Molecular modelling

Zdeněk Kříž
National Centre for Biomolecular Research
Brno, Czech Republic

zdenek@chemi.muni.cz

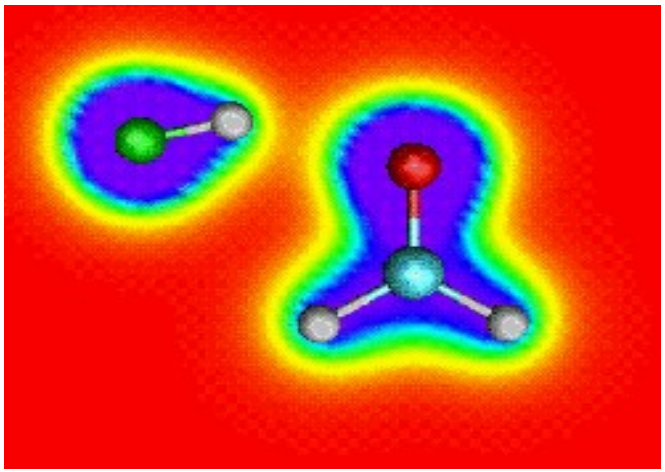
Computational chemistry

Computational chemistry is a branch of chemistry that uses principles of computer science to assist in solving chemical problems. It uses the results of theoretical chemistry, incorporated into efficient computer programs, to calculate the structures and properties of molecules and solids. Its necessity arises from the well-known fact that apart from relatively recent results concerning the Hydrogen molecular ion, the quantum n-body problem cannot be solved analytically, much less in closed form.

Molecular modelling

Molecular modelling encompasses all theoretical methods and computational techniques used to model or mimic the behaviour of molecules. The techniques are used in the fields of computational chemistry, computational biology and materials science for studying molecular systems ranging from small chemical systems to large biological molecules and material assemblies.

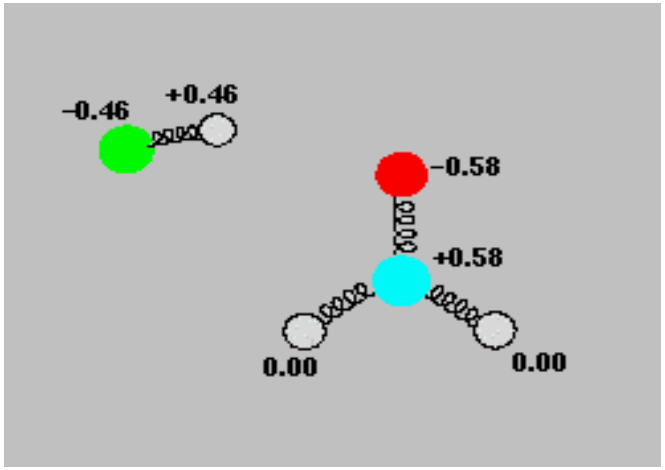
Quantum mechanics approach



Exact solution of Schroedinger Equation.

vs.

Molecular mechanics approach

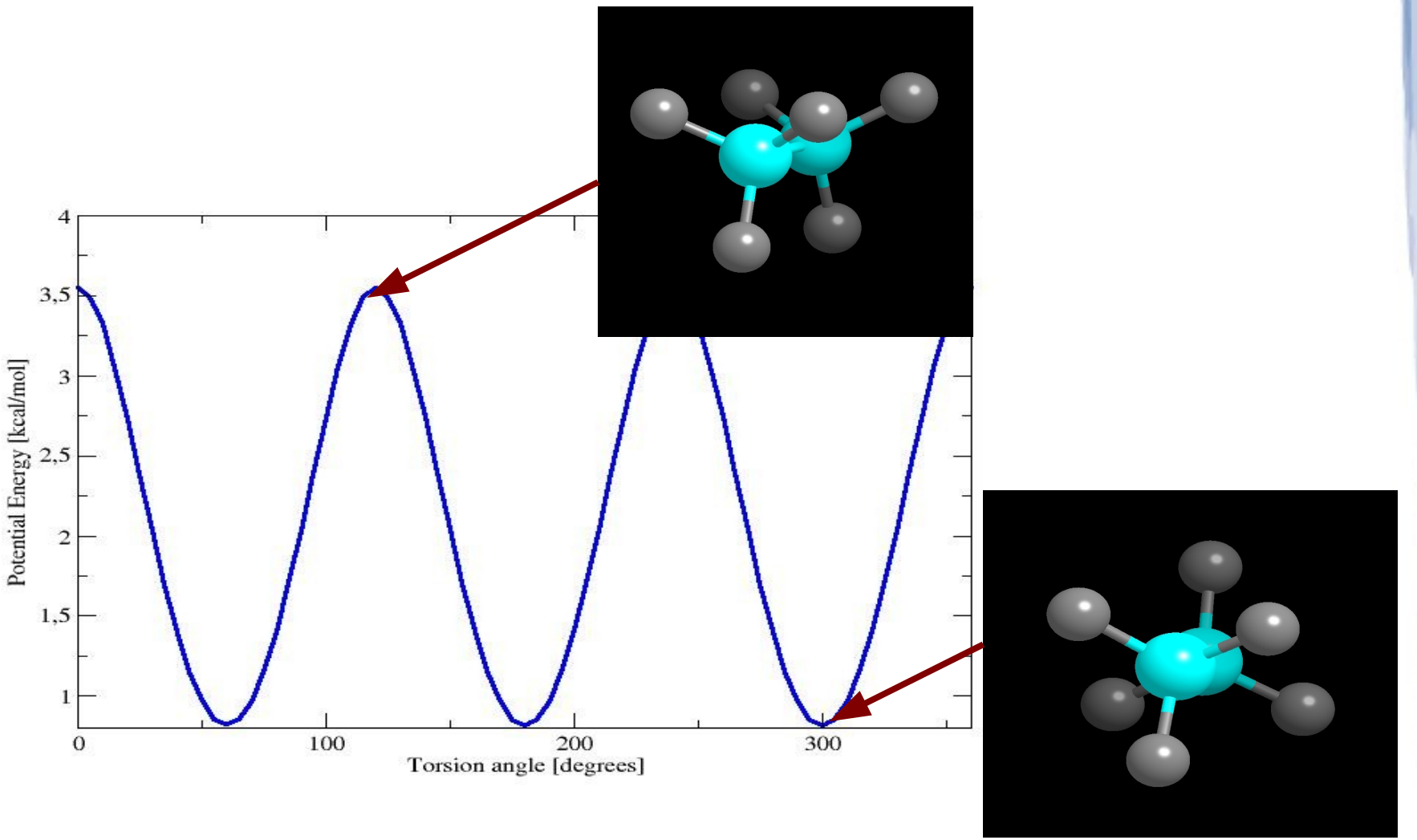


Using parameters. Fast, but not usable for chemical reactions study

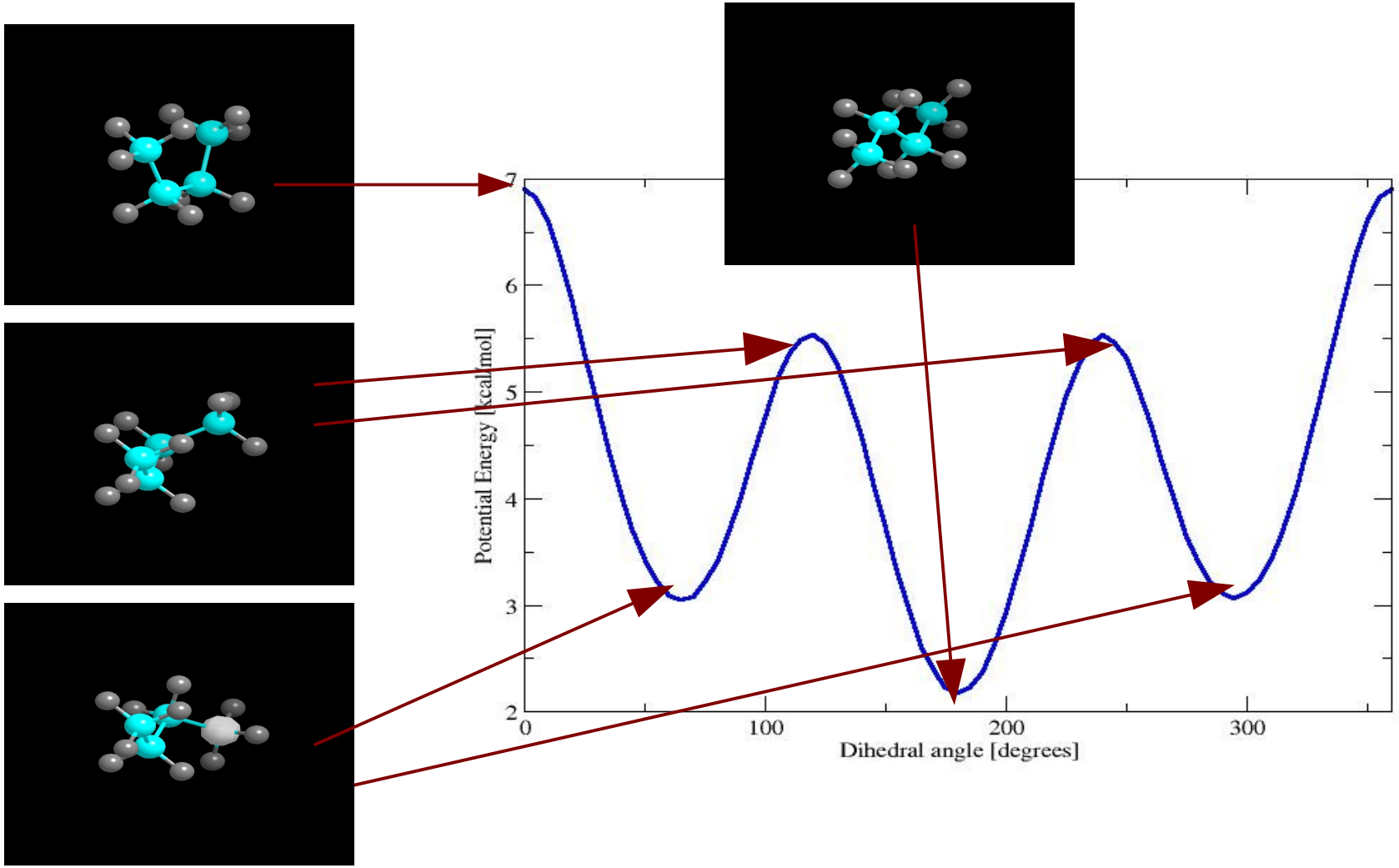
Outline

- Conformational analysis
- Potential energy (hyper)surface/Free energy (hyper)surface
- Searching methods
- Single coordinate driving method - SCD
- Results of CICADA searching
- Molecular Mechanics
- Molecular Dynamics

Conformational analysis



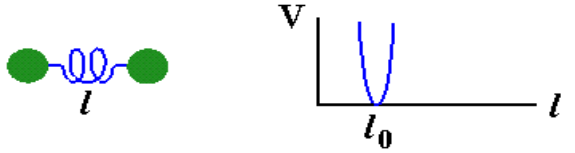
Conformational analysis



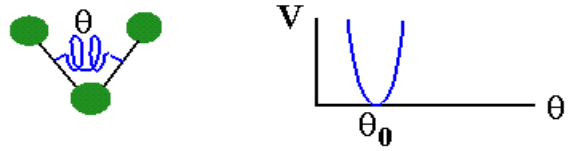
Potential / free energy

Empirical Potential Energy Function

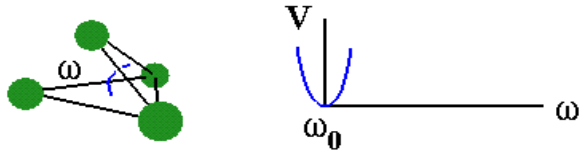
Bonds



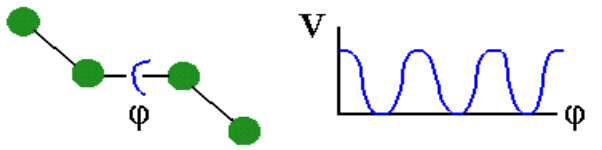
Angles



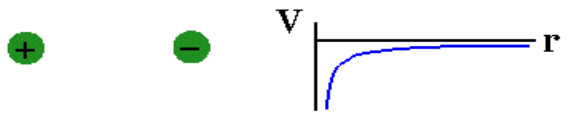
Improper Dihedral



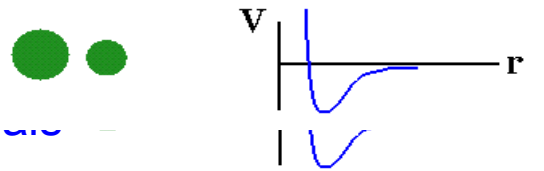
Torsions



Electrostatics



Van der Waals



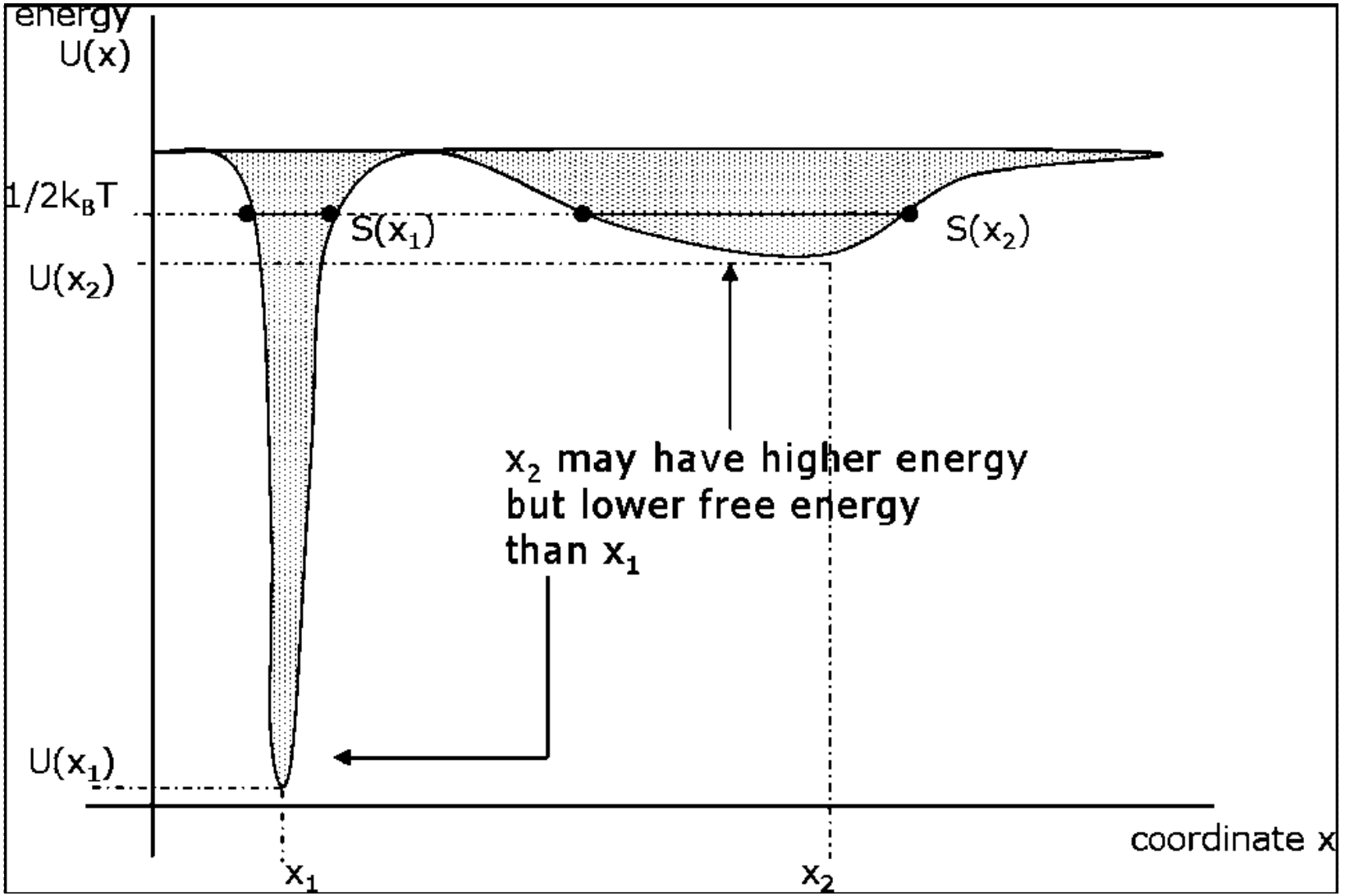
Free Energy

Constant NVT:

Helmholtz free energy
A=U-TS ,
Where U=<E> - average energy

Constatnt NPT
Gibbs free energy
G=U+PV-TS=H-TS

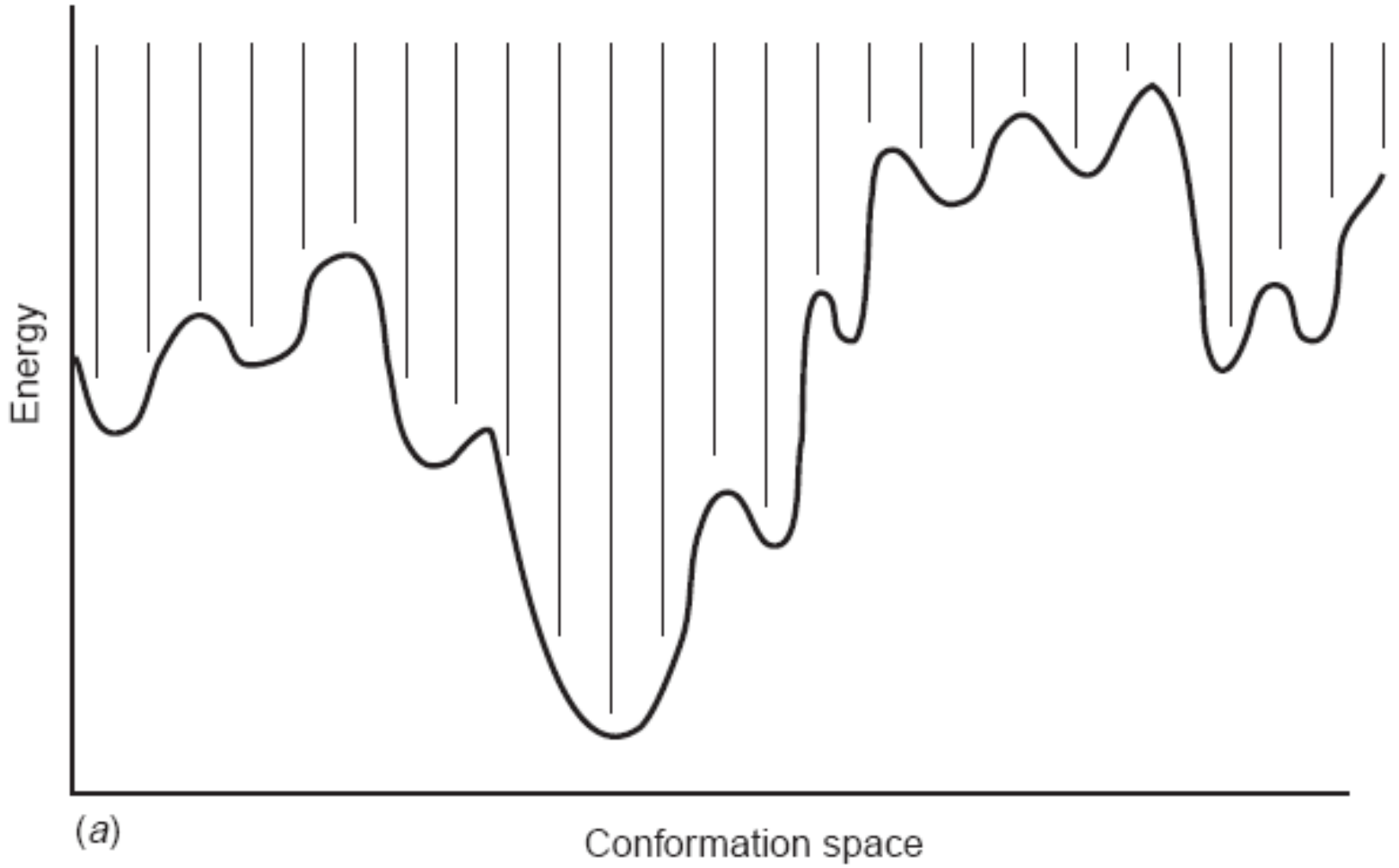
Potential / free energy hypersurface



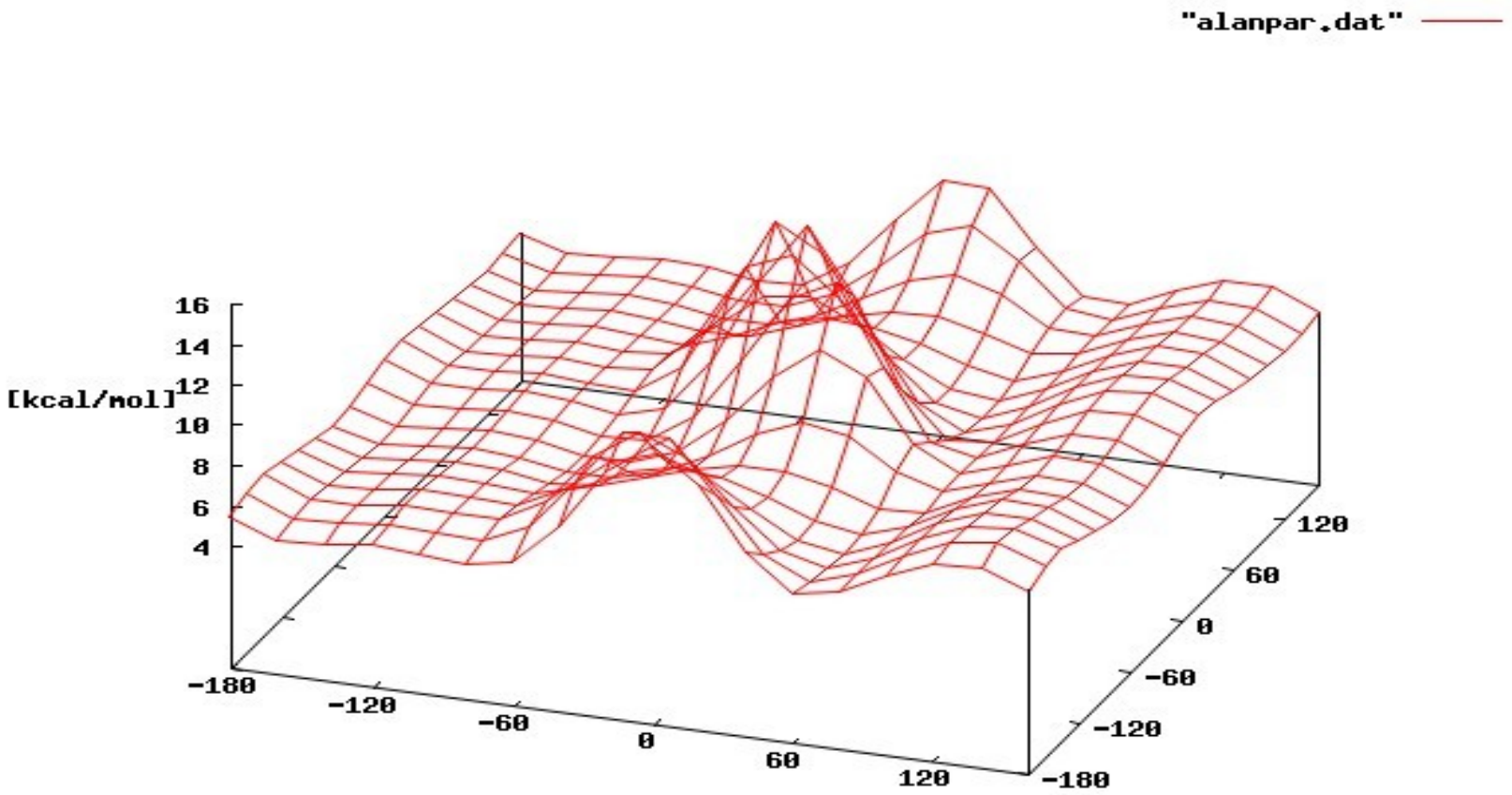
Searching PES

- Grid search
- Metropolis Monte Carlo method
- Simulated annealing
- Distance geometry search
- Homology modelling
- Fragment approach
- Genetic algorithms
- Chain growth algorithms

Grid search method



Grid search method



Grid search method

Problem of combinatorial explosion

$$N_{\mathrm{GP}} = \prod_{i=1}^{k} 360/torsion\ step_i$$

Step size 30 degrees

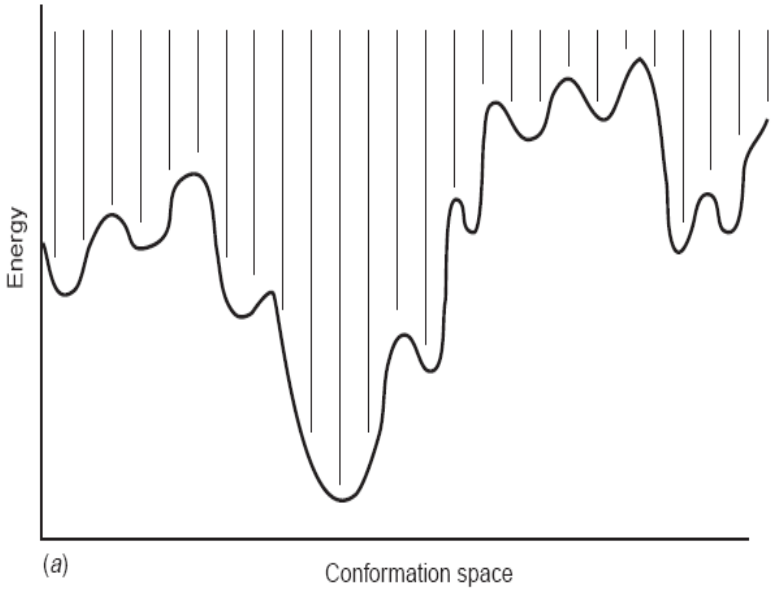
N torsions	3	4	5	6	7
N conformers	1728	20736	248832	2985984	3.6 E7

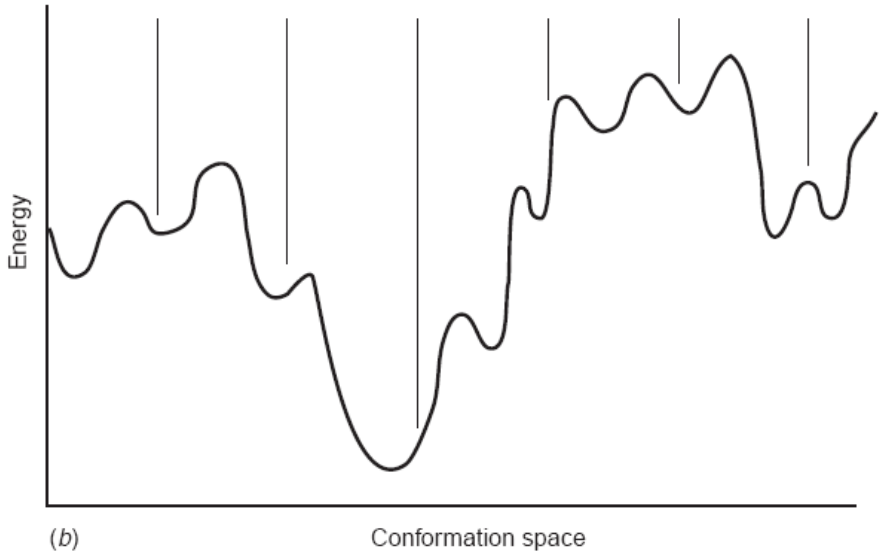
Time: 1 conformer per 1 second:
3 torsions – 29 min, 4 torsions – 6 hours,
5 torsions – 69 hours, 7 torsion – 417 days

Grid search method

- How to solve problem with combinatorial explosion?
  - Energy cutoff
  - Coarse grain grid
  - Fragment base approach
  - Single coordinate driving method

Grid search method





Grid search method

- Advantage:
  - Explore all conformational space systematically.
  - All possible minima can be found.
- Lack:
  - Time consuming combinatorial explosion.
  - Can not be used for large and flexible systems.
  - Limitation for ring systems.

Metropolis Monte Carlo

1 generate state x0

2 find state x1 next to x0

3 calculate E(x1), E(x0)

if (E(x1)<E(x0))

{ x0 = x1 … cycle from **2**}

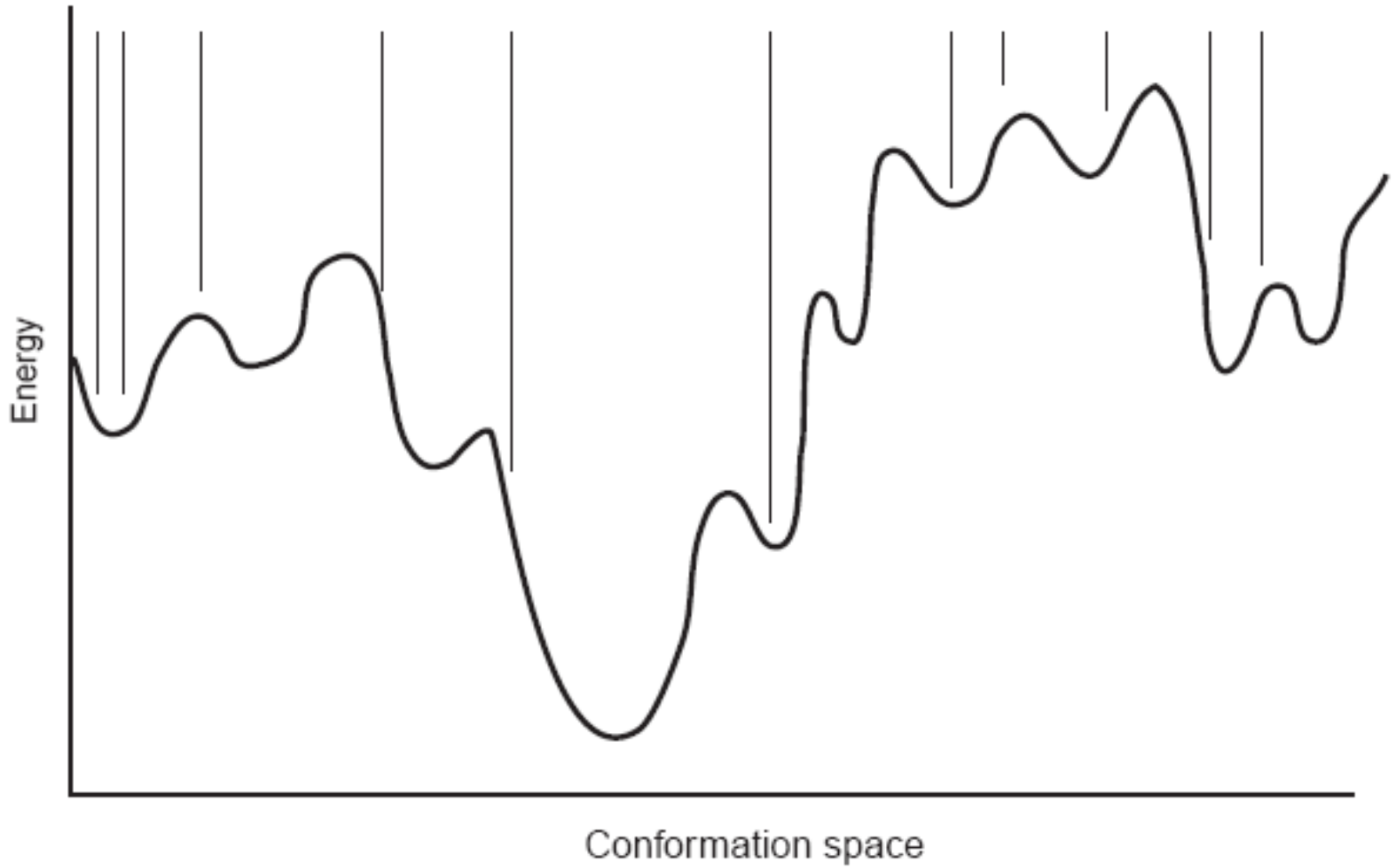
if (E(x1)>E(x0))

{ make x0 = x1 with probability }

$$P(x_0 \rightarrow x_1) = \frac{P_{x_1}}{P_{x_0}} = \mathrm{e}^{\frac{-E(x_1)-E(x_0)}{kT}}$$

4 cycle until x1 does not change

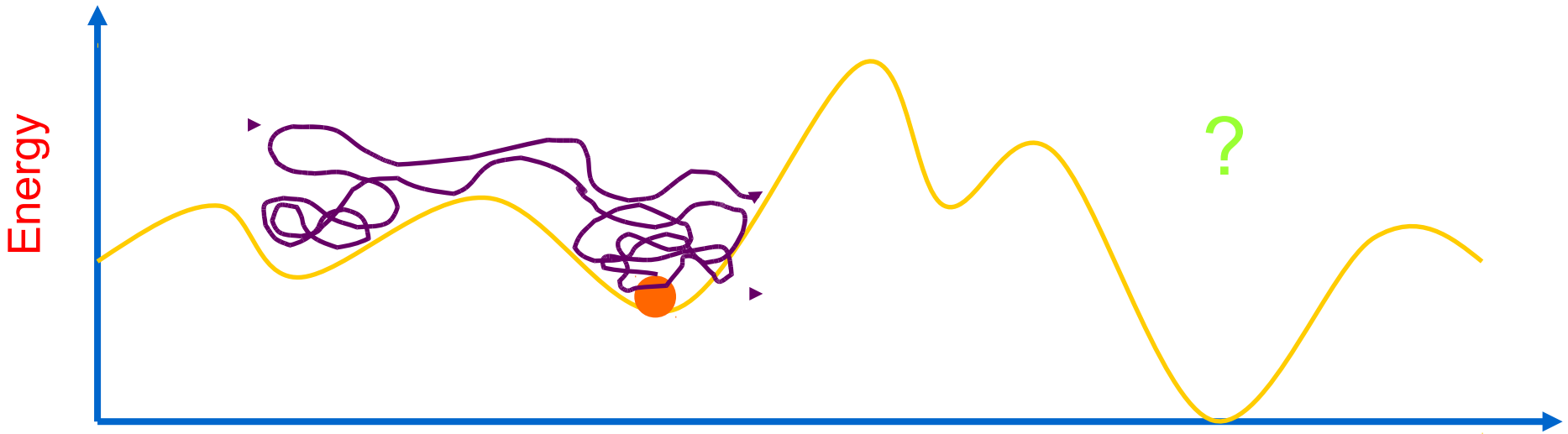
Metropolis Monte Carlo



Metropolis Monte Carlo

- Advantage:
  - Fast and powerful method for large flexible systems.
  - Useful for ring systems.
  - As an additional option chiral centers can be preserved to their original geometry or inverted during conformational search.
- Lack:
  - Only energy criterion for search.
  - No real end point like for systematic search.

Molecular dynamics



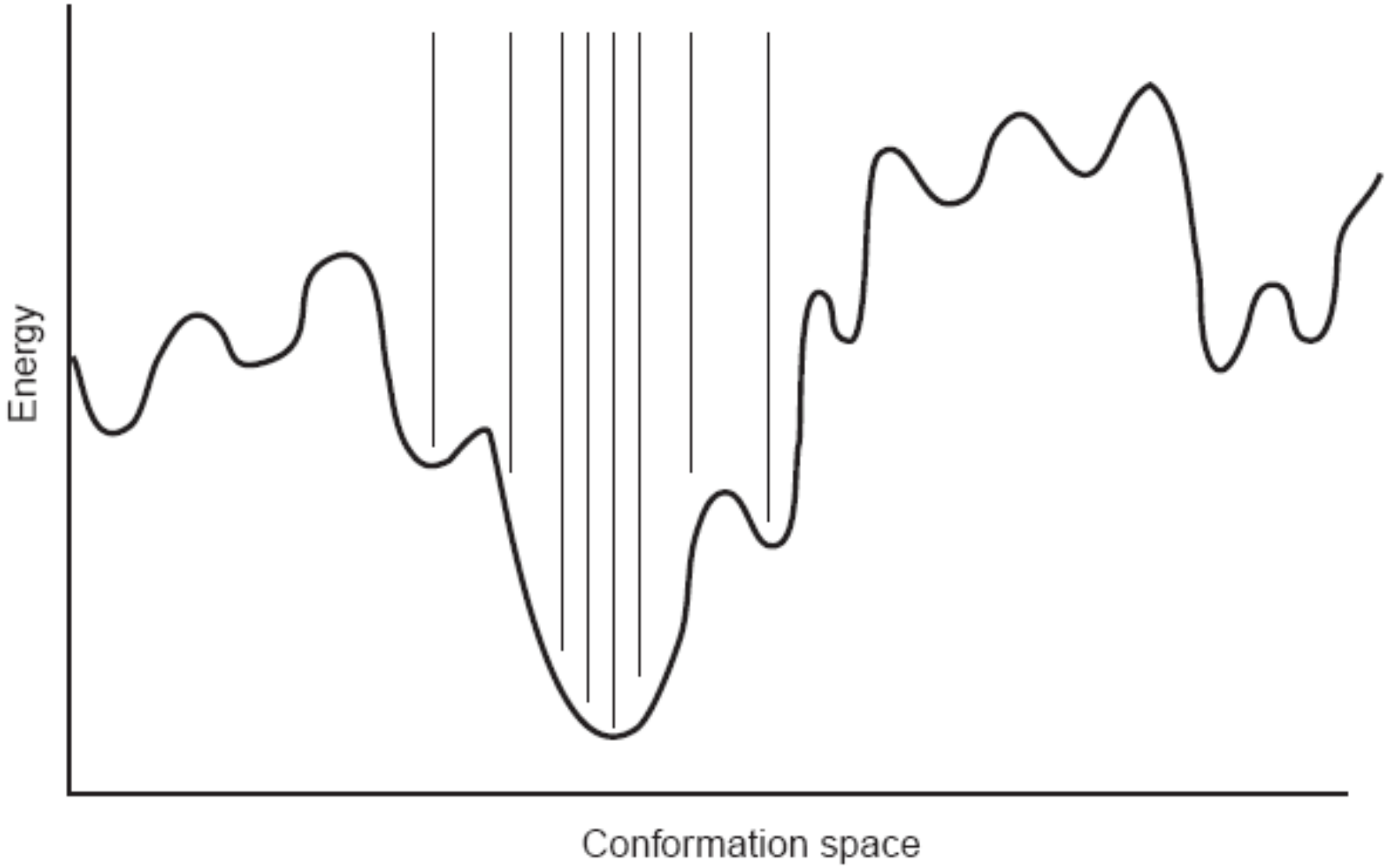
Simulated anealing

- Global optimization technique based on Monte Carlo method.
- Number of accepted conformations is dependent on simulation temperature.
  - High simulation temperature = many states will be accepted.
  - Low simulation temperature = majority of generated states will be rejected.

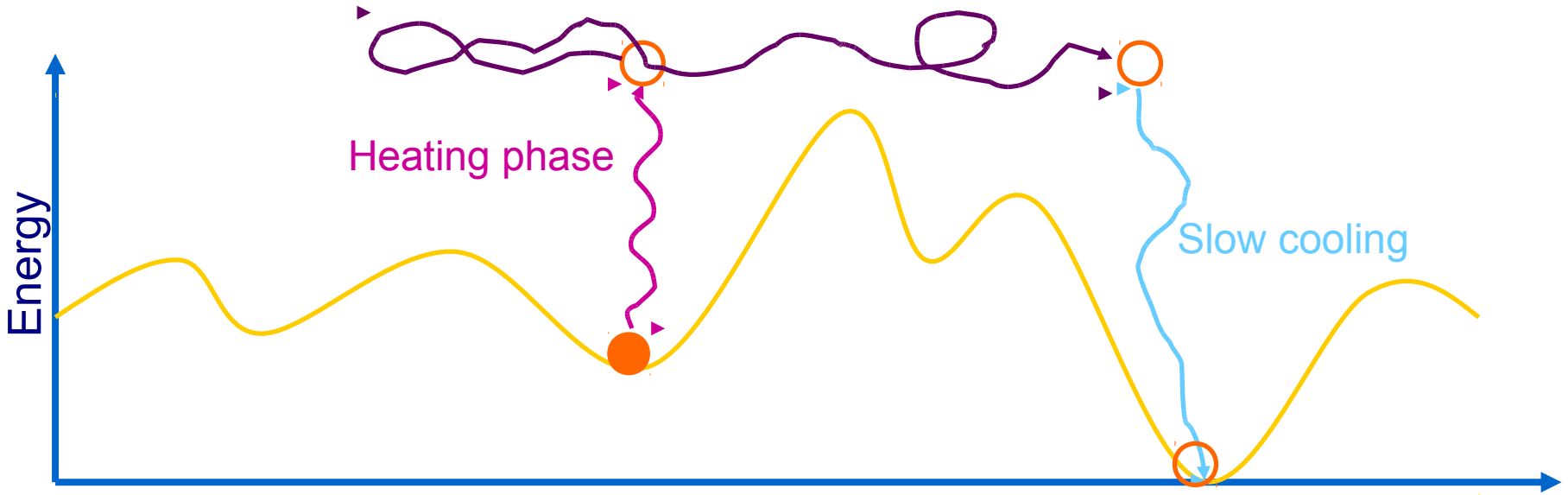
Simulated anealing

```
0 set Tmax
1 generate state x0
2 find state x1 next to x0
3 calculate E(x1), E(x0)
if (E(x1)<E(x0))
{ x0 = x1 … cycle from 2}
if (E(x1)>E(x0))
{ make x0 = x1 with Metropolis criterion }
4 cycle until x1 does not change
5 cool down system and cycle from 1 with x1 state
```

Simulated anealing



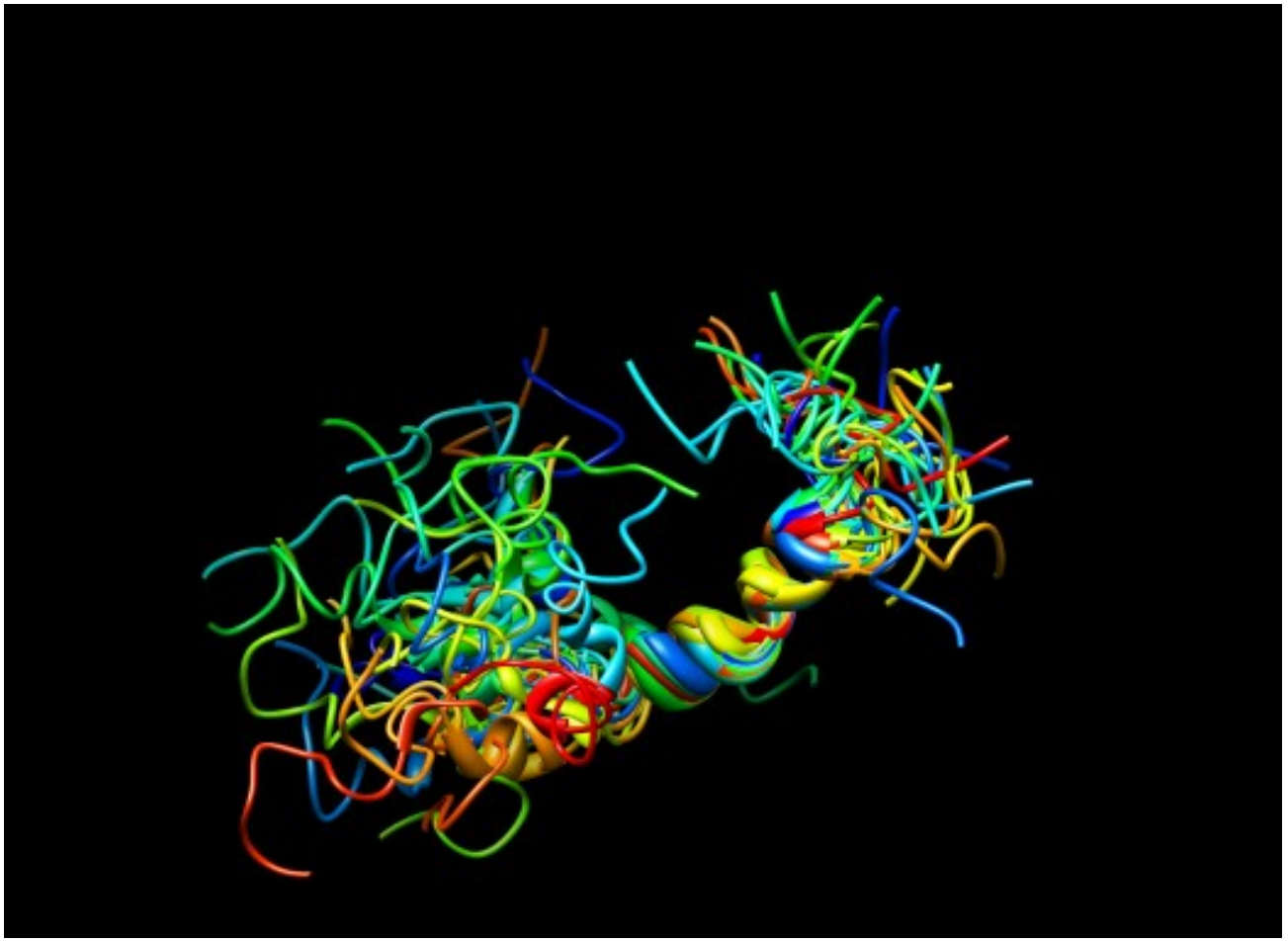
Simulated anealing



Distance geometry

- Reduction of degrees of freedom by experimentally known geometry data
- Crystallography, NMR (NOE)
- Constraints and penalty functions
- Time-averaged constraints

Distance geometry



Distance geometry

- Advantage:
  - Very accurate in the systems for which experimental data are available (NMR, X-ray).
  - Useful for refining structure of proteins and nucleic acids.
  - Can generate several conformations that are consistent with experimental data – additional information about flexibility of the system.
- Lack:
  - Requires experimental data.

Single coordinate driving method

CICADA – **C**hannel **I**n **C**onformational Space **A**nalalysed
by **D**river **A**pproach

CICADA approach

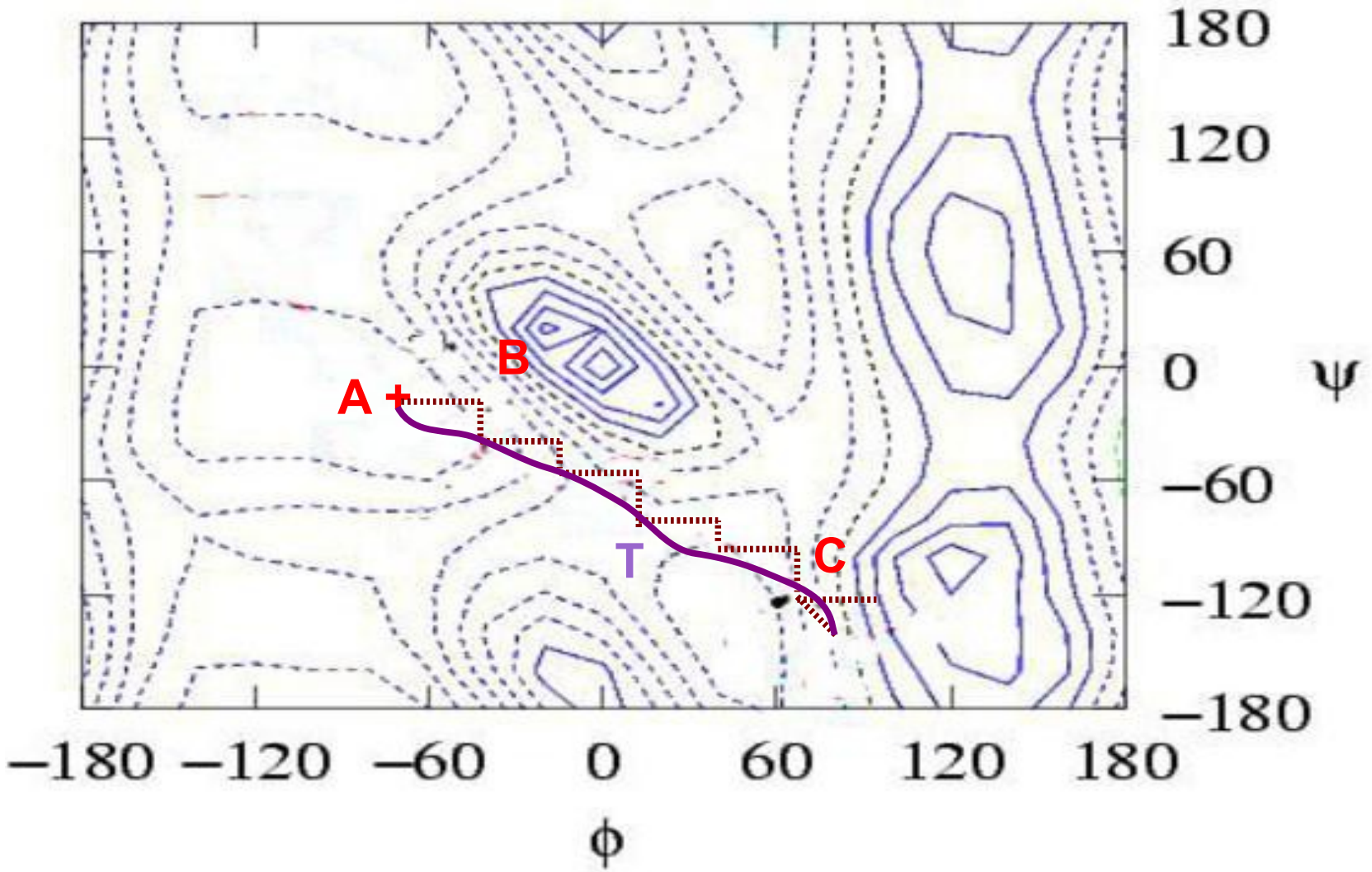
- Systematic search method.
- Solve problem of combinatorial explosion.
- Explore only low energy areas of energy surface.
- Good parallelization = useful for larger systems.
- Used also as docking method.
- Minimization using TINKER software

CICADA approach

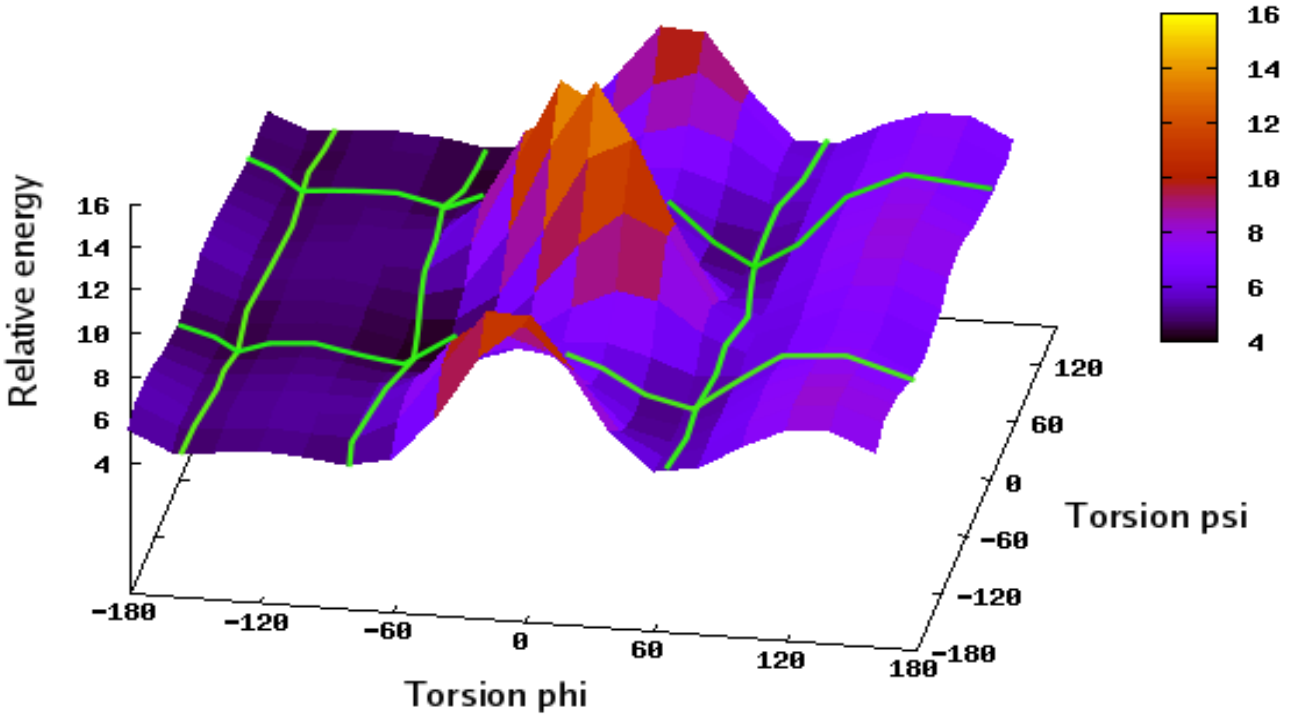
- Why TINKER?

- Free with source code – easy to implemented into CICADA.
- Variety of minimize routines.
- Several common parameter sets:
  - Amber (ff94, ff96, ff98, ff99, ff99SB),
  - CHARMM (19, 22, 22CMAP),
  - Allinger MM (MM2-1991 and MM3-2000),
  - OPLS (OPLS-UA, OPLS-AA),

CICADA approach



CICADA approach



CICADA approach

Control parameters:

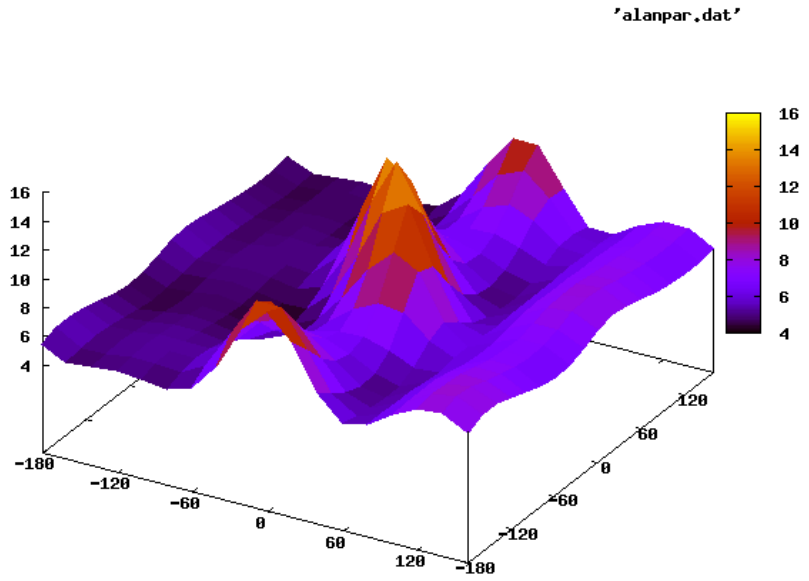
```
DYNAM      0.
ROCRIT     30.
ECRIT      0.20
SEED       pept
PATH       /home/zdenek/PCKI/
DOCKING    F
1./LAST    2   301
BTCHACT    0
DRIVSTEP   20.
STARTTIM   16   37   2   0   09   9   30
BACK       F
MULTCOEF   5
ECUT       150.
CONFCUT    50.00
```

```
NUMBER OF TORSIONS:   36
  1   2   3  18     1   0   0.
  3  18  19  20     1   0   0.
 18  19  20  34     1   0   0.
 20  34  35  36     1   0   0.
 34  35  36  45     1   0   0.
 36  45  46  47     1   0   0.
 45  46  47  69     1   0   0.
 47  69  70  71     1   0   0.
 69  70  71  88     1   0   0.
 71  88  89  90     1   1   0.
```

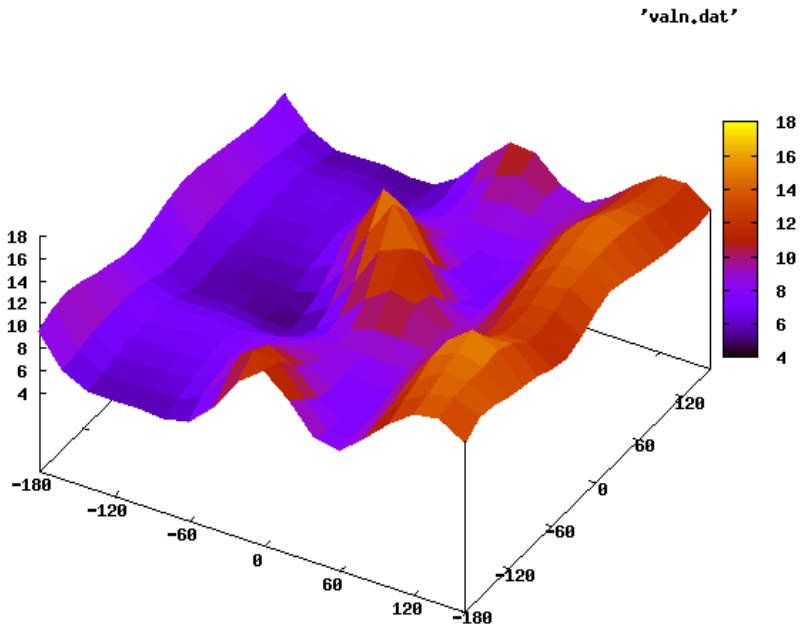
CICADA results

Single amino acids

Alanine



Valine

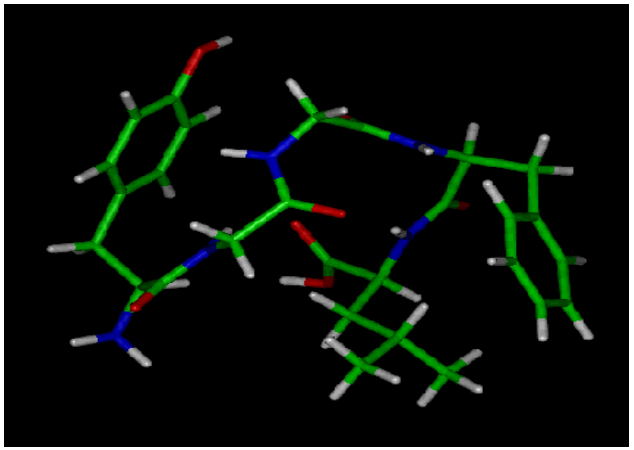


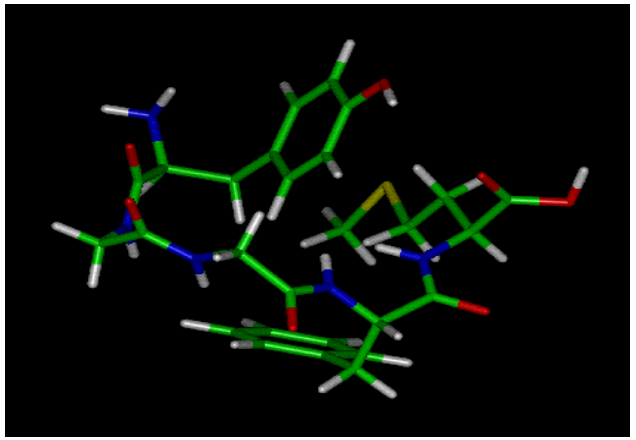
CICADA results

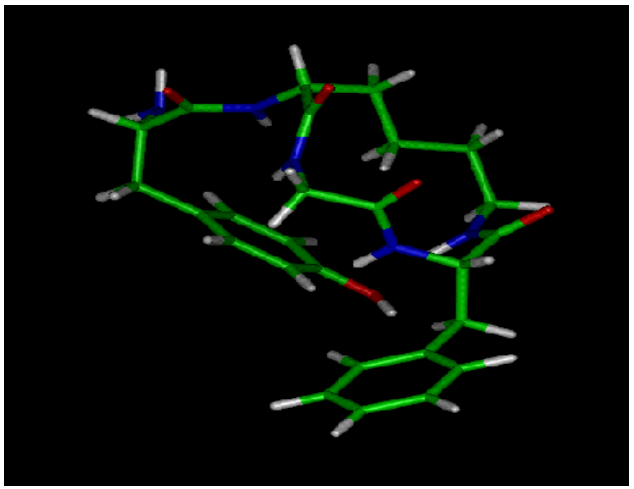
- Cysteine, cystine (L-Cys -S-S-L-Cys and L-Cys-S-S-D-Cys)
- Nucleic acids fragments
- Monosaccharides and oligosaccharides
- Small peptides – enkephalins and their cyclic analogues
- Small organic molecules

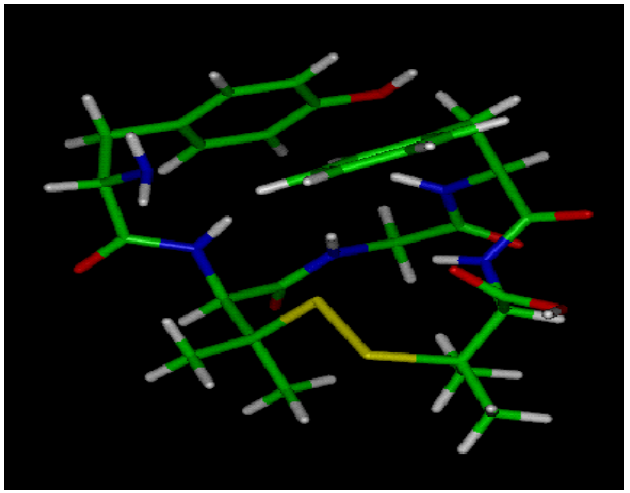
CICADA results

Enkephalins





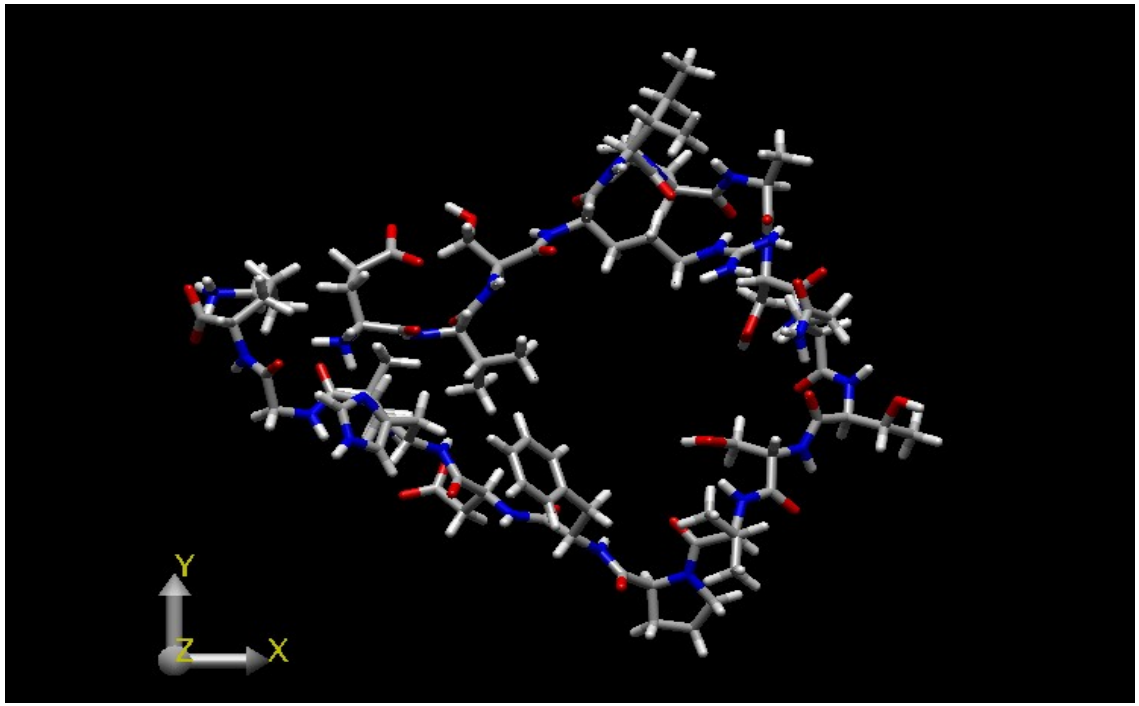




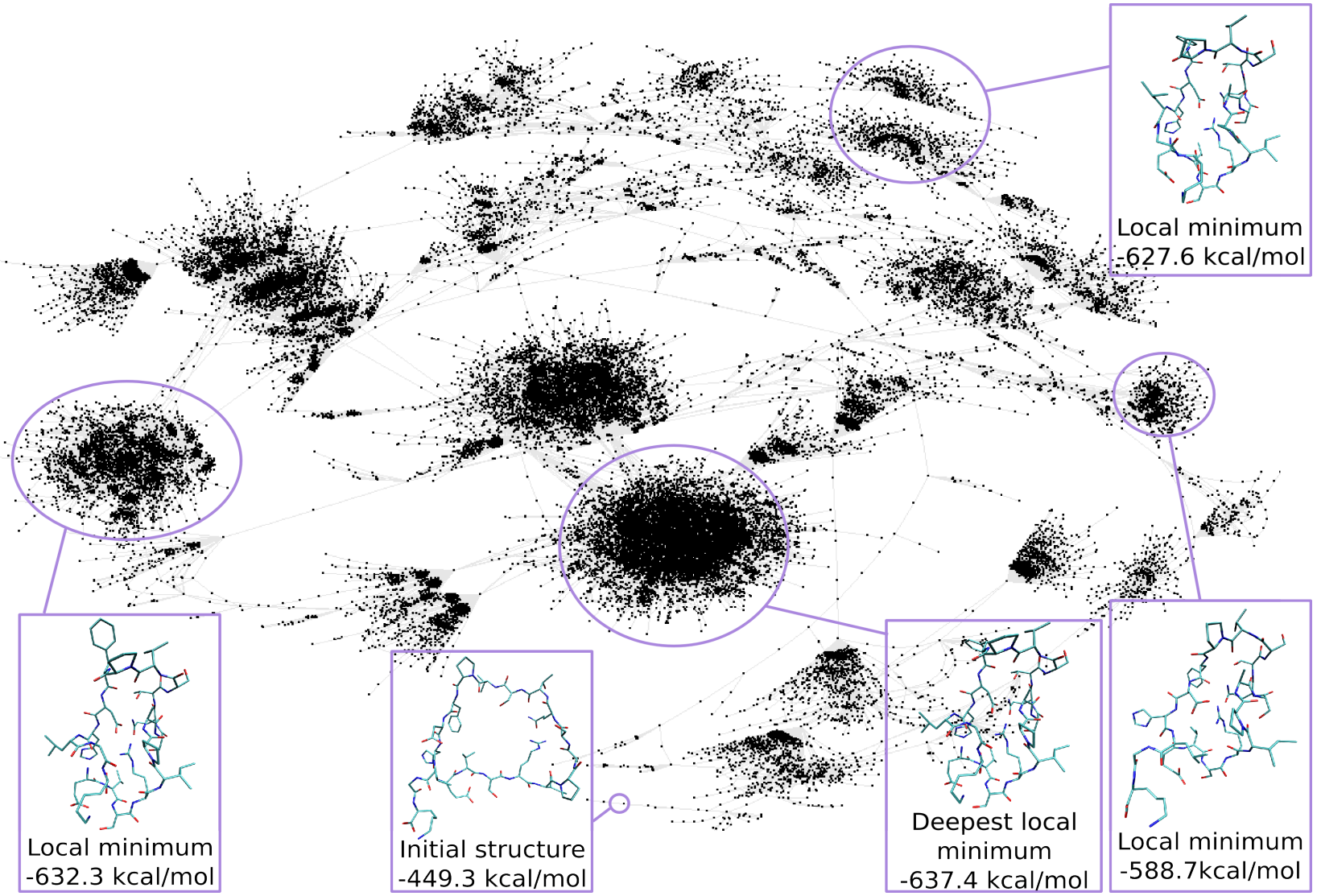
CICADA results

The largest system we have studied using CICADA program

19 amino acids – C terminal domain of Casein Kinase I CKI ε
– natural and polyphosphorylated form

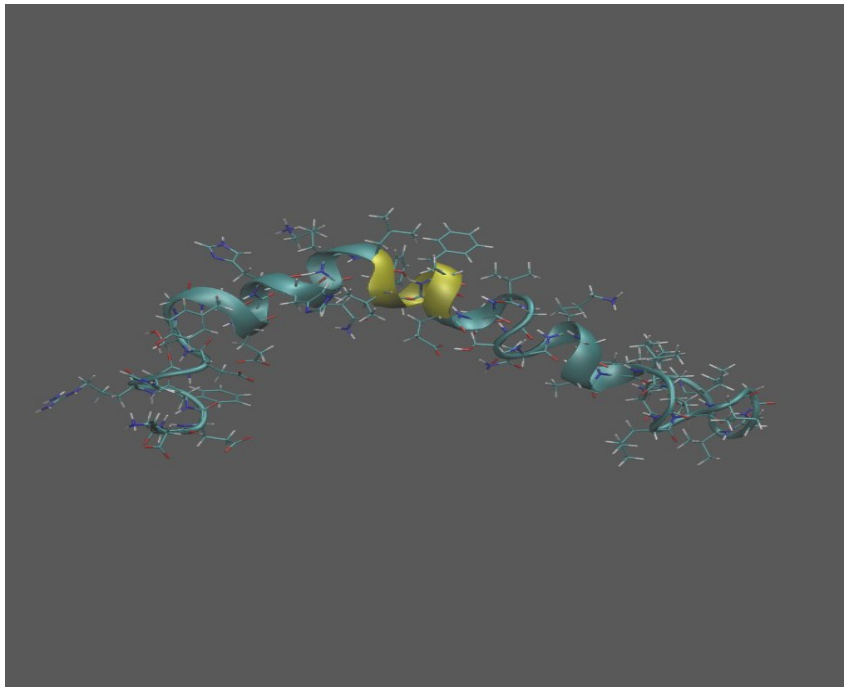


CICADA results

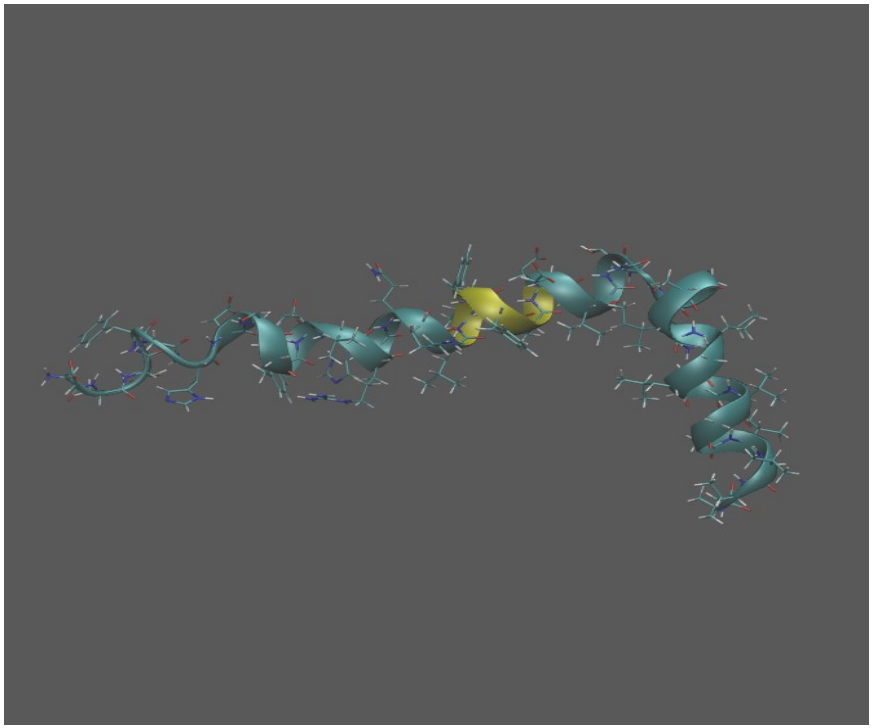


CICADA results

Amyloid β – human and rat form



42 amino acids
Difference in only 3 amino acids



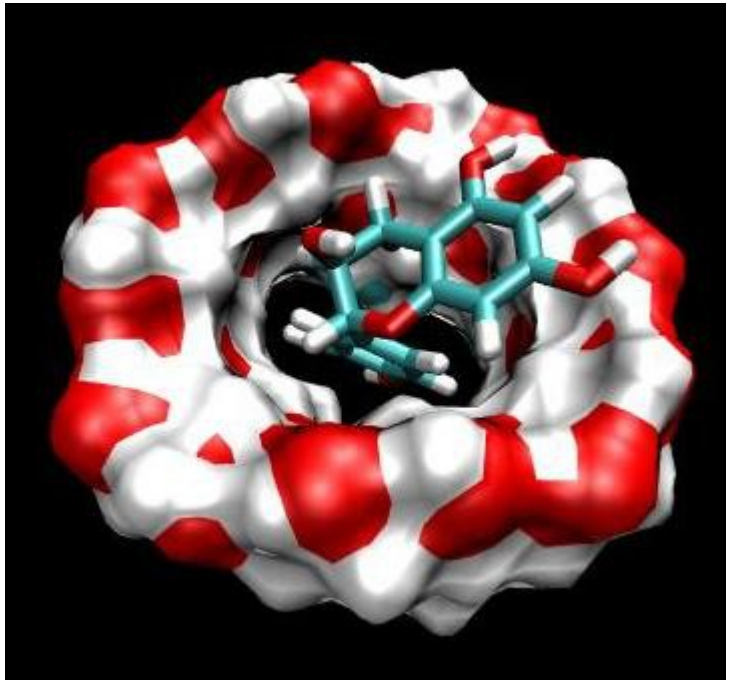
CICADA - docking

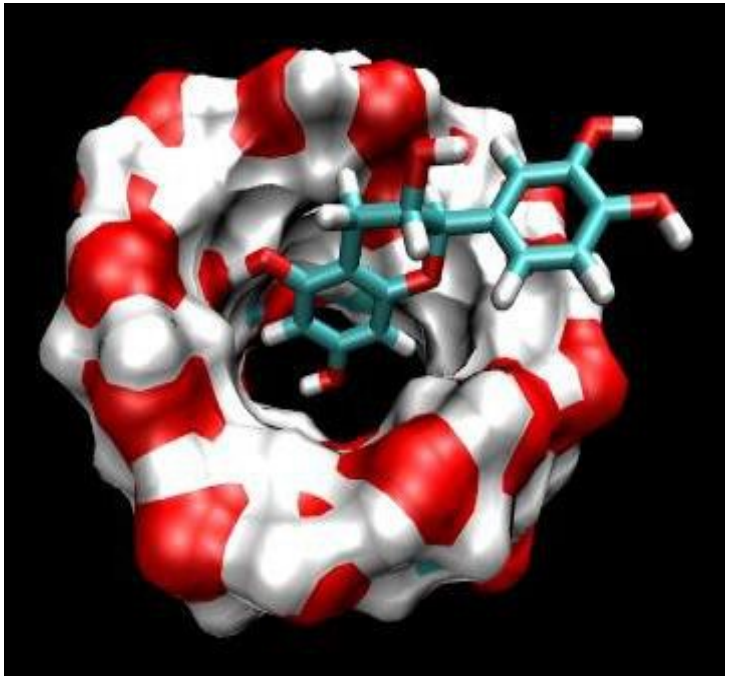
CICADA - docking

- CICADA – docking
  - Normal conformational search of receptor or ligand
  - SCD docking
    - rotations of ligand (3 directions)
    - Translation of ligand (3 directions)

CICADA - docking

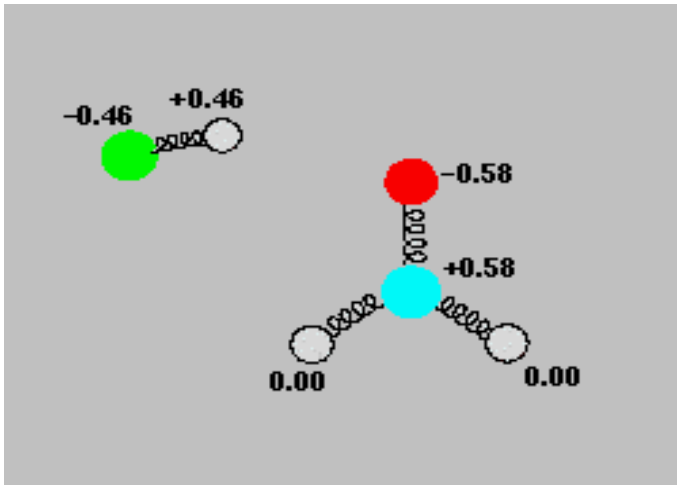
Catechine – β-cyclodextrin docking using CICADA software



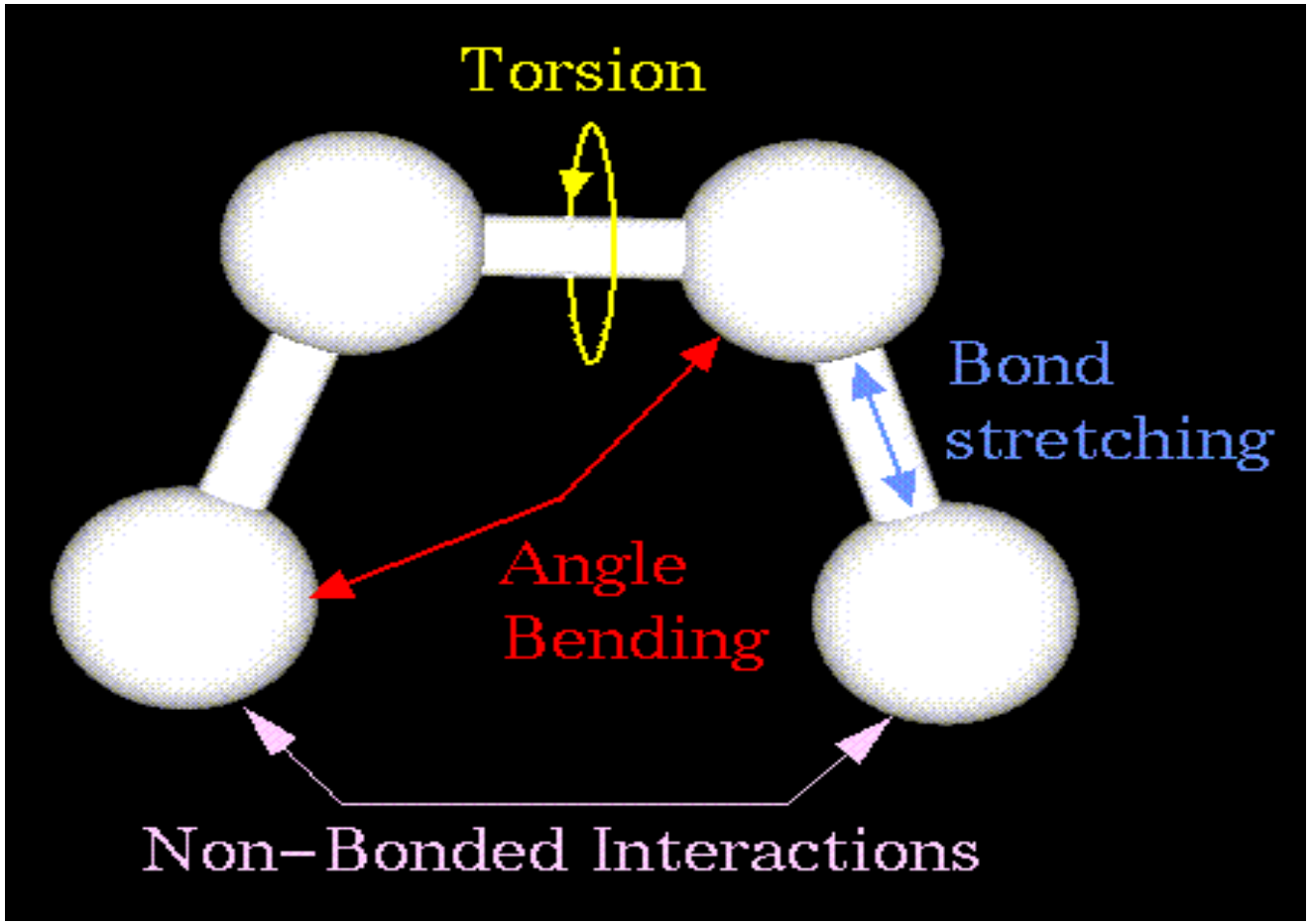


Molecular Mechanics

- Atoms in molecules described using classical mechanics (balls and springs)
- Usable for large systems such a proteins and nucleic acids
- No real energy based on physical background
- Force fields

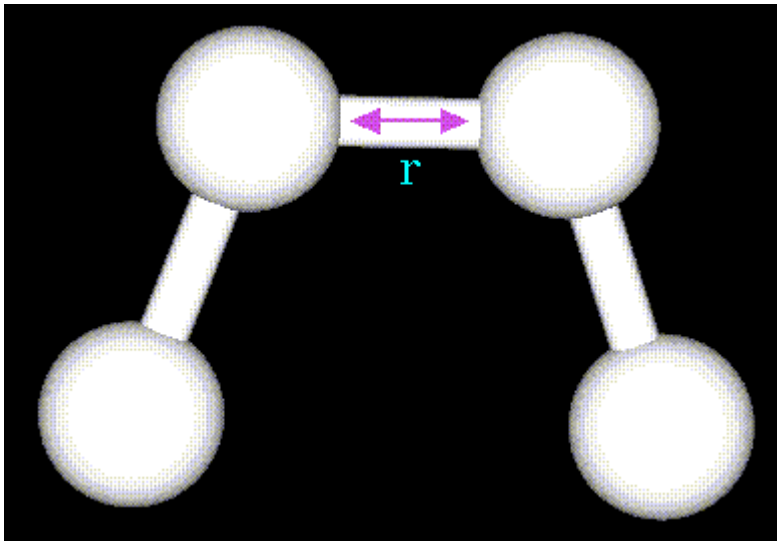


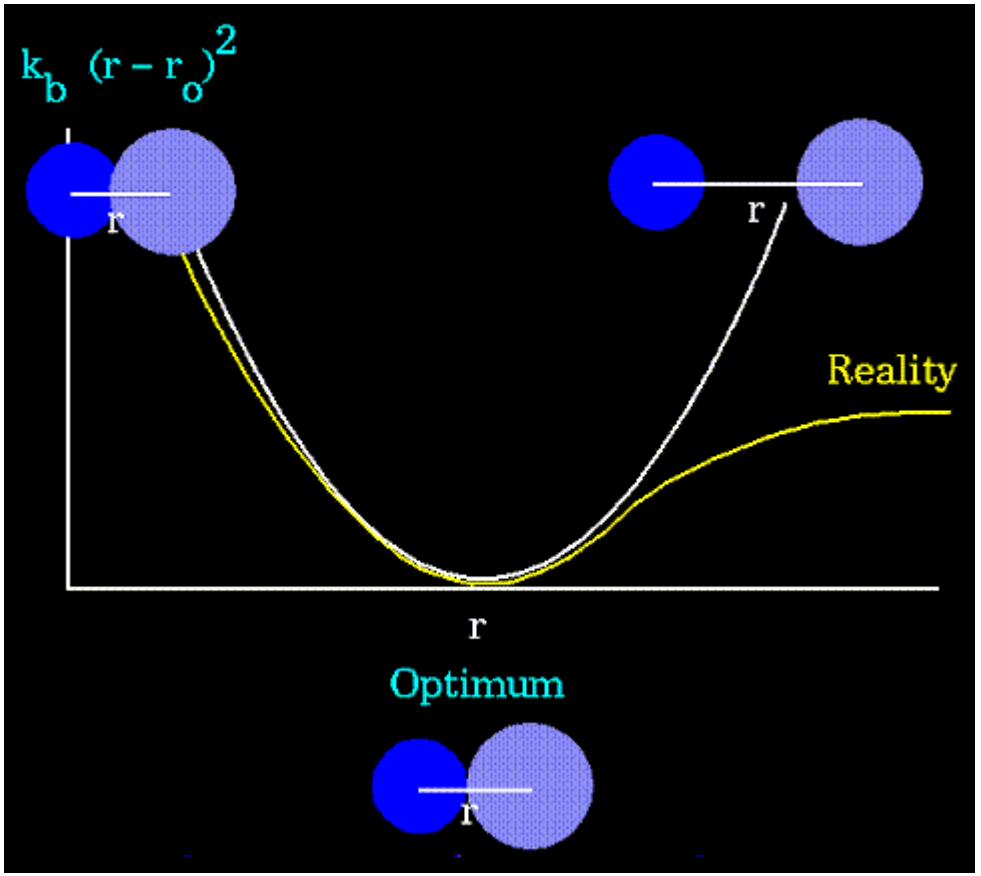
Force fields



Force fields

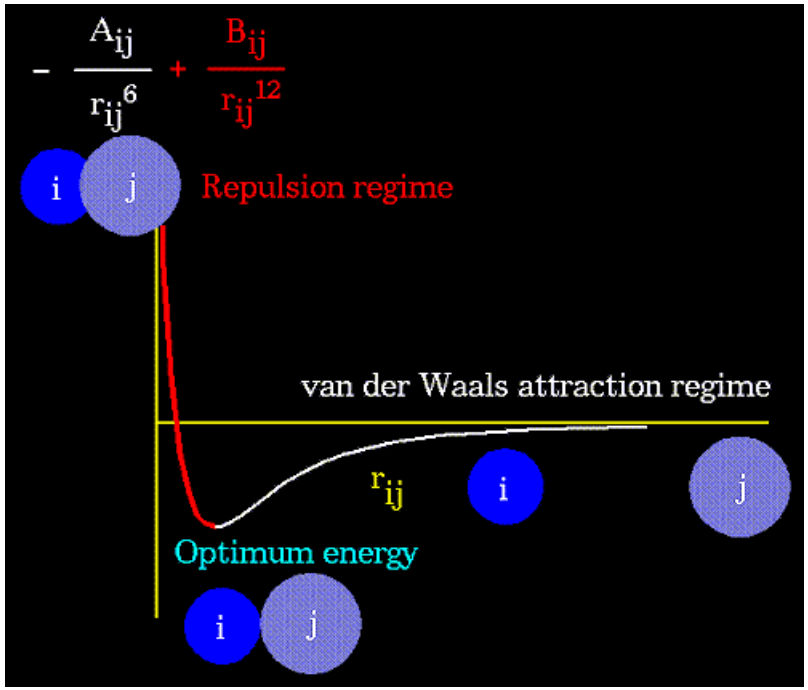
$$E = \sum_{\text{bonds}} k_b \, (r - r_o)^2$$

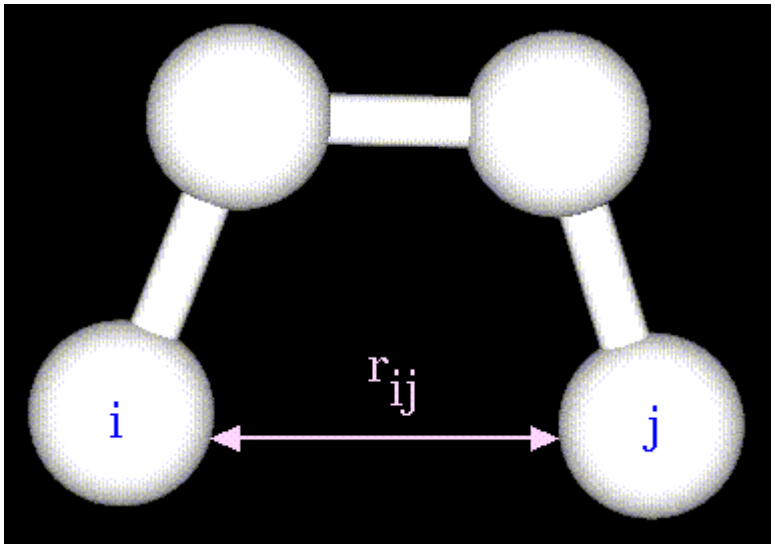




Force fields – non bonded

VdW





Electrostatic

$$+\sum_i\sum_j \frac{q_i\, q_j}{r_{ij}}$$

Electrostatic term

Force fields

```
PARM94 for DNA, RNA and proteins with TIP3P Water. USE SCEE=1.2 in energy progs
C  12.01                              sp2 C carbonyl group
CA 12.01                              sp2 C pure aromatic (benzene)
H  1.008                              H bonded to nitrogen atoms
HC 1.008                              H aliph. bond. to C without electrwd.group
NB 14.01                              sp2 N in 5 memb.ring w/LP (HIS,ADE,GUA)
P  30.97                              phosphate
S  32.06                              sulphur in disulfide linkage
...
OW-HW  553.0     0.9572     ! TIP3P water
C -CA  469.0     1.409        JCC,7,(1986),230; TYR
C -CB  447.0     1.419        JCC,7,(1986),230; GUA
C -N*  424.0     1.383        JCC,7,(1986),230; CYT,URA
...
HW-OW-HW    100.       104.52    TIP3P water
HW-HW-OW      0.       127.74    (found in crystallographic water with 3 bonds)
CB-C -NA    70.0       111.30    NA
CB-C -O     80.0       128.80
...
X -C -CA-X    4   14.50          180.0              2.          intrpol.bsd.on C6H6
X -C -CB-X    4   12.00          180.0              2.          intrpol.bsd.on C6H6
CT-CT-OS-CT   1    0.383           0.0             -3.
N -CT-C -N    1    0.40          180.0             -4.
...
  H           0.6000  0.0157             !Ferguson base pair geom.
  HW          0.0000  0.0000              TIP3P water model
  O           1.6612  0.2100              OPLS
  S           2.0000  0.2500              W. Cornell CH3SH and CH3SCH3 FEP's
  IP          1.8680  0.00277             Na+ Aqvist JPC 1990,94,8021. (adapted)
```

Force field - equations

$$E_{\text{total}} = \sum_{\text{bonds}} K_r(r - r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_{eq})^2$$

$$+ \sum_{\text{dihedrals}} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}\right]$$

Force fields -types

- MM2, MM3 – N. A. Allinger
- AMBER – P. Kollmann
- OPLS – Jorgensen
- MMFF
- CHARMM
- Amoeba

Optimization procedures

- Simplex method
- Steepest descent – first derivatives
- Conjugated gradient – first derivatives
- Newton – Raphson method – second derivatives

Acknowledgment

- Grand agency of the Czech Republic
- Ministry of education, youth and sports

- Supercomputer center Brno

- Prof. Jaroslav Koča
- Dr. Petr Kulhanek
- Jakub Štěpán