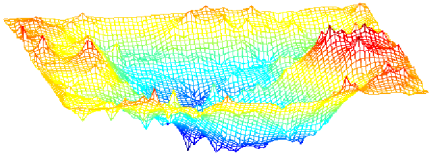
Chemical processes are characterized by a transition state, a configuration with the lowest possible (free) energy that links the product(s) with the reactant(s). This state is normally not experimentally accessible, but there are theoretical methods to search for such structures. Consequently theory is a necessary complement to experiment.

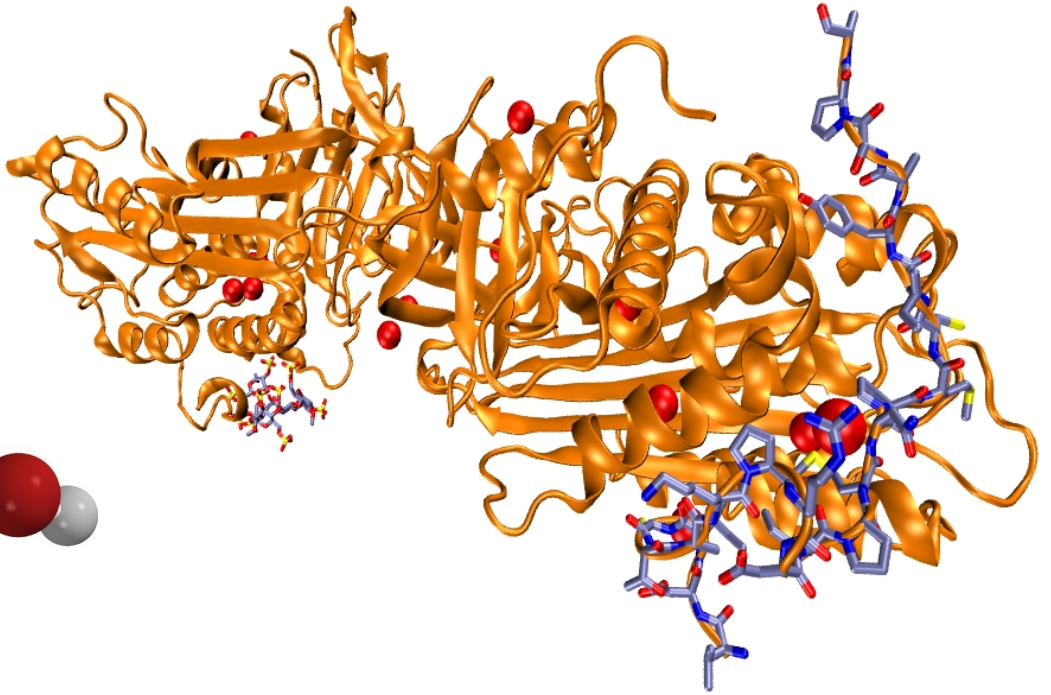
Wintersemester 2016/2017
Biomolecular Engineering/Nanobiophysics Modules

BIOMOLECULAR MODELLING:

METHODOLOGY AND CASE STUDIES IN

COMPUTATIONAL BIOLOGY





$$\Phi = \sum B_{\perp} \Delta A \quad E = mc^2 \quad E = -13.6 \frac{(Z^*)^2}{n^2} (eV)$$

$$E = -N \frac{\Delta \Phi}{\Delta t} \quad \tau = \tau_0 \Big/ \sqrt{1 - v^2/c^2} \quad r = .53 \frac{n^2}{Z} (\text{Å}) \quad {}^1_1p \quad {}^1_0n \quad {}^4_2He \quad e^- \; e^+$$

$$E = -L \frac{\Delta I}{\Delta t} \quad L = L_0 \sqrt{1 - v^2/c^2} \quad p = h/\lambda \quad 938\,\text{MeV} \quad .511\,\text{MeV}$$

$$V = I X \quad E = h f \quad E = \frac{p^2}{2m} + V(x) \quad N = N_0 \left(\frac{1}{2}\right)^{t/t_{1/2}}$$

$$X_C = \frac{1}{\omega C} \quad h f = \varphi_0 + \frac{1}{2} m v^2 \quad \psi(x) \quad |\psi(x)|^2$$

$$X_L = \omega L \quad h = 6.63\,(10)^{-34}\,\text{Js} \quad L = n \frac{\lambda}{2}$$

$$\omega_0 = 1/\sqrt{LC} \quad \Delta p \, \Delta x \geq h/2\pi$$

$$\Delta E \, \Delta t \geq h/2\pi$$

$$c = \lambda f \quad \Delta f/f = \pm v/c \quad \frac{d}{2} \sin(\theta) = \ldots \quad n = 1, 2, 3, \ldots$$

$$C = 3 \times 10^8 \text{m/s} = \frac{1}{\sqrt{\varepsilon_0 \mu_0}} \quad l = 0, 1, 2, \ldots$$

$$u = \frac{1}{2} \varepsilon_o E^2 + \frac{1}{2\mu_o} B^2 \quad d \sin(\theta) = \ldots \quad l = n-1, n-2, \ldots$$

$$m = -l, -l+1, \ldots, l-1, l$$

$$\theta_i = \theta_r \quad n = c/v \quad m = -\frac{i}{o} \quad 2l + 1$$

$$n_1 \sin(\theta_1) = n_2 \sin(\theta_2) \quad \frac{1}{f} = \frac{1}{o} + \frac{1}{i}$$

PLAN OF THE COURSE: LECTURES

- Lecture 1. Introduction to the course. Force field. Docking.
- Lecture 2 (double). Molecular Dynamics (MD).
- Lecture 3. Solvent in biomolecular modelling.
- Lecture 4. Protein folding.
- Lecture 5. Computational glycobiology.
- Lecture 6. Basics of QM.
- Lecture 7. MD, QM and NMR.
- Lecture 8 (double). DNA and DNA-protein interactions.

+ Case studies

PLAN OF THE COURSE: SEMINARS

- **Research papers related to lectures**

- **Presentation of the paper: MAX 10'+5'**
  - Introduction (Motivation)
  - **Methodology**
  - Results + Conclusions
  - *Critics: strong/weak points

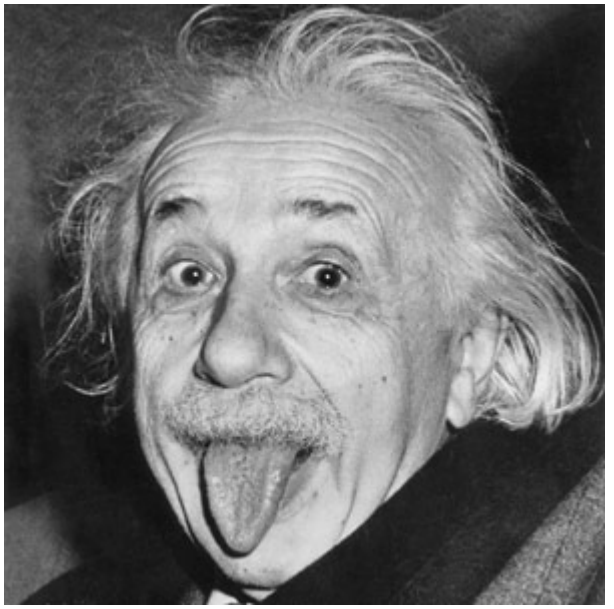
- **Questions and discussion**

LECTURE 1: OUTLINE
FORCE FIELD AND DOCKING

- Biomolecular modelling
- Force field:
  - Forces at the molecular level
  - Parameters derivation
  - Force field applicability
- Molecular docking
- Case study: inhibition of angiotensine converting enzyme

WHY NOT JUST MAKING EXPERIMENTS?

- **Time**
- **Money**
- **Complementarity to experiments**



"*If the facts don't fit the theory, change the facts*".

Albert Einstein

MOLECULAR MODELLING OBJECTIVES

- Use of theoretical methods and computational techniques for modelling and mimicking the behaviour of molecules

Areas:

- Computational chemistry
- Computational biology
- Material science

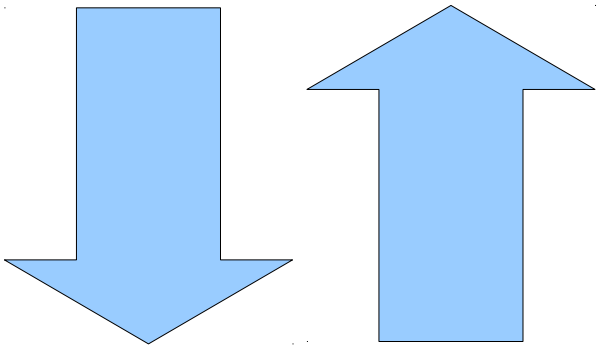
- Force field determines the behaviour of each individual atom and, consequently, of the whole system

WHY FORCE FIELD IS NEEDED?

- To distinguish different atoms with different properties
- To describe the environment of the atoms
- To model physical nature of the interactions between the atoms

Model

Testing force field



Creation of force field

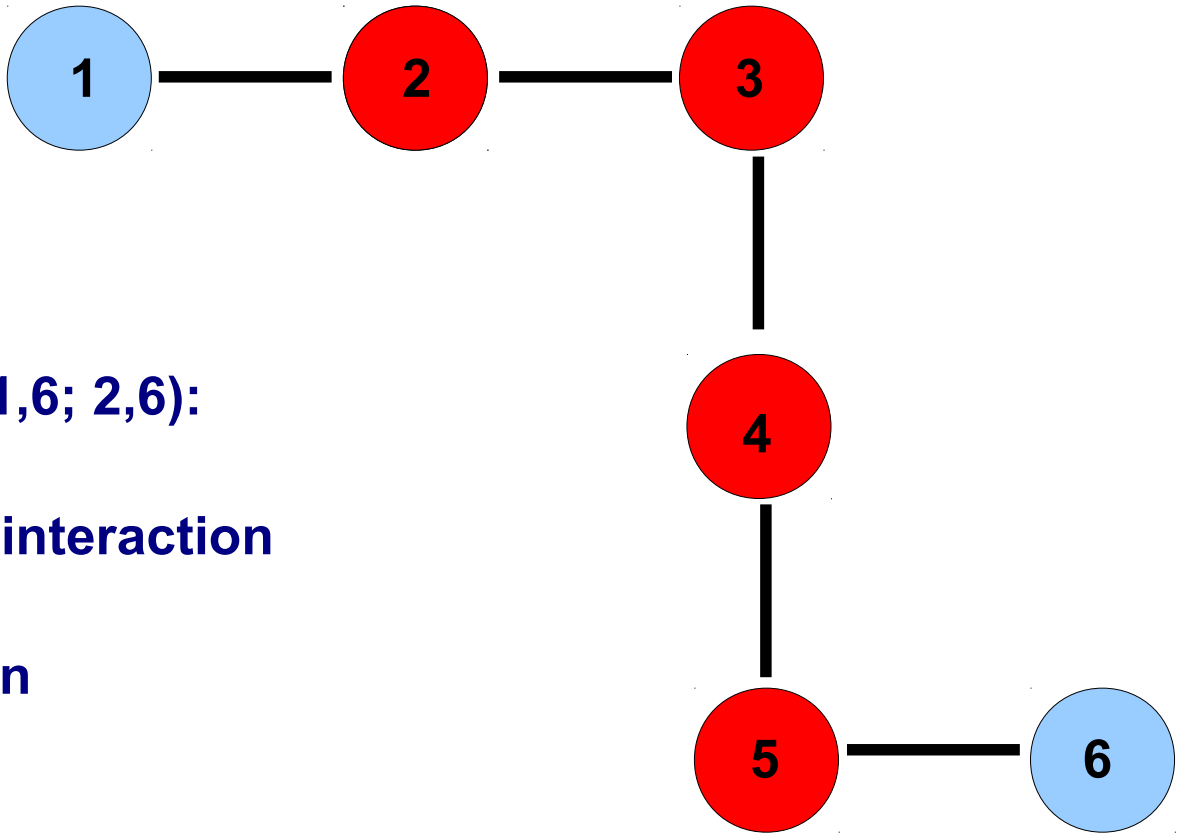
Experiment

- Force field = form + parameters set (atomic types), which describe potential energy of a system of particles

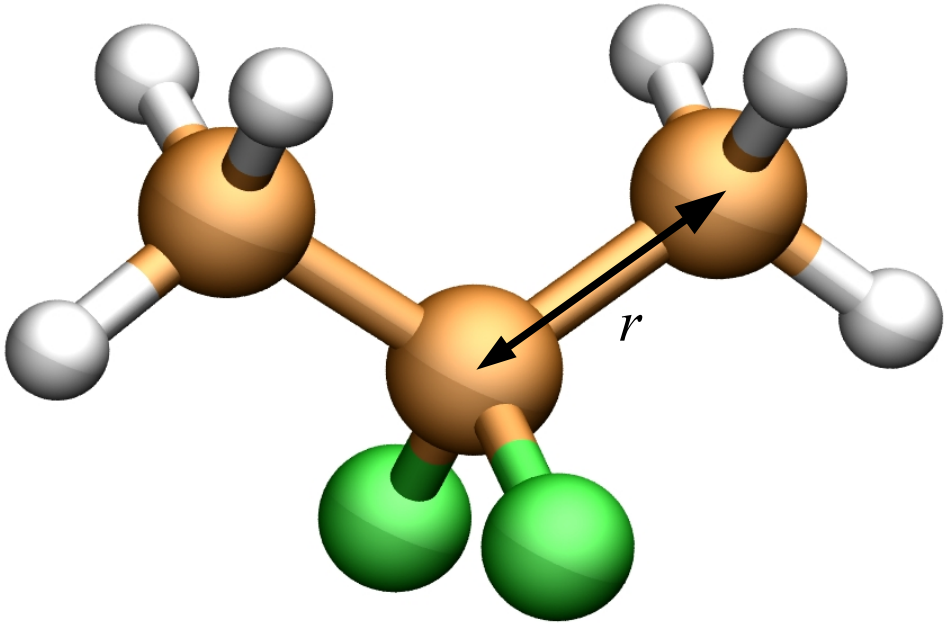
MOLECULAR FORCES

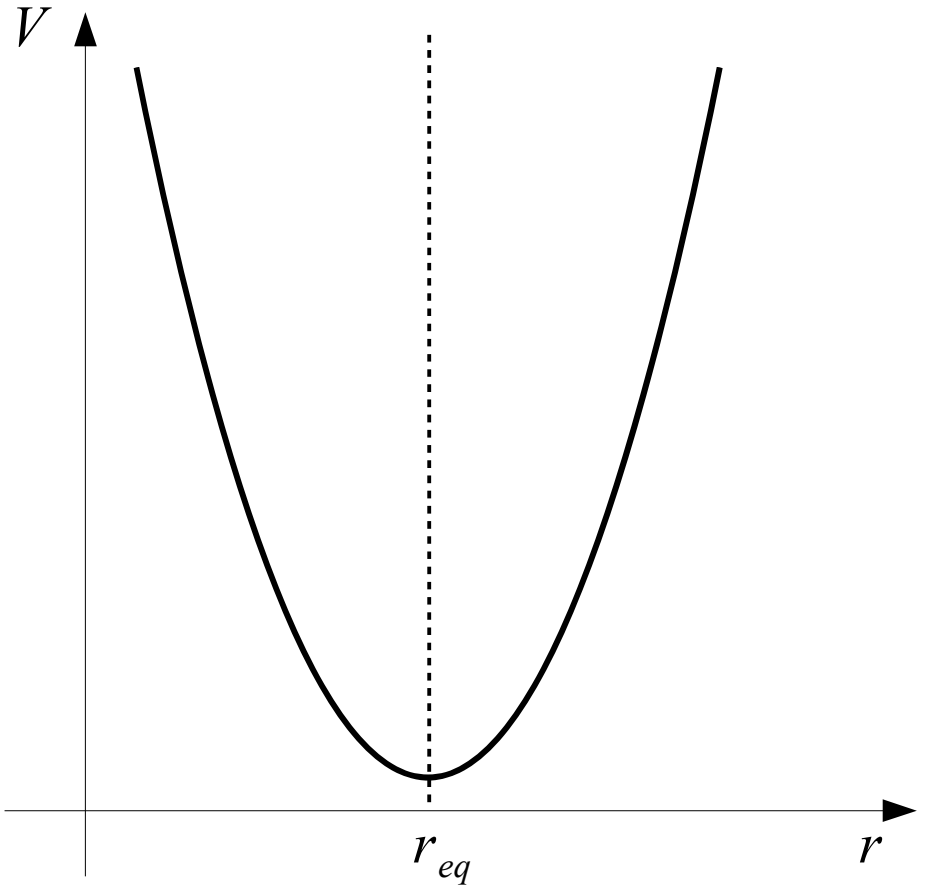
$$\vec{F}_i = \frac{-\delta V}{\delta \vec{r}_i}$$

- For atoms separated by 1, 2, 3 covalent bonds (2,3,4,5; 3,4,5,6; 1, 2, 3, 4):
  - Bond
  - Angle
  - Dihedral
- For all other atoms (1,5; 1,6; 2,6):
  - Electrostatic (Coulomb) interaction
  - Van der Waals interaction



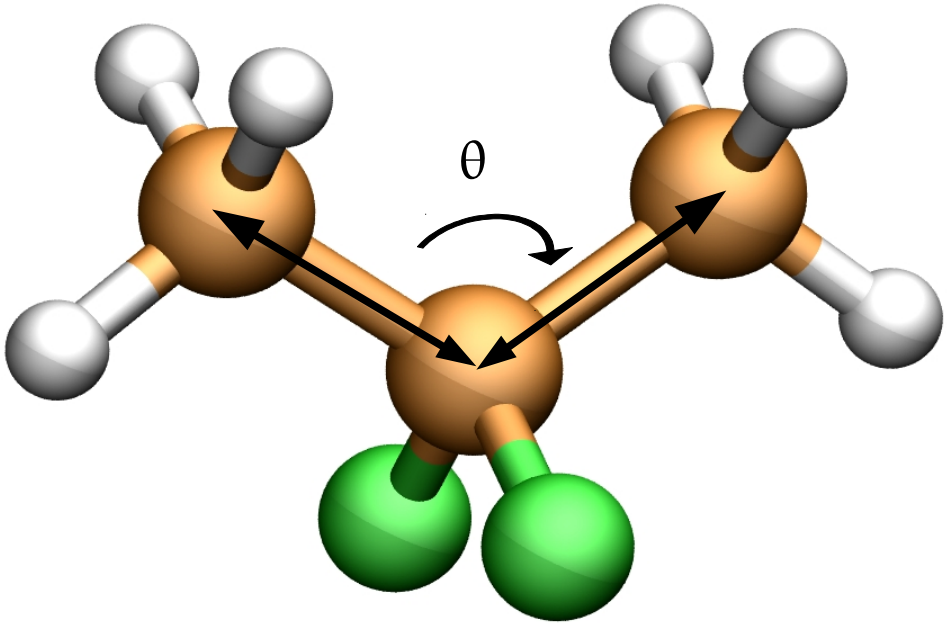
BONDS

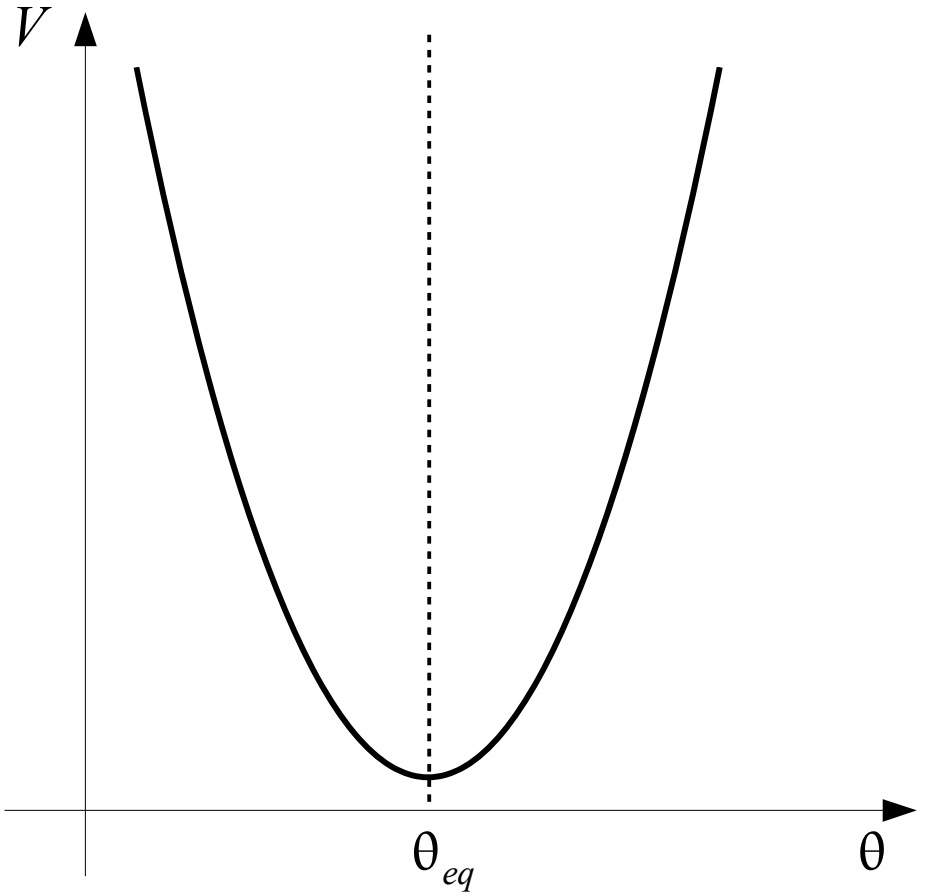




- **No approximation:** $V(r)=C_2(r-r_{eq})^2+C_3(r-r_{eq})^3+C_4(r-r_{eq})^4+\ldots C_n(r-r_{eq})^n$
- **Harmonic approximation:** $V(r)=K(r-r_{eq})^2,\ K=\frac{d^2V}{2dr^2}$
- **Each pair of atoms: 2 parameters** K, r_{eq}

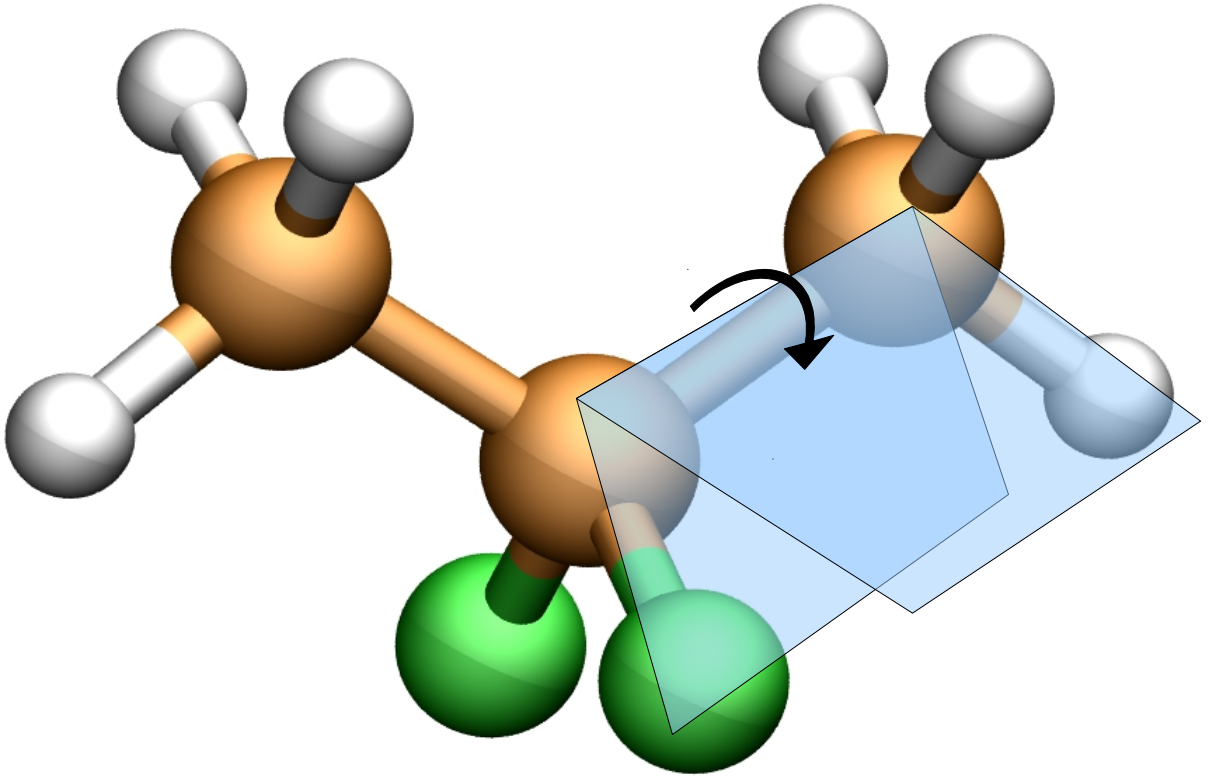
ANGLES

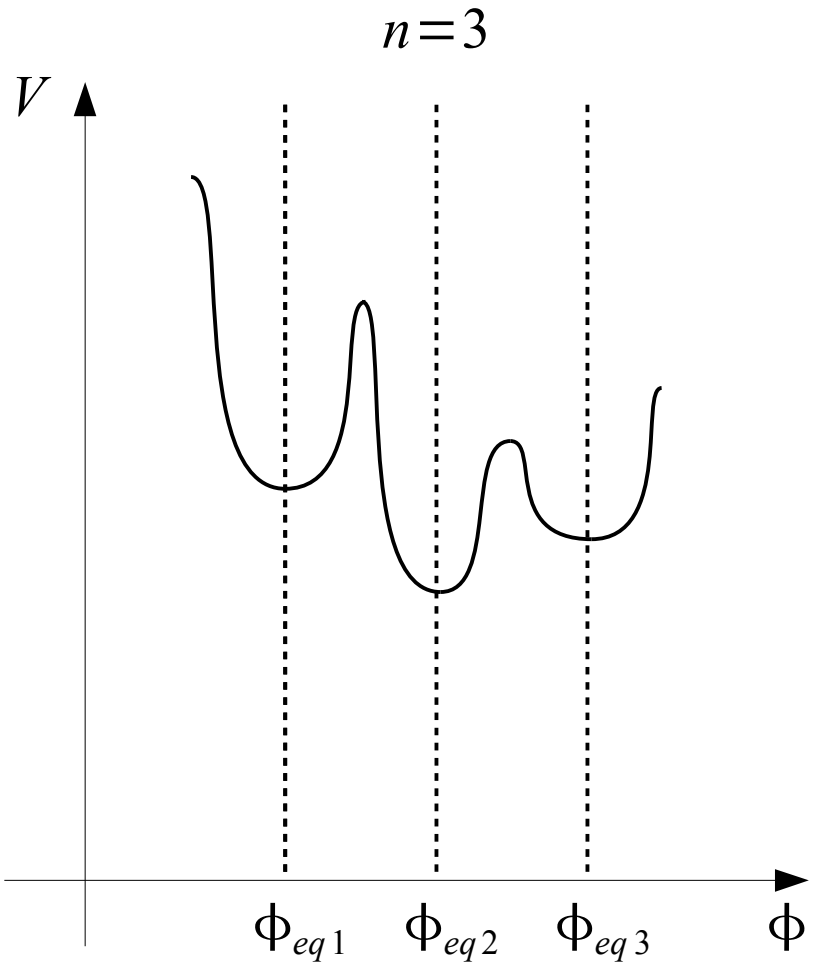




- **Harmonic approximation:** $V(\theta)=K(\theta-\theta_{eq})^2 \qquad K,\theta_{eq}-?$
- **Each three different atoms: 3!=6 parameters (A-B-C, A-C-B, B-A-C)**

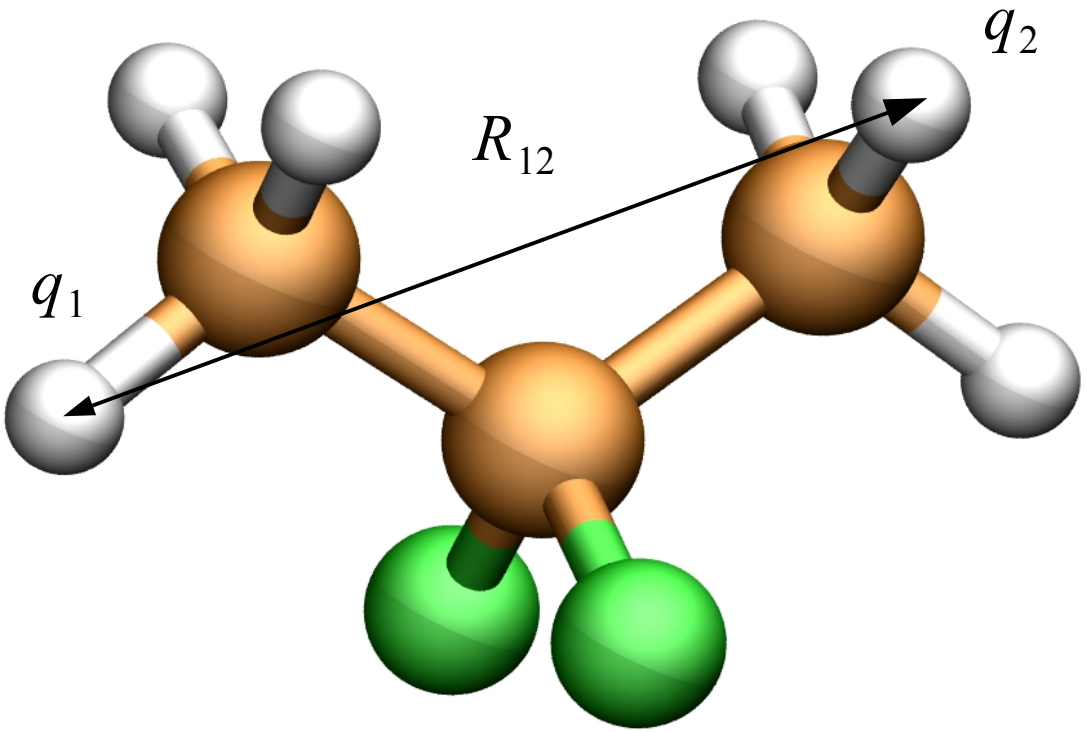
DIHEDRALS

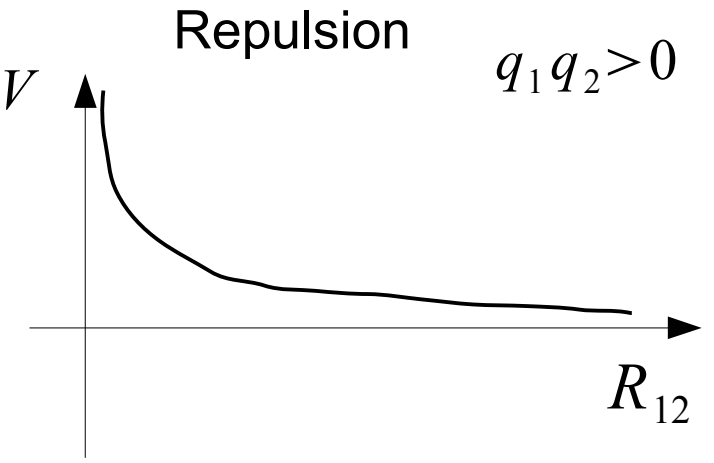


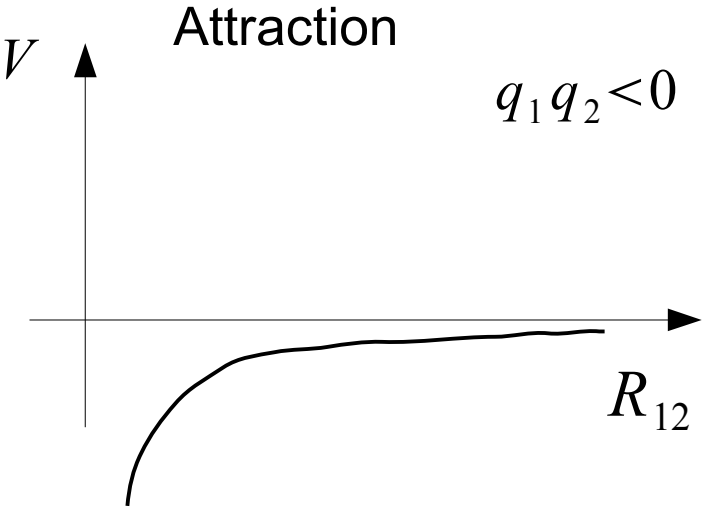


- **Harmonic approximation:** $V(\phi)=\frac{V_n}{2}(1+\cos[n\phi-\gamma])$
- **Each four different atoms: 4!(n+1) parameters**

ELECTROSTATICS

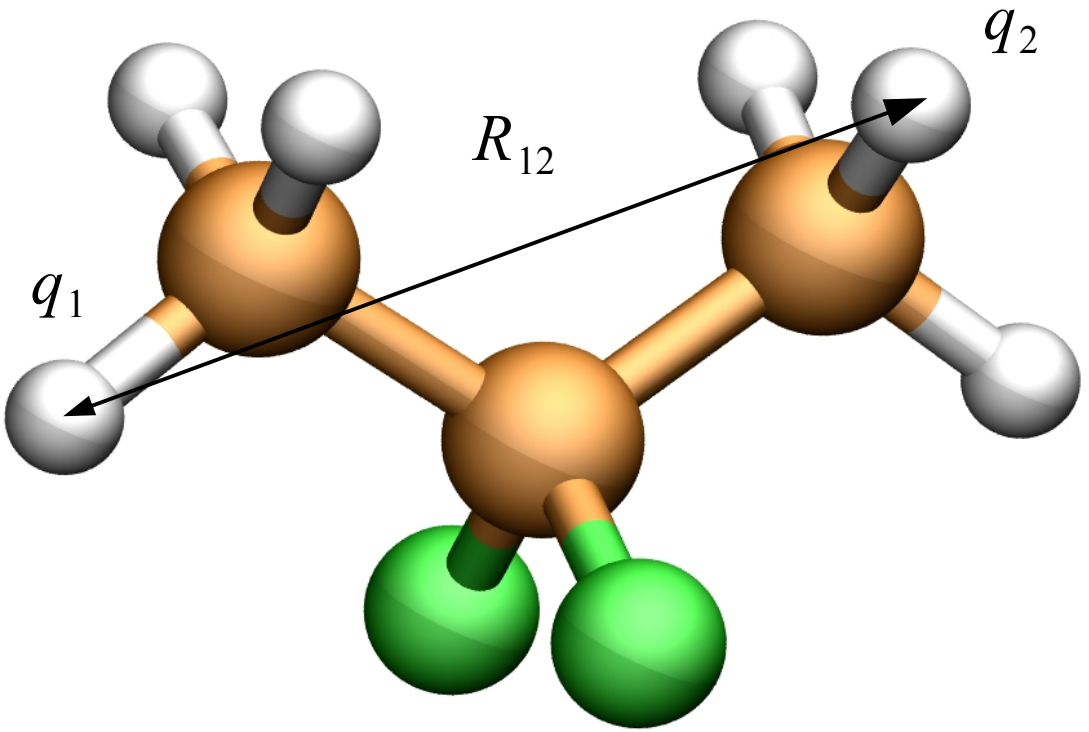


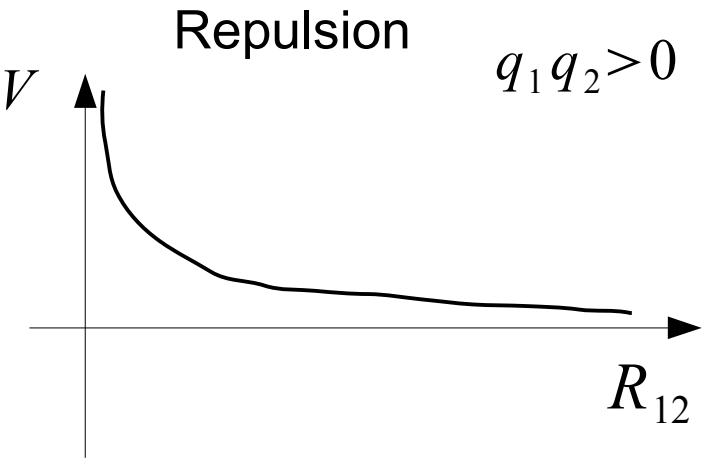


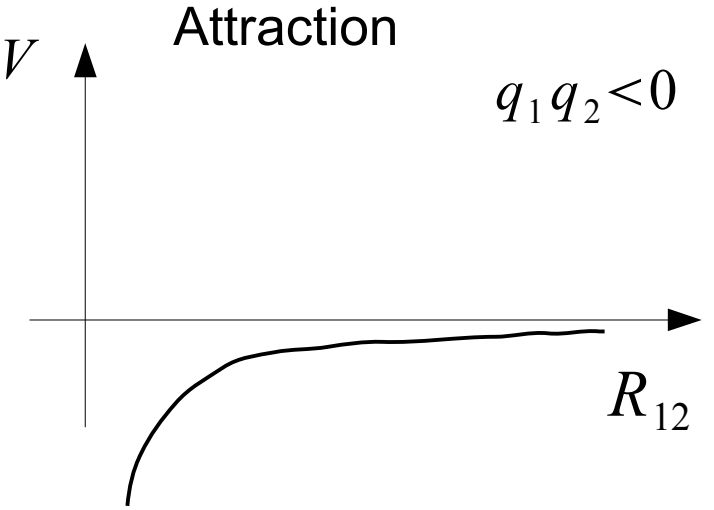


- **Coulomb interaction:** $V(R_{12}) = \frac{q_1 q_2}{\epsilon R_{12}}$
- **How to define dielectric constant** ϵ **?**
- **For n atoms (n-1)n/2 ~ $n^2/2$ contributions**
- **Each pair of atoms: 2 parameters** q_1, q_2

ELECTROSTATICS

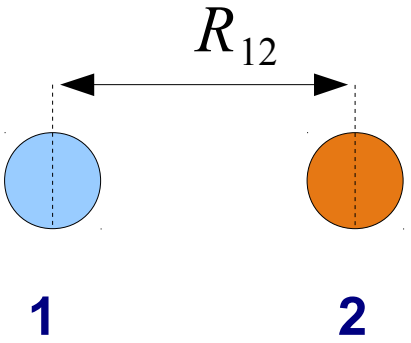


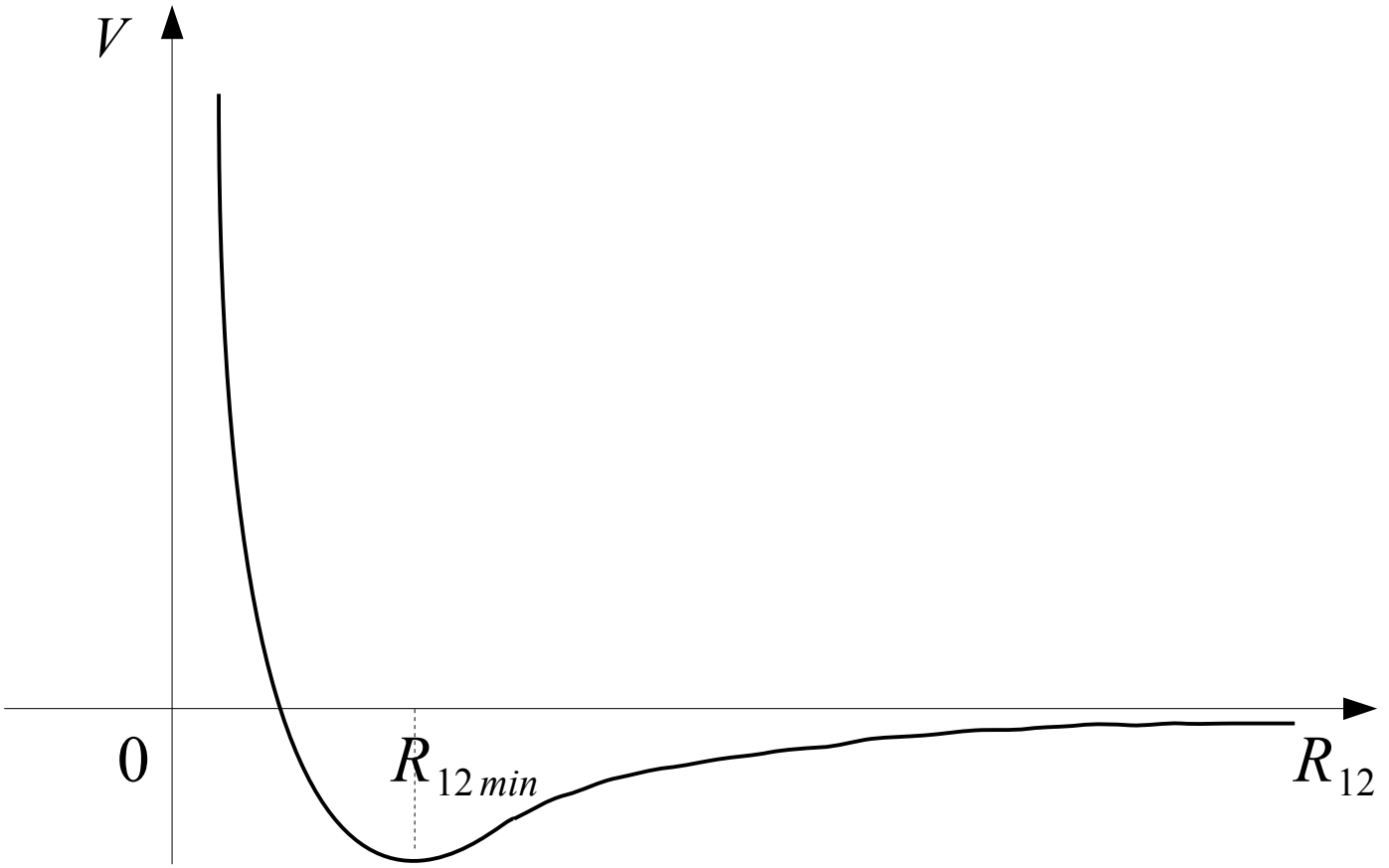




- **Coulomb interaction:** $V(R_{12}) = \frac{q_1 q_2}{\epsilon R_{12}}$
- **How to define dielectric constant** ϵ **?**
- **For n atoms (n-1)n/2 ~ $n^2/2$ contributions**
- **Each pair of atoms: 2 parameters** q_1, q_2

VAN DER WAALS



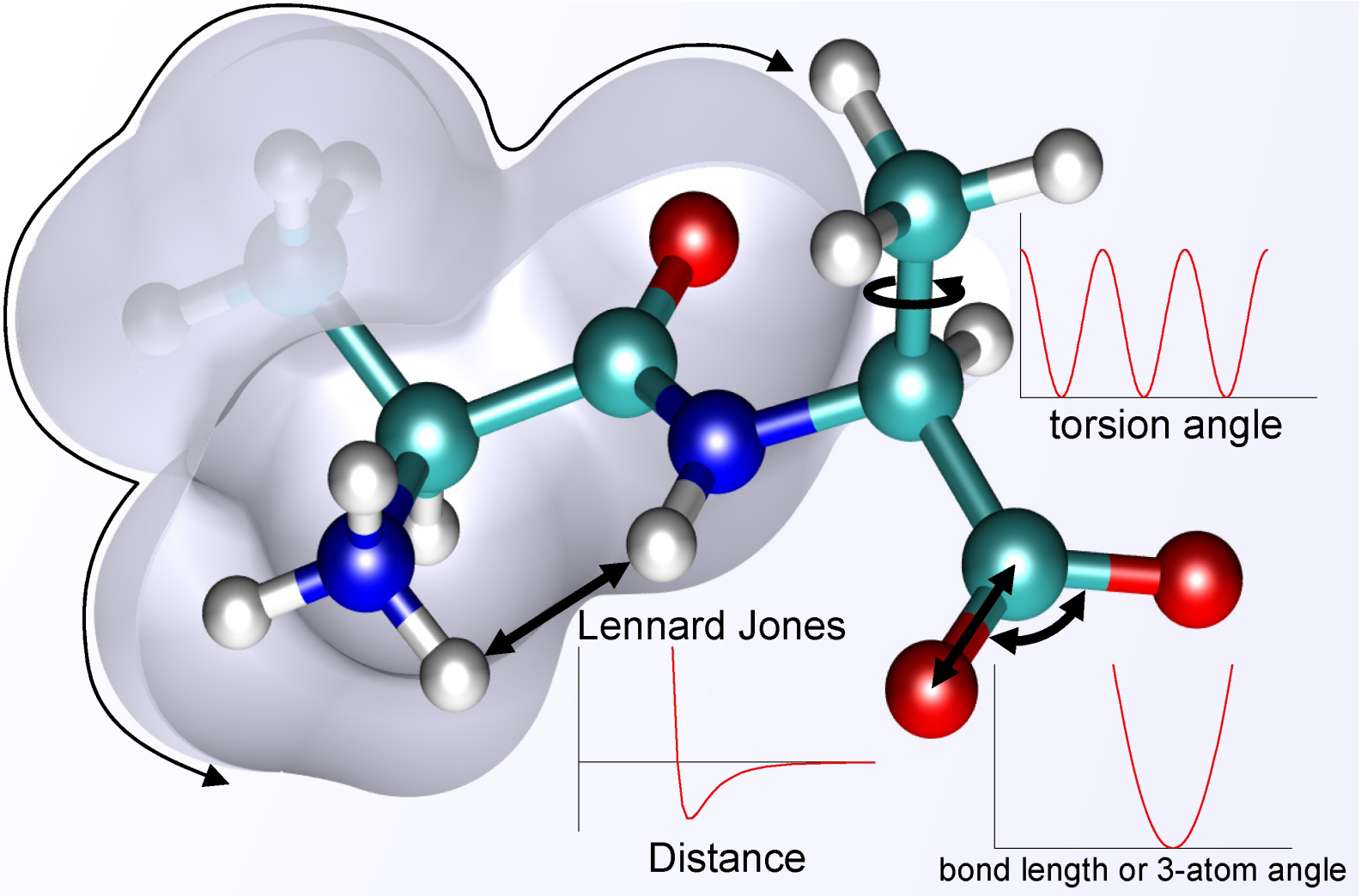


- **Lennard-Jones potential:** $V(R_{12}) = \frac{A_{12}}{R_{12}^{12}} - \frac{B_{12}}{R_{12}^{6}}$
- **For n atoms (n-1)n/2 ~ $n^2/2$ contributions**
- **Each pair of atoms: 2 parameters**

FORCE FIELD: SUMMING UP

$$V(\vec{r}) = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}(1 + \cos[n\phi - \gamma]) + \sum_{i<j}^{atoms} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right) + \sum_{i<j}^{atoms} \frac{q_i q_j}{\epsilon R_{ij}}$$

$$\vec{F}_i = \frac{-\delta V}{\delta \vec{r}_i}$$

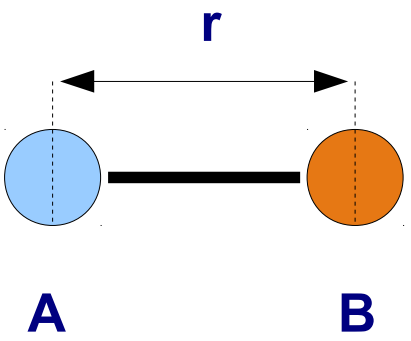


- **Force field = form + parameters set (atomic types), which describe potential energy of a system of particles**

FORCE FIELD: PARAMETERS DERIVATION

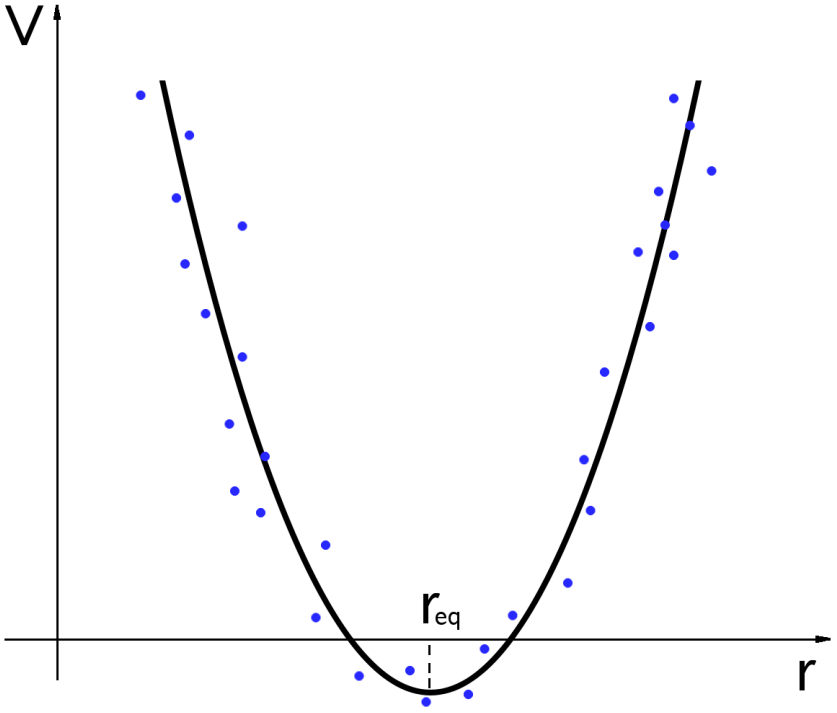
- *Ab initio* calculations (quantum chemistry)
- Experimental data (X-Ray, NMR, Spectroscopy, Calorimetry etc.)
- Learning algorithms to fit parameters to describe macroscopic properties (density, viscosity, energies of phase transitions etc.)

EXAMPLE: bond length r (A-B)



$$V(r) = K(r - r_{eq})^2$$

$$K, r_{eq} - ?$$



FORCE FIELD FILE EXAMPLE

Atom types

```
C  12.01        0.616  !        sp2 C carbonyl group
CA 12.01        0.360           sp2 C pure aromatic (benzene)
CB 12.01        0.360           sp2 aromatic C, 5&6 membered ring junction
CC 12.01        0.360           sp2 aromatic C, 5 memb. ring HIS
```

Bonds

```
…
C -C   310.0     1.525      Junmei et al, 1999
C -CA  469.0     1.409      JCC,7,(1986),230; (not used any more in TYR)
C -CB  447.0     1.419      JCC,7,(1986),230; GUA
C -CM  410.0     1.444      JCC,7,(1986),230; THY,URA
```

Angles

```
…
C -C -O       80.0      120.00    Junmei et al, 1999 acrolein
C -C -OH      80.0      120.00    Junmei et al, 1999
CA-C -CA      63.0      120.00    changed from 85.0  bsd on C6H6 nmodes; AA
CA-C -OH      70.0      120.00    AA (not used in tyr)
```

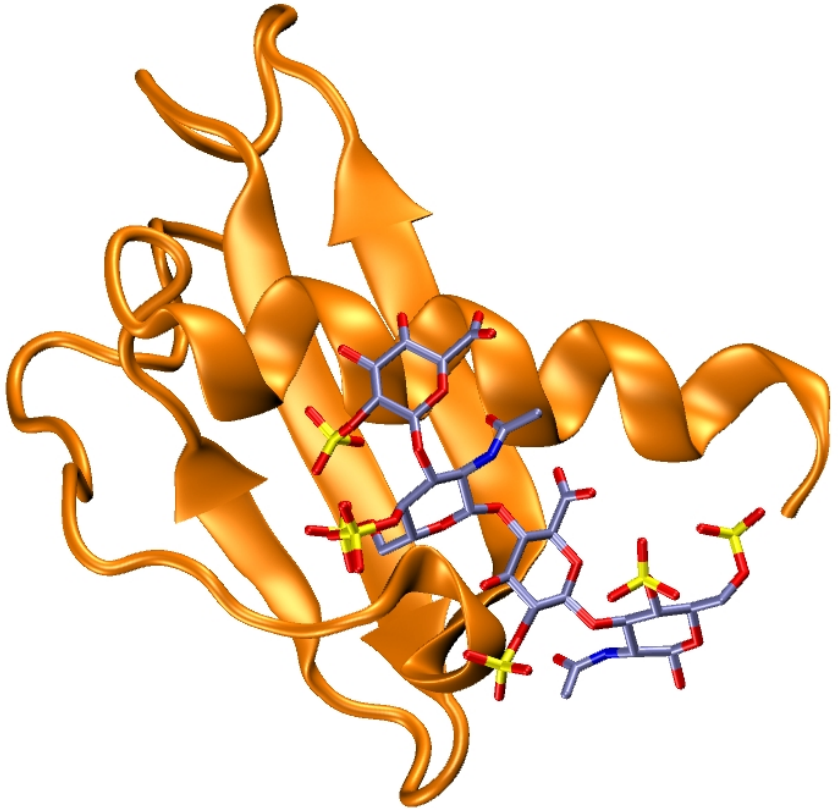
Dihedrals

```
...
CK-CB-N*-CT       1.0        180.        2.
CM-C -N*-CT       1.0        180.        2.          dac guess, 9/94
CM-C -CM-CT       1.1        180.        2.
CT-O -C -OH       10.5       180.        2.
```

Lennard-Jones

```
…
H          0.6000  0.0157        !Ferguson base pair geom.
HO         0.0000  0.0000         OPLS Jorgensen, JACS,110,(1988),1657
HS         0.6000  0.0157        W. Cornell CH3SH --> CH3OH FEP
HC         1.4870  0.0157        OPLS
```

FORCE FIELD: APPLICATIONS



Relaxation → **?**

Behaviour in time → **?**

Ligand position → **?**

- **Minimization**
- **Molecular dynamics (MD)**
- **Docking**

MM and MD OVERVIEW

Atomic positions (coordinate file)

Covalent structure (topology file)

Potential energy function (parameter file)

Additional atoms (hydrogens, heteroatoms, solvent, counterions)

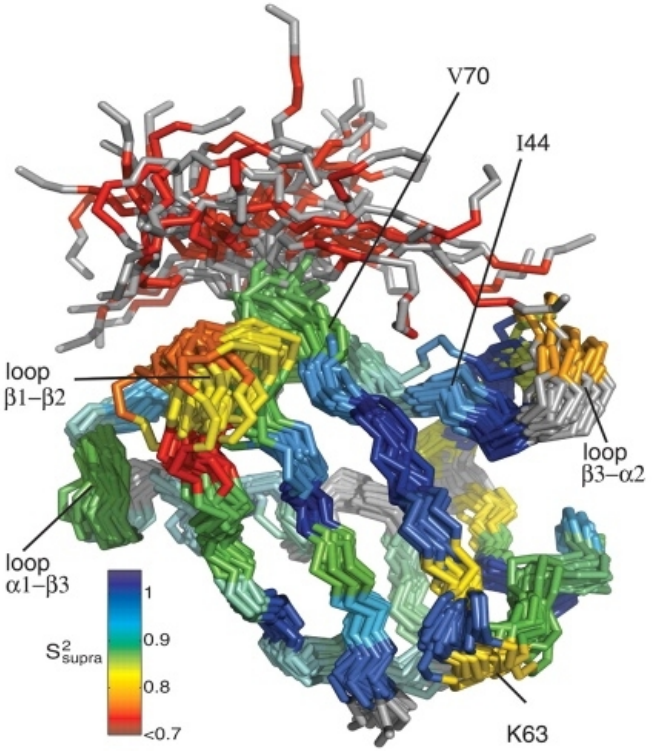
*Special features (periodic boundary conditions, *constant temperature *constant pressure)*

**Atomic velocities*

Total potential energy

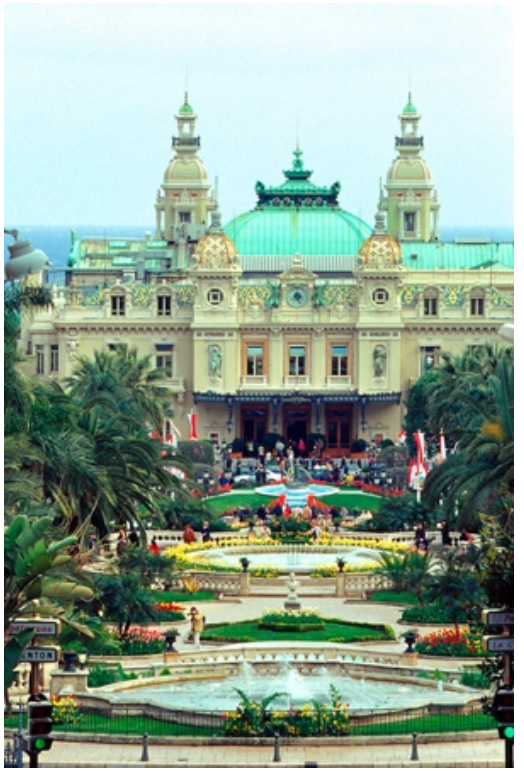
Forces on each atom

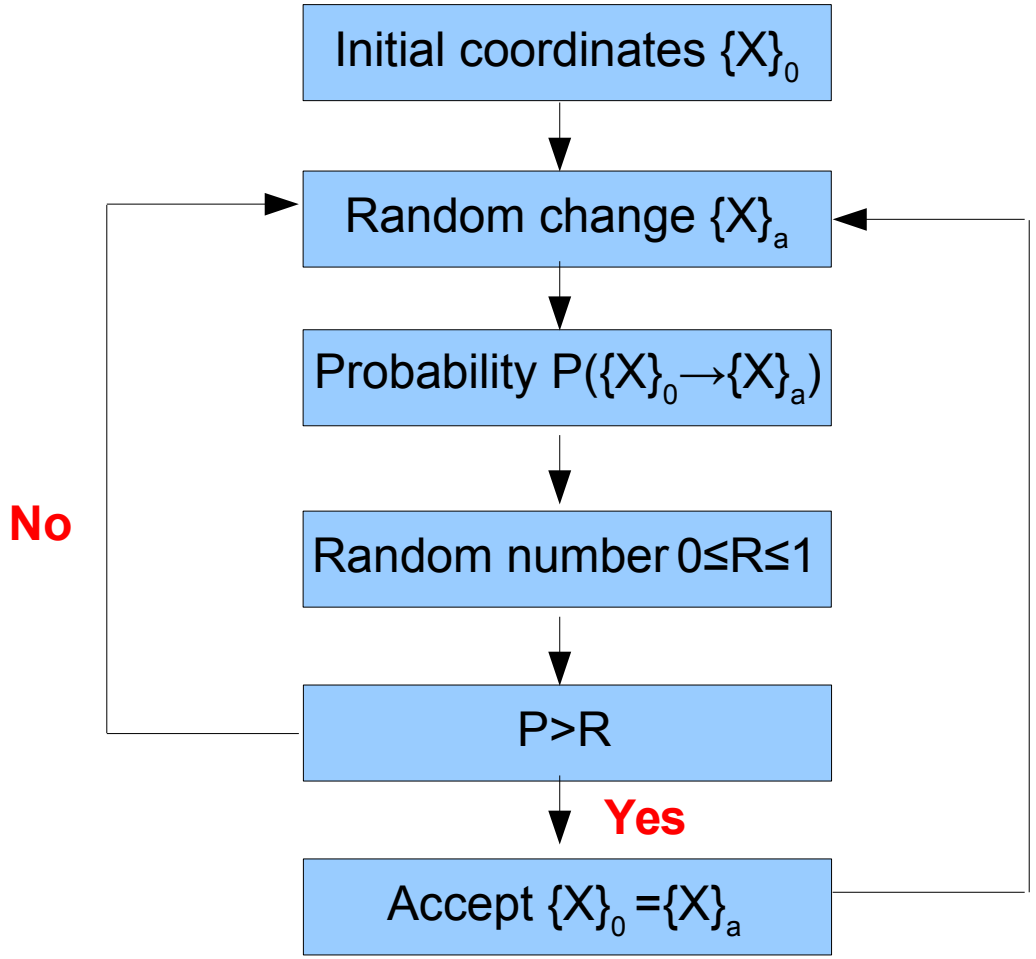
**Effective temperature*



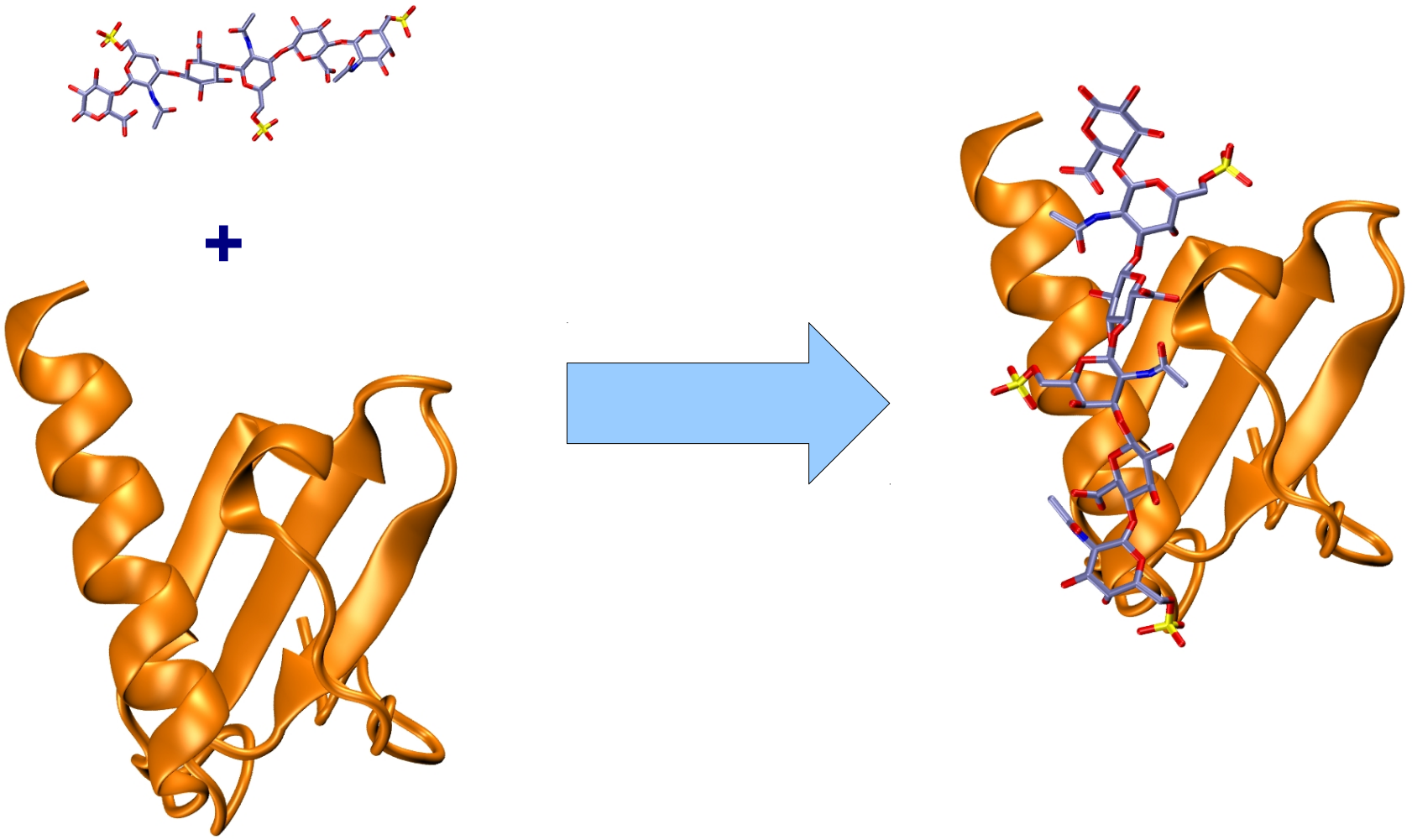
Monte Carlo Method

- **Monte Carlo Methods make up a class of computational algorithms that rely on repeated random sampling.**





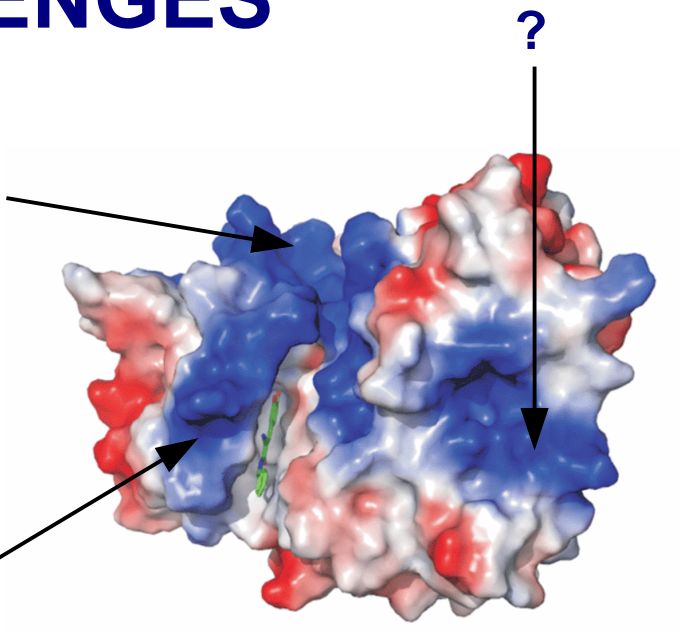
MOLECULAR DOCKING: DEFINITION

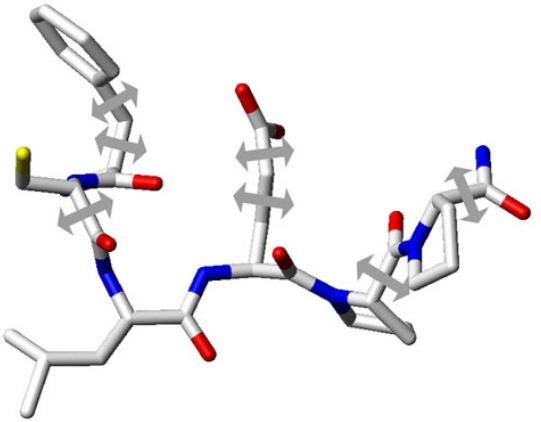


- **Molecular Docking is a computational approach, which predicts the binding site and binding conformation of one molecule in relation to a second when they are bound to each other to form a stable complex**

DOCKING CHALLENGES

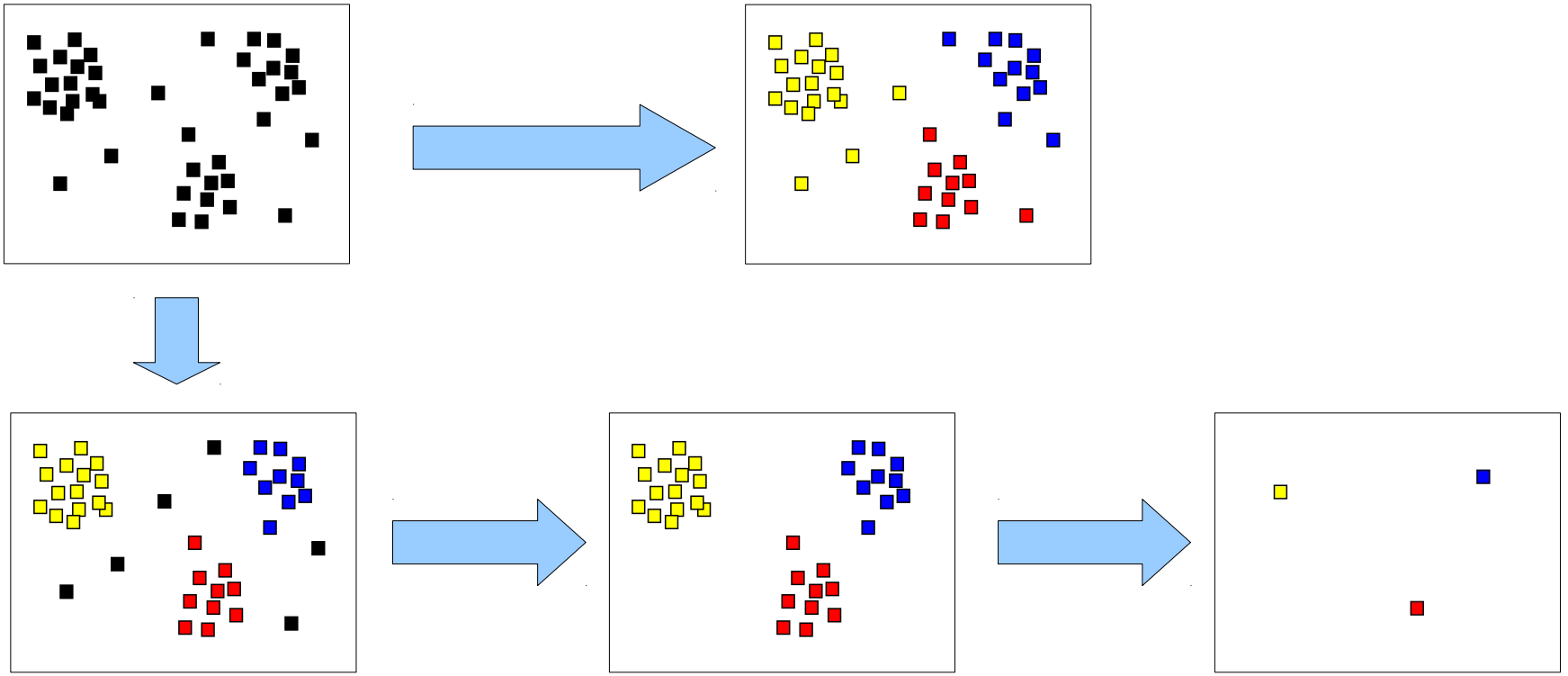
- **Search for the binding site:**
  - **Size/flexibility of the receptor**
  - **Heterogeneity of the receptor surface**
  - **Accuracy of prediction/size of ligand**
- **Search for the best conformation of ligand:**
  - **Size of ligand (degrees of freedom)**
  - **Symmetry of ligand**
  - **Receptor flexibility**
- **Scoring and clustering**
  - **Force field**
  - **Clustering procedure**





$$V(\vec{r}) = \sum_{bonds} K_r (r - r_{eq})^2 + \sum_{angles} K_\theta (\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}(1 + \cos[n\phi - \gamma]) + \sum_{i<j}^{atoms} \left(\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}}\right) + \sum_{i<j}^{atoms} \frac{q_i q_j}{\epsilon R_{ij}}$$

CLUSTERING

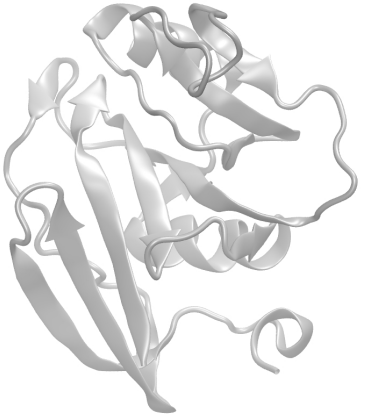


- Clustering allows to discard false positives and to find representative solutions

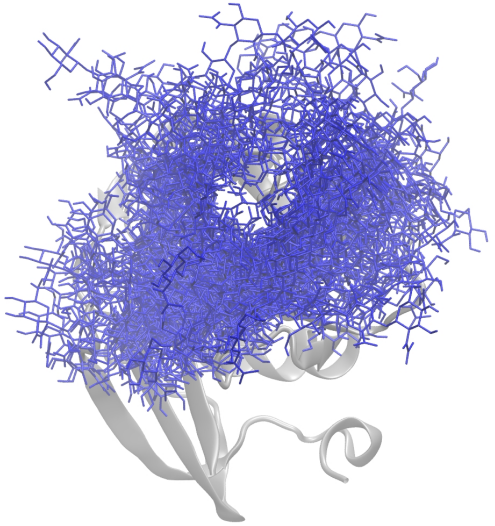
CLUSTERING

Solution 1 – Score 1
Solution 2 – Score 2
...
Solution n – Score n
…
Solution N – Score N

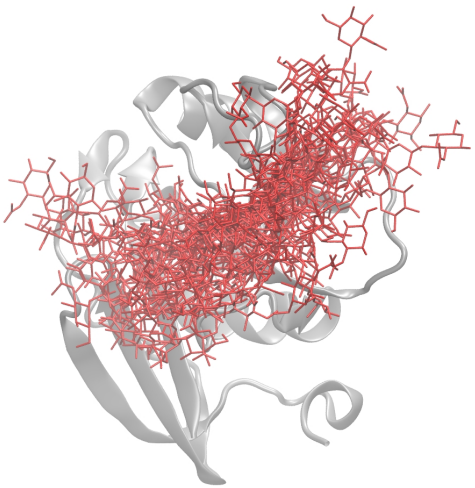
N structures



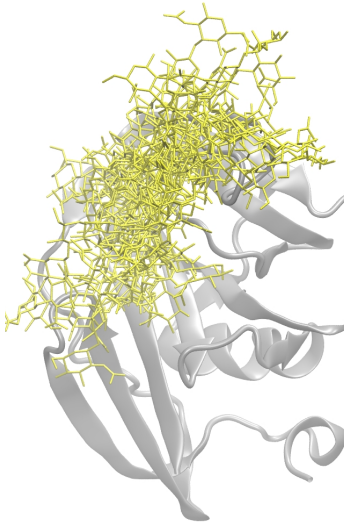
Receptor



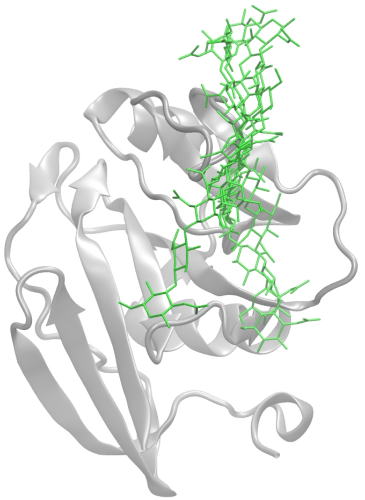
100 top solutions



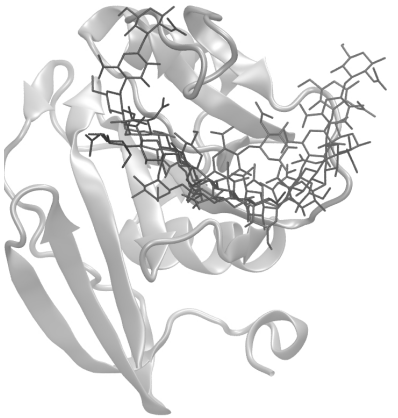
Cluster 1



Cluster 2



Cluster 3



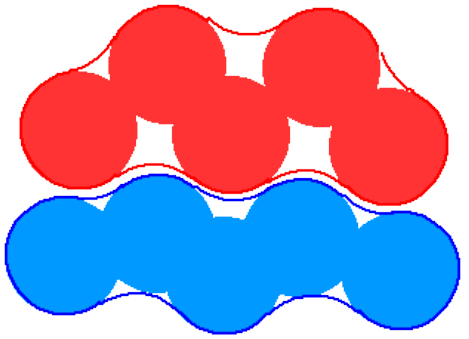
Cluster 4

DOCKING APPROACHES CLASSIFICATION

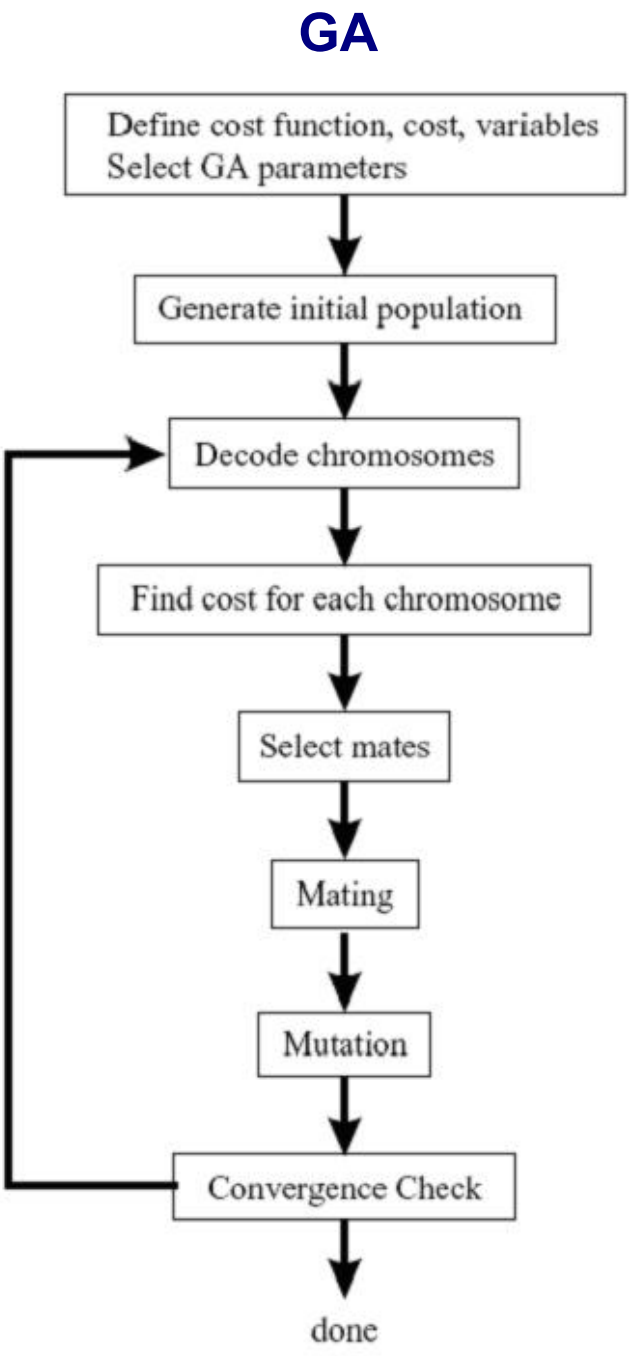
- Global/local
- Flexibility of the receptor
- Flexibility of ligand
- Taking solvent into account (explicitly/implicitly)
- By classes of molecules they are optimized for
- By algorithms

DOCKING ALGORITHMS

- Shape complementarity
- Genetic algorithms
- Simulated annealing
- Molecular dynamics



- Criteria for docking algorithm quality:
  - Precision/Recall/Accuracy/True negative rate
  - Reproducibility
  - Speed



DOCKING PIPELINE (Autodock)

Receptor structure

Force field 1, defining flexible residues

Receptor + charges + types

Running an atomic probe

Grid

Ligand structure

Force field 1, defining flexible angles

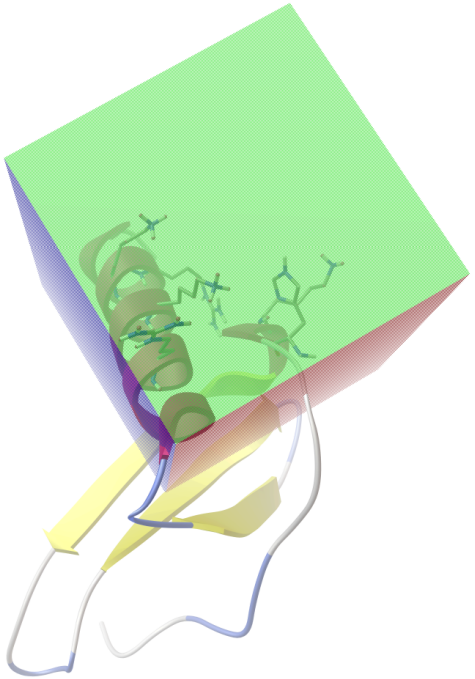
Ligand + charges + freedom degrees

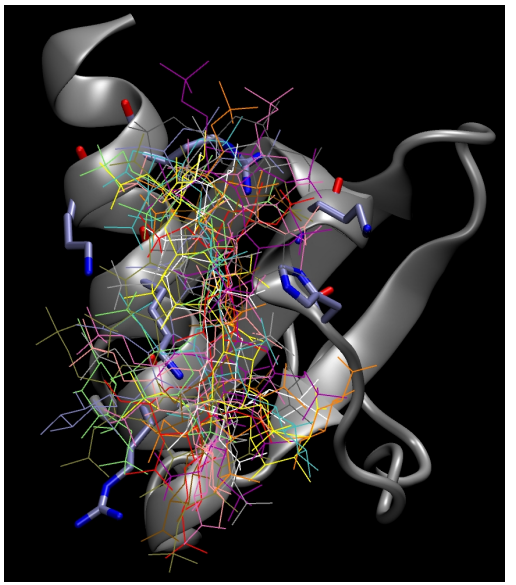
Fitting ligand to grid

All docking poses

Scoring/ranking with force field 2, (refinement)

Best docking poses





Autodock 3

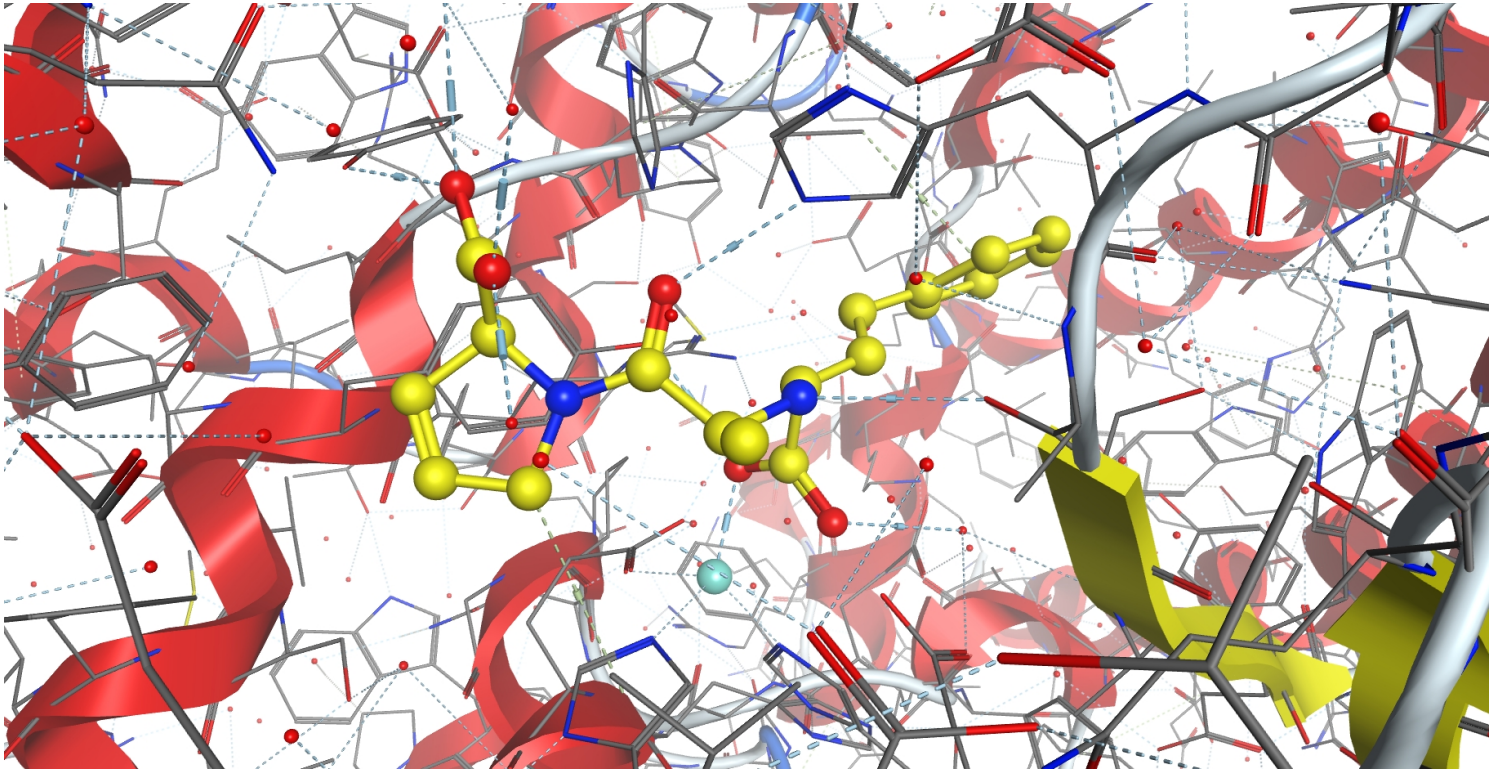
Force field 1 has a simpler form than force field 2

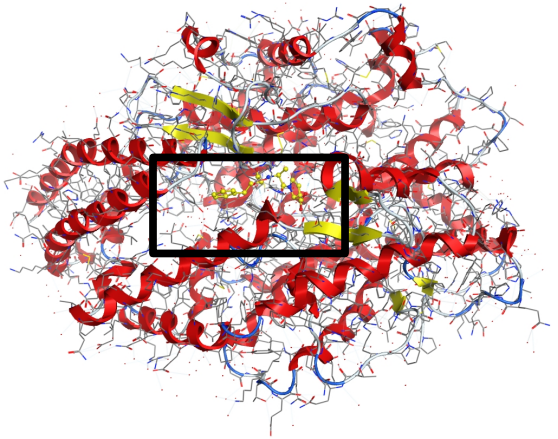
DOCKING PROGRAMS

- Autodock (Open Source)
- DOCK (Open Source)
- GOLD
- FlexX
- eHiTS
- Glide
- HADDOCK (Open Source)
- SLICK (Open Source, for sugars)

CASE STUDY: ACE ENZYME

- Aim: design and analysis of inhibitors
- Object: testicular angiotensin I-converting enzyme
- Methods: docking (Autodock 4), MD (AMBER 10) => energy calculations
- Comparison to experiment: inhibition activity (IC_{50})

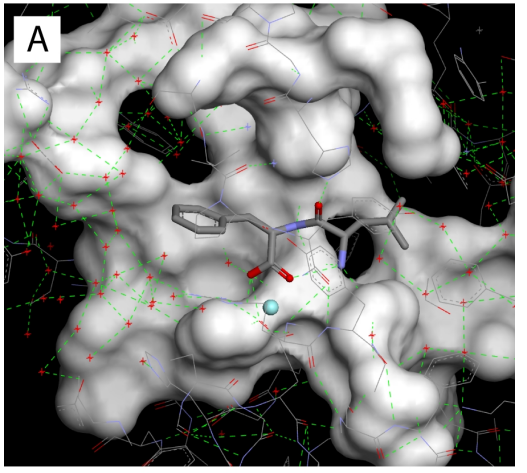




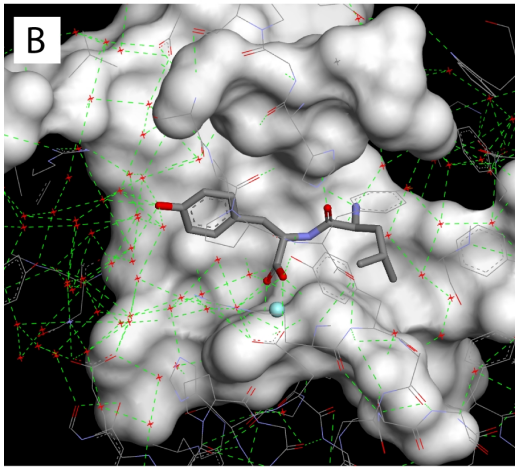
PDB ID: 1UZE, 1.82 Å

RESULTS: 1st GROUP OF LIGANDS

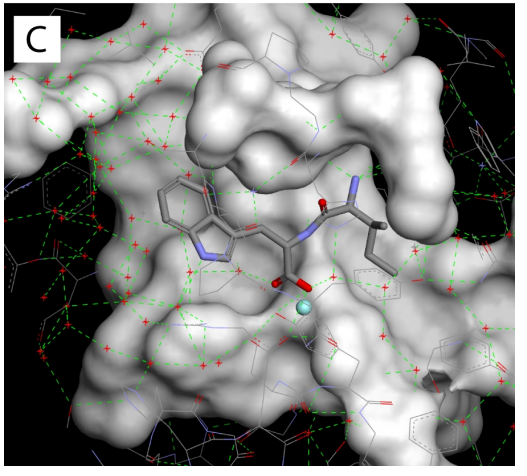
Leu-Phe: IC_{50} =349 μM, ΔG_{calc}= -63 kcal/mol



Leu-Tyr: IC_{50} =44 μM, ΔG_{calc}= -74 kcal/mol



Leu-Trp: IC_{50} =1.5 μM, ΔG_{calc}= -125 kcal/mol



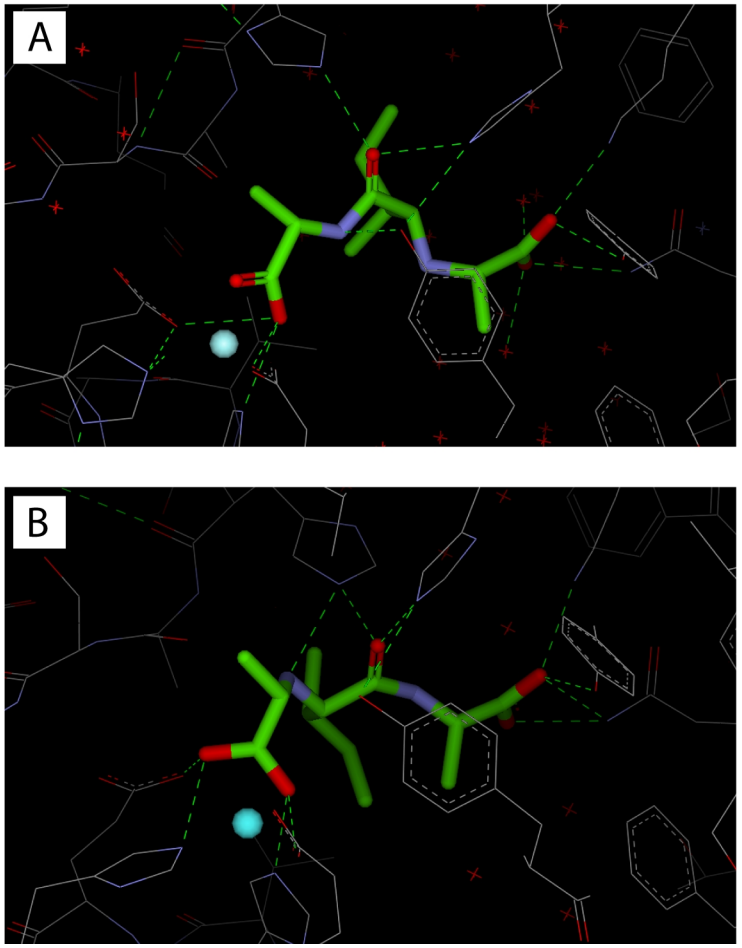
- **Docking+MD agree with the experiment and explain it in terms of physical interactions: EL+VDW**

RESULTS: 2nd GROUP OF LIGANDS

Ile-Ala: IC_{50} =910 μM, ΔG_{calc}= -72 kcal/mol

COO^--Ile-Ala: IC_{50} =25 μM, ΔG_{calc}= -225 kcal/mol

COO^-CH_3-Ile-Ala: IC_{50} =0.5 μM, ΔG_{calc}= -275 kcal/mol

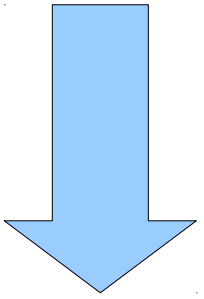


Symmetric binding of the carboxylated dipeptide

- Docking+MD agree with the experiment and explain it in terms of physical interactions: EL+VDW

HOWs IN THE CASE STUDY

- How to choose appropriate binding poses?
- How to decide if the analyzed binding pose is stable?
- How to calculate the energies?
- How to treat the solvent?
- How important is the electrostatic effect?



Molecular dynamics help

LECTURE 1: OUTLINE
FORCE FIELD AND DOCKING

- Biomolecular modelling
- Force field:
  - Forces at the molecular level
  - Parameters derivation
  - Force field applicability
- Molecular docking
- Case study: inhibition of angiotensine converting enzyme