# HP-SEE

www.hp-see.eu

# Virtual Research Community Computational Chemistry

HP-SEE

High-Performance Computing Infrastructure for South East Europe's Research Communities

**Branko J. Drakulić, Department of Chemistry**
**IChTM, University of Belgrade**

The HP-SEE initiative is co-funded by the European Commission under the FP7 Research Infrastructures contract no. 261499

# Overview

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

## Computational Chemistry Applications

*CFD Analysis of Combustion* – OpenFOAM – test – UoBL -BiH

*CompChem* – NAMD, Desmond, Gaussian, Autodock, Cheminformatics applications – production – UoBg - Serbia

*FMD-PA* – Non-linear optics materials – Gaussian, NWChem - production - Computational Chemistry Group of NHRF - Greece

*HC-MD-QM-CS* – MM/QM simulation in condensed phase – - in house parallel code - production - UKIM - FYROM

# Overview

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

## Computational Chemistry Applications

*ISyMAB* – MD simulation of biomacromolecules by parallel codes - NAMD, in house pre/post processing – beta - FIN-HH, Department of Elementary Particles and Informational Technologies - Romania

*MDCisplatin* – Design of platinum group metal complexes – Gaussian, GAMESS (Firefly) – beta - Institute of Molecular Biology, BAS - Bulgaria

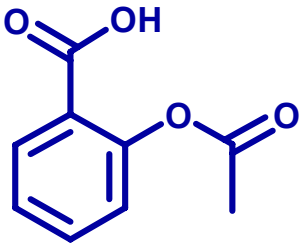
*PACIC* - PCA of the conformational interconversions in ring systems – AMBER, in house pre/post processing – production – IOCCP, BAS - Bulgaria

- Along with this, some of life science applications

# In common....

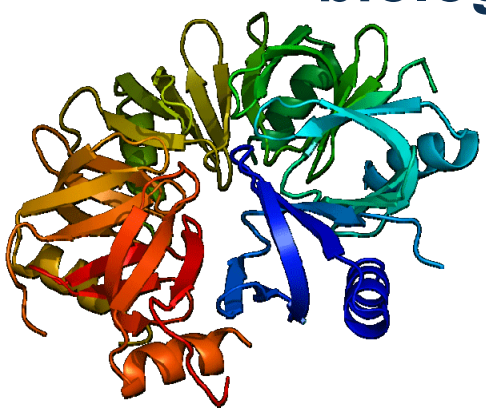
HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

**OK, for the majority of applications**

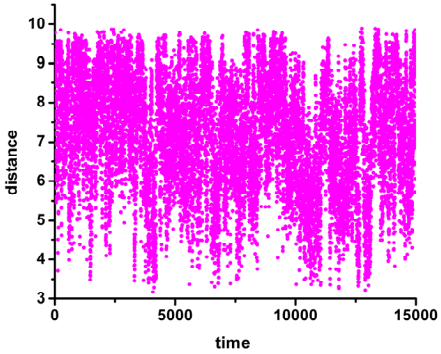



**- small**

**- biomacromolecules**



**- biologically active molecules**




**- their dynamical behavior**

**- analysis of such results could offer insight in properties that influence their biological activity, or the mode of action within living system (cell, tissue, whole animal)**

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- So far:

  - molecular dynamics code NAMD (versions 2.6 to 2.8) http://www.ks.uiuc.edu/Research/namd/

  - *ab initio*/DFT code Gaussian (version 03) http://www.gaussian.com/

  - DFT-based pseudopotentials (Car-Parrinello molecular dynamics) CPMD http://www.cpmd.org/

  - AutoDock Vina (version 1.0) http://vina.scripps.edu/

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- In progress:

  - molecular dynamics code Desmond (version 3.0) http://www.deshawresearch.com/

  - cheminformatics applications OMEGA, ROCS, EON, QUACPAC http://www.eyesopen.com/

- Interest exist for material modeling software, for example DL_Poly http://www.cse.scitech.ac.uk/ccg/software/DL_POLY/

# CompChem (RS)

HP-SEE

- (Why) NAMD

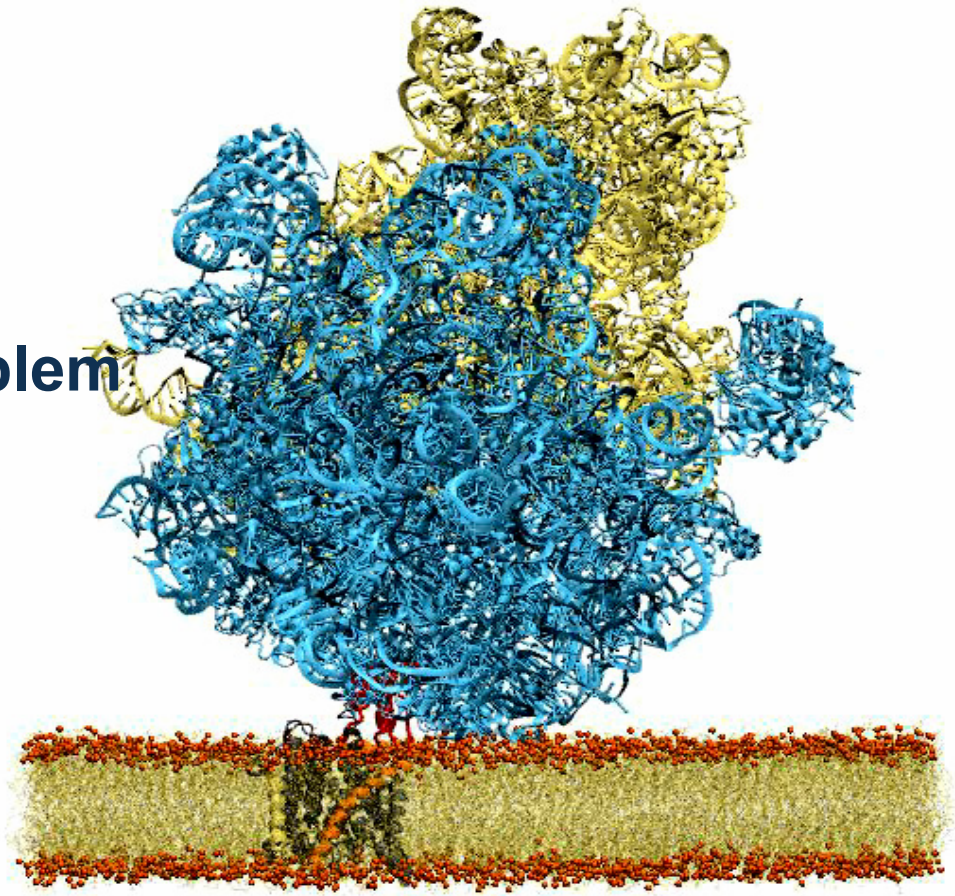
- size of the system is not a problem

- time of simulation is a big problem

- some acceleration procedures exist

- still, we are far from real time time scales

*PNAS*, 2011, 108, 3596-3601



# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

**In USA**

**National Center for Supercomputing Applications (NCSA)**

* NCSA Intel 64 Linux Cluster (Abe)
* TeraGrid Cluster (Mercury)
* NCSA SGI Altix (Cobalt)
* NCSA Xeon Cluster (Tungsten)
* NCSA IBM p690 (Regatta)

**Pittsburgh Supercomputer Center (PSC) - including Cray XT3**

**Texas Advanced Computing Center (TACC)**

**San Diego Supercomputer Center (SDSC)**

- ***J. Comp. Chem*. 2005, 26 (16) 1781–1802, > 1000 citation, recorded on March 2010**

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- Gaussian
  - semiempirical, DFT, and *ab initio* calculations
  - including 'static' behavior of compound under study
  - or reaction pats – mechanism of chemical reaction
  - allow MM/QM calculations
  - in such studies part of the system where, for example, covalent chemistry are involved is treated on QM level, while the rest of the system are treated on the lower level of theory.

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- AutoDock
  - allow relatively fast docking of compounds library to chosen target
  - proved as accurate
  - run in parallel mode
  - high reproducibility of results
  - results obtained can be re-scored by other tools

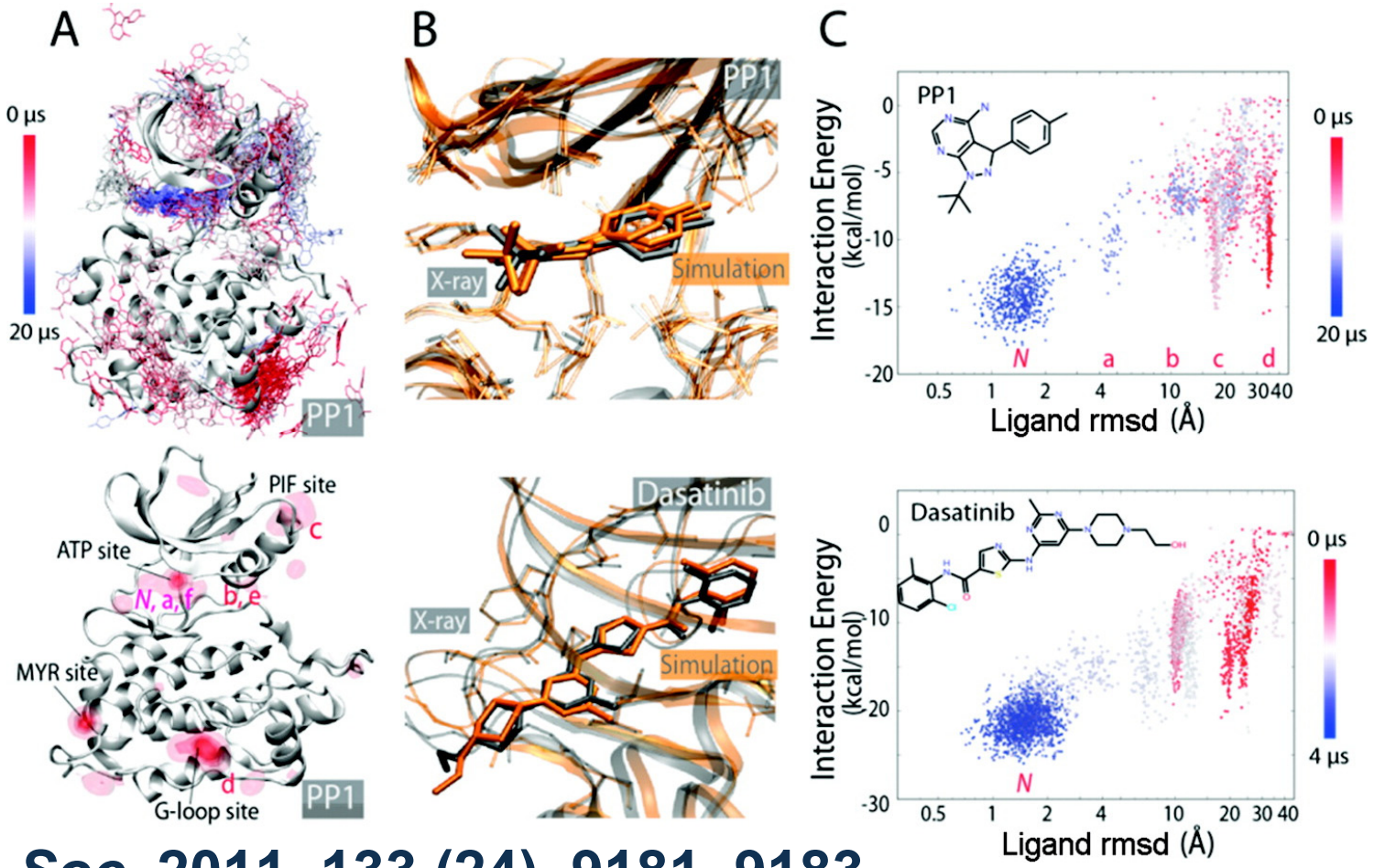
*J. Comp. Chem*. 31 (2010) 455-461

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- Desmond

- milisecond MD




*J. Am. Chem. Soc.* 2011, 133 (24), 9181–9183

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- Cheminformatics applications
  - handle large number of molecules
  - handle large number of conformers
  - most often - both
- can use multinode/multicore CPUs', as MPI or PVM
- proved as accurate
- input can be analyzed on user terminal in a fast manner

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- by application

  OMEGA – conformer generation

  ROCS – shape fitting

  EON – electrostatic fitting

  QUACPACK – charge assignment

- pre/post processing of the data needed will be discussed in the next section

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- main goal

- to build *software infrastructure* that enable efficient usage of molecular modelling applications, cheminformatics applications, and applications of the 'first principle' MO tools to researchers in virtual community.

- not so trivial, as look like on the first sight

- along with financing, which is bottleneck for the majority of scientific communities, we are faced with different scope of interest among different research groups, poor organization and *many subjective circumstances* 'in heritage' to south-eastern European region…

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

*Part two*

- Who is a target group, or who have interest to use applications in CompChem project ?

- Advantages ?

- About theory in background

- About pre/post processing

- Many oversimplification due to time limit and audience

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- **Who is a target group, or who have interest to use applications in CompChem project ?**
  - **Chemists at the first place, but of 'what kind' ?**
  - **Organic, biochemists, medicinal, chemical biologists**
  - **Physical, physical organic**
- **Advantages ?**
  - **Same applications could be used in your lab's PC, still…**
  - **Run significantly slower**
  - **Significantly less data can be processed**

# CompChem (RS)

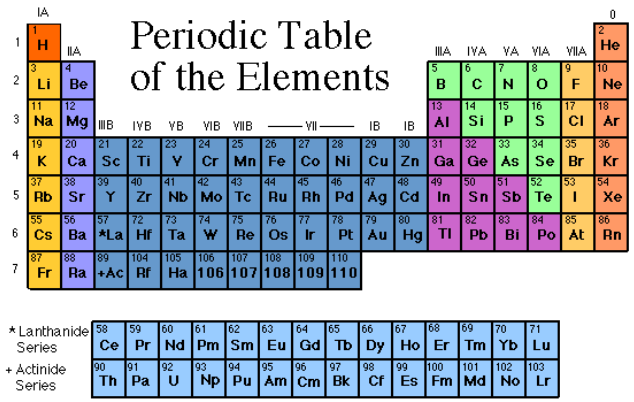
HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- About theory in background

  - Chemistry is the science of matter, especially its properties, structure, composition, behavior, reactions, interactions and the changes it undergoes. (Merriam-Webster's Medical Dictionary)

- To make things simpler

  - There are atoms




  … and molecules composed of atoms bounded to each other

- Artificial, but useful classification

  - organic, main elements C, H, O, N, S, P

  - inorganic

# CompChem (RS)

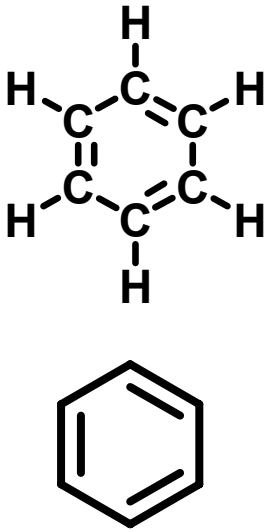

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

**50th millionth small molecule recorded 2008, September**

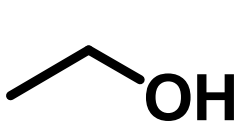
# CompChem (RS)

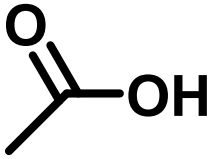
HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

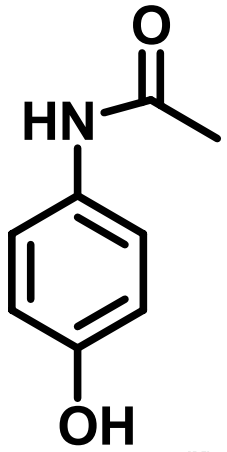
**Molecular representation** **1D, 2D, 3D**

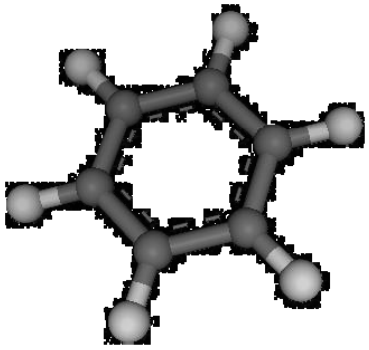


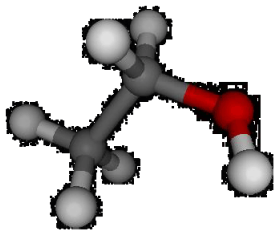

$CH_3CH_2OH$ $CH_3COOH$ $HOC_6H_4NHC(O)CH_3$

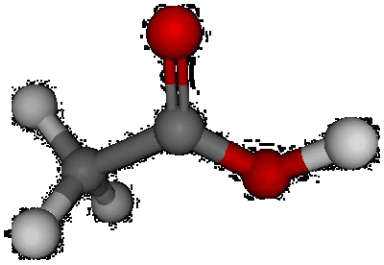


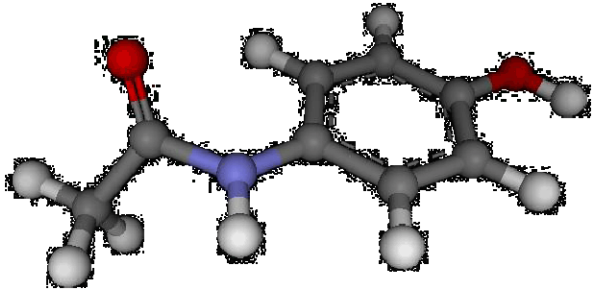





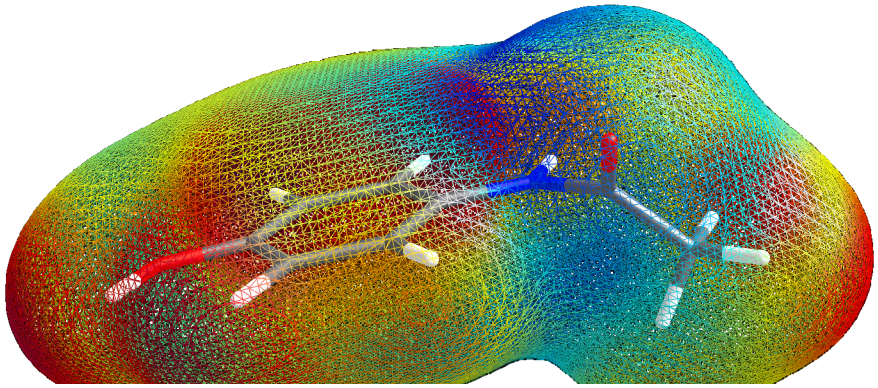


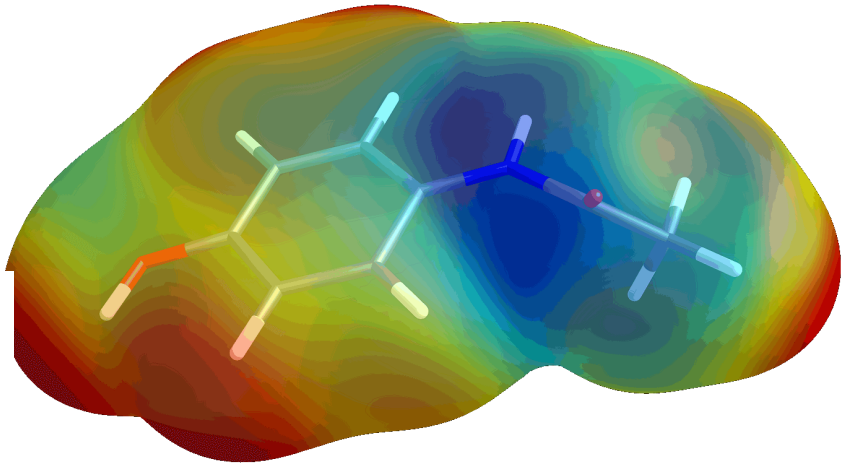


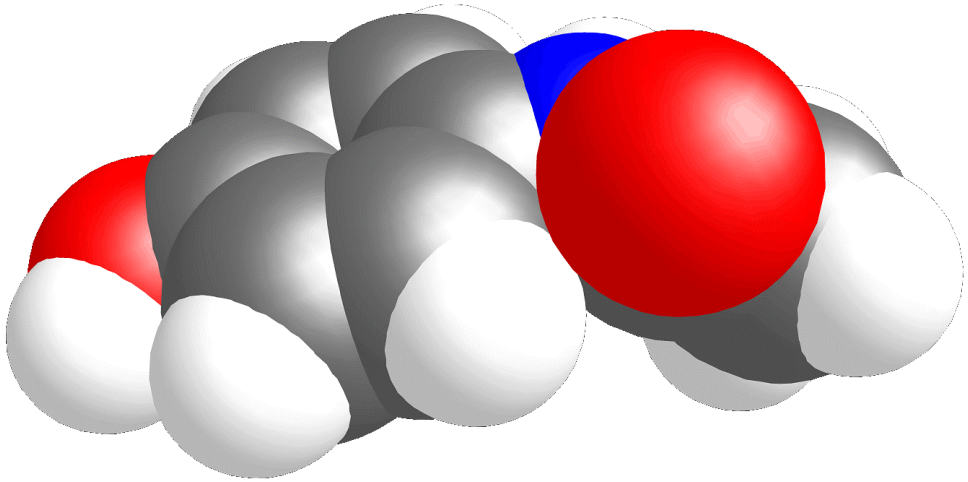
# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

## Molecular volume and/or surface



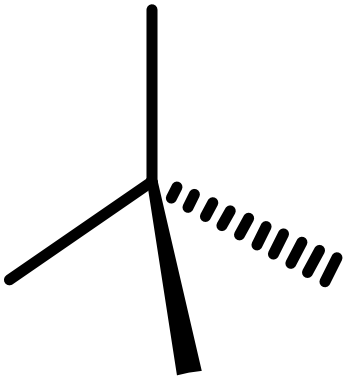




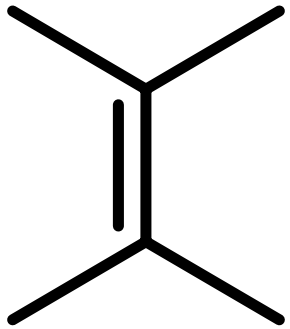
# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

## Geometries

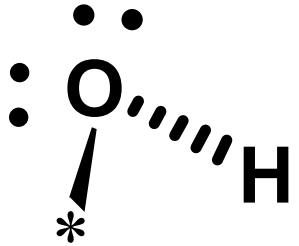


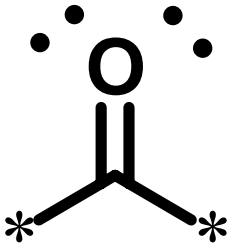
$sp^3$

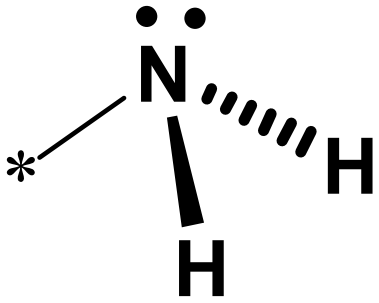


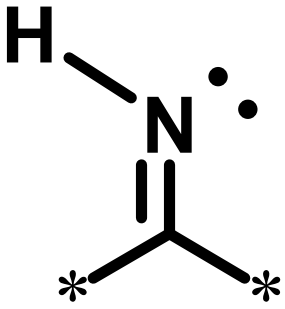
$sp^2$

$sp$



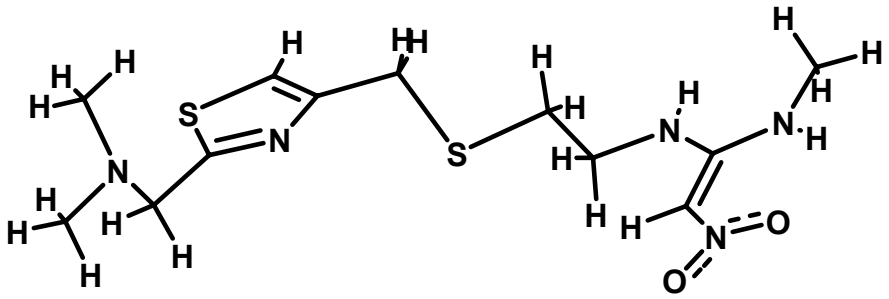



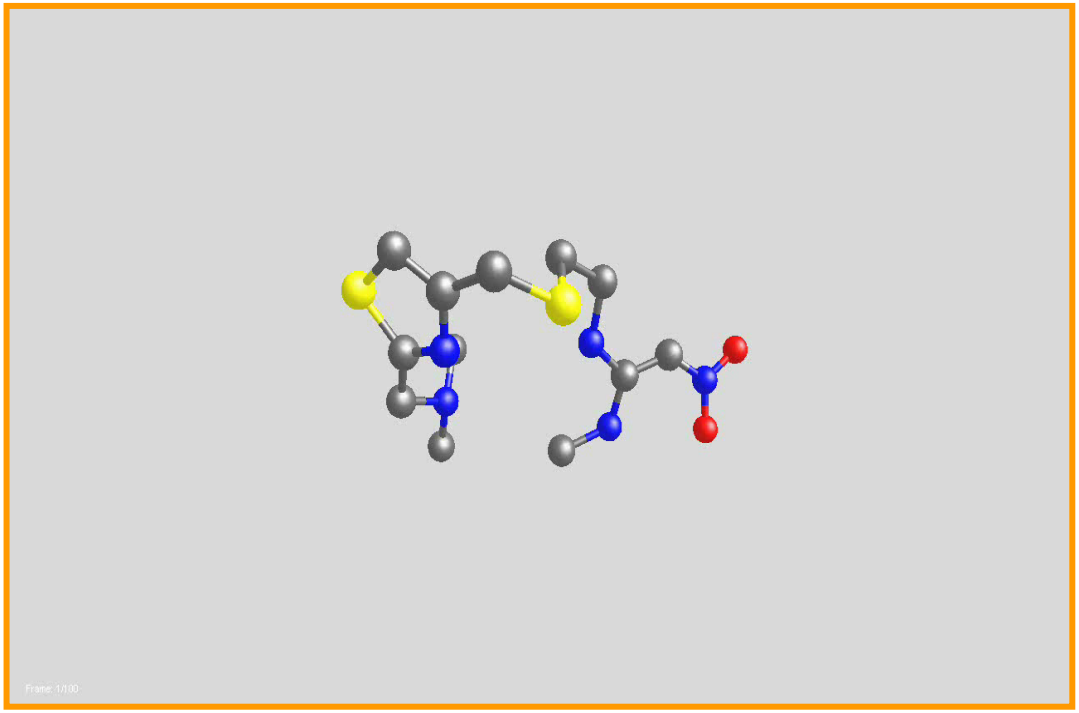



# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

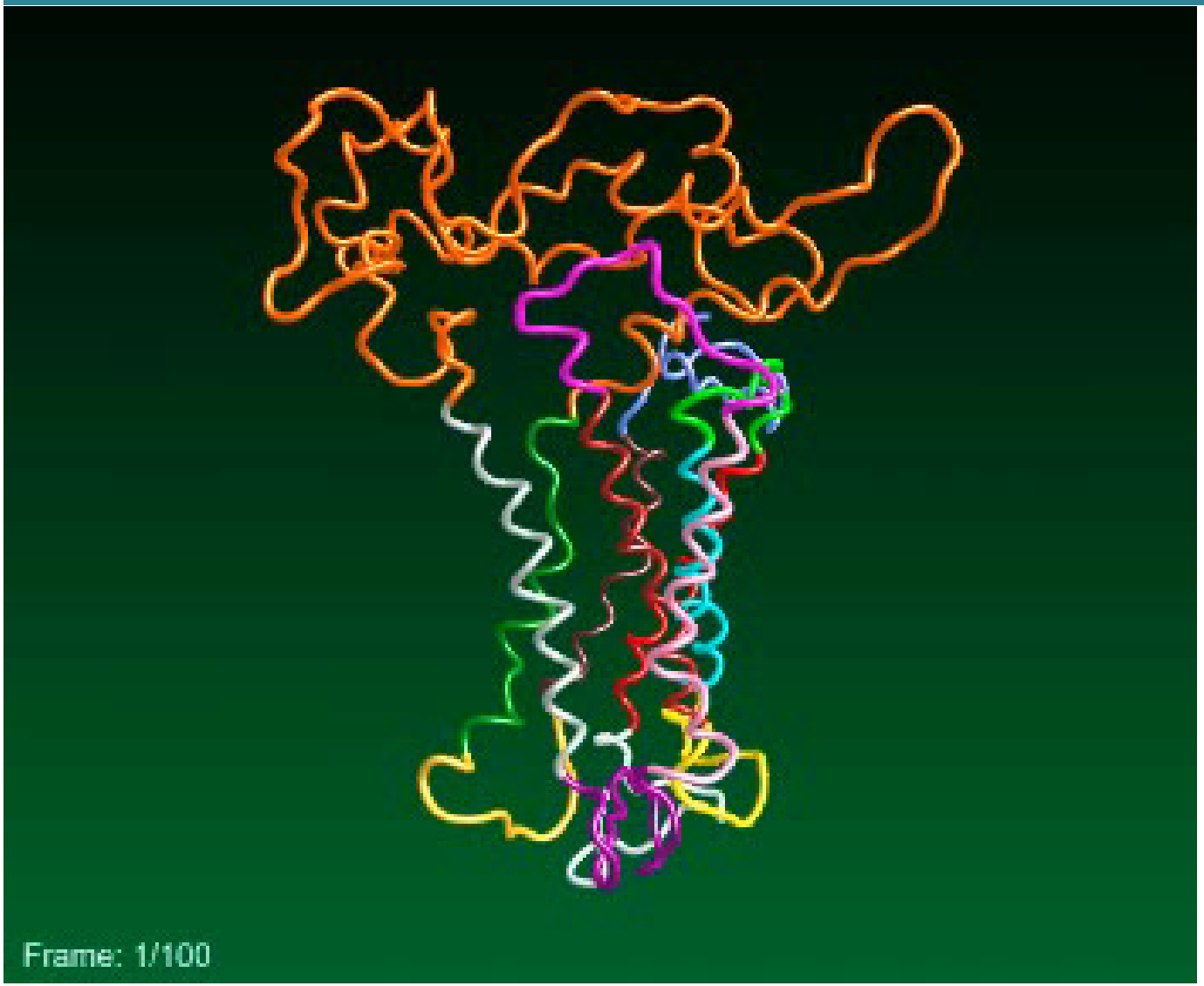
## Molecular flexibility





the ‘small molecules’

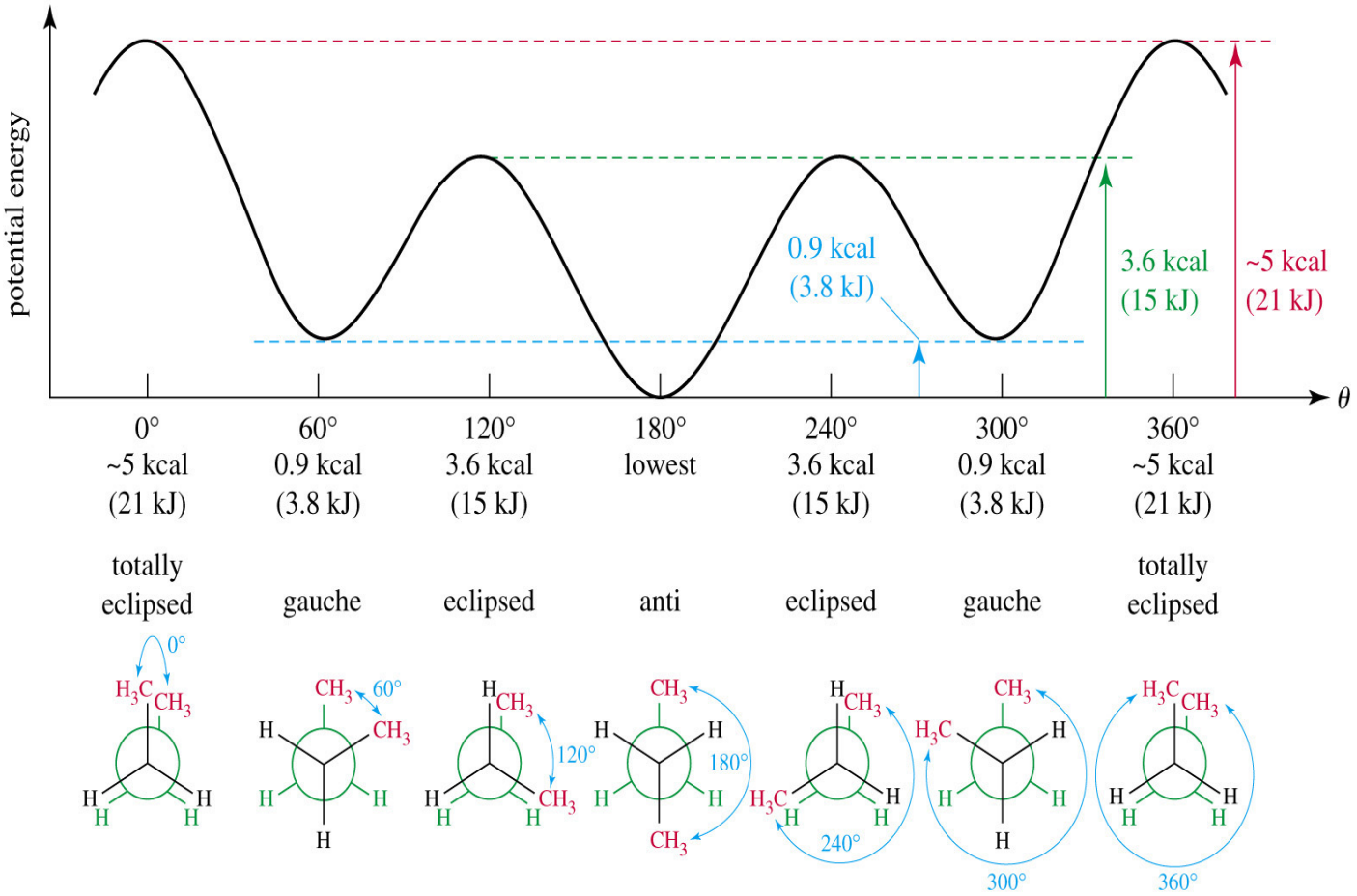
# CompChem (RS)

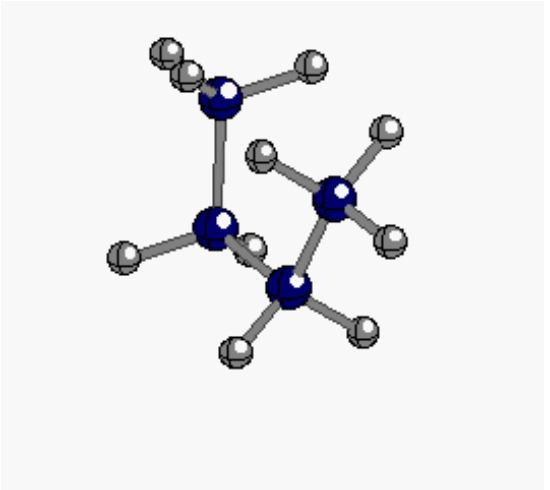



**and big ones**

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

## Different conformations have different energies

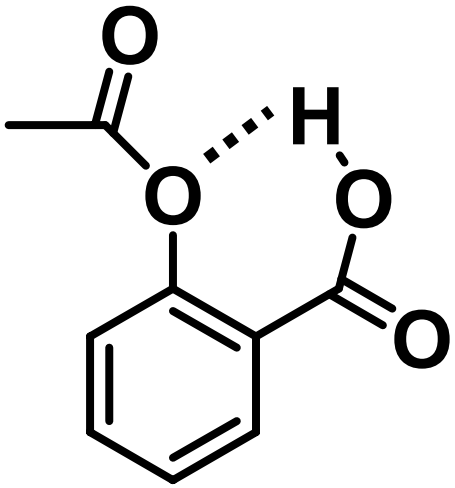


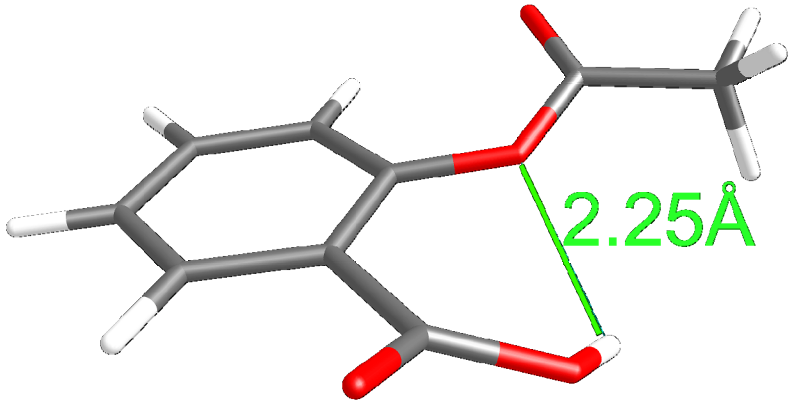
# CompChem (RS)

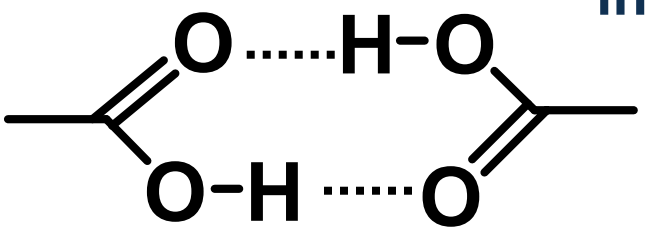
HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

## Non-covalent interactions

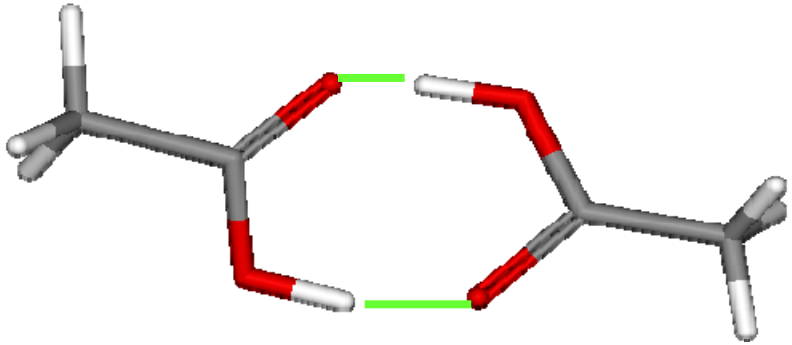



intramolecular



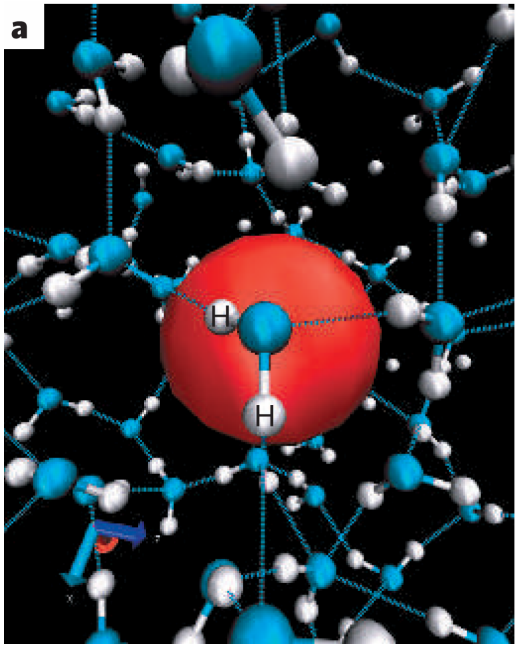

intermolecular

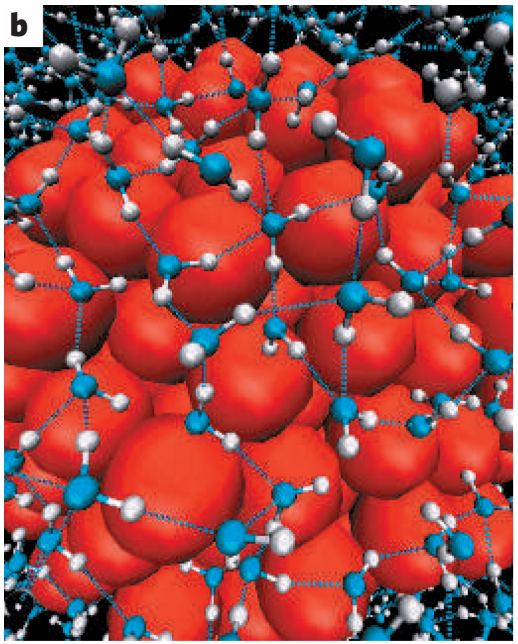


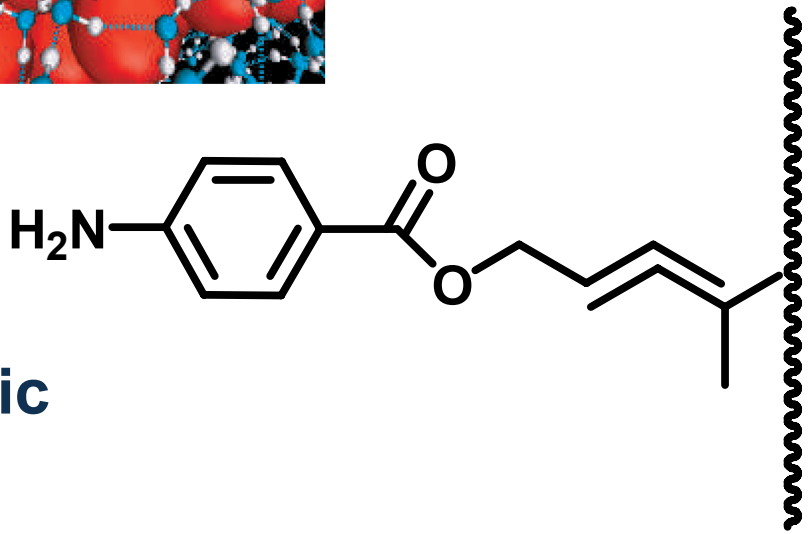
# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

## Hydrophobic interactions

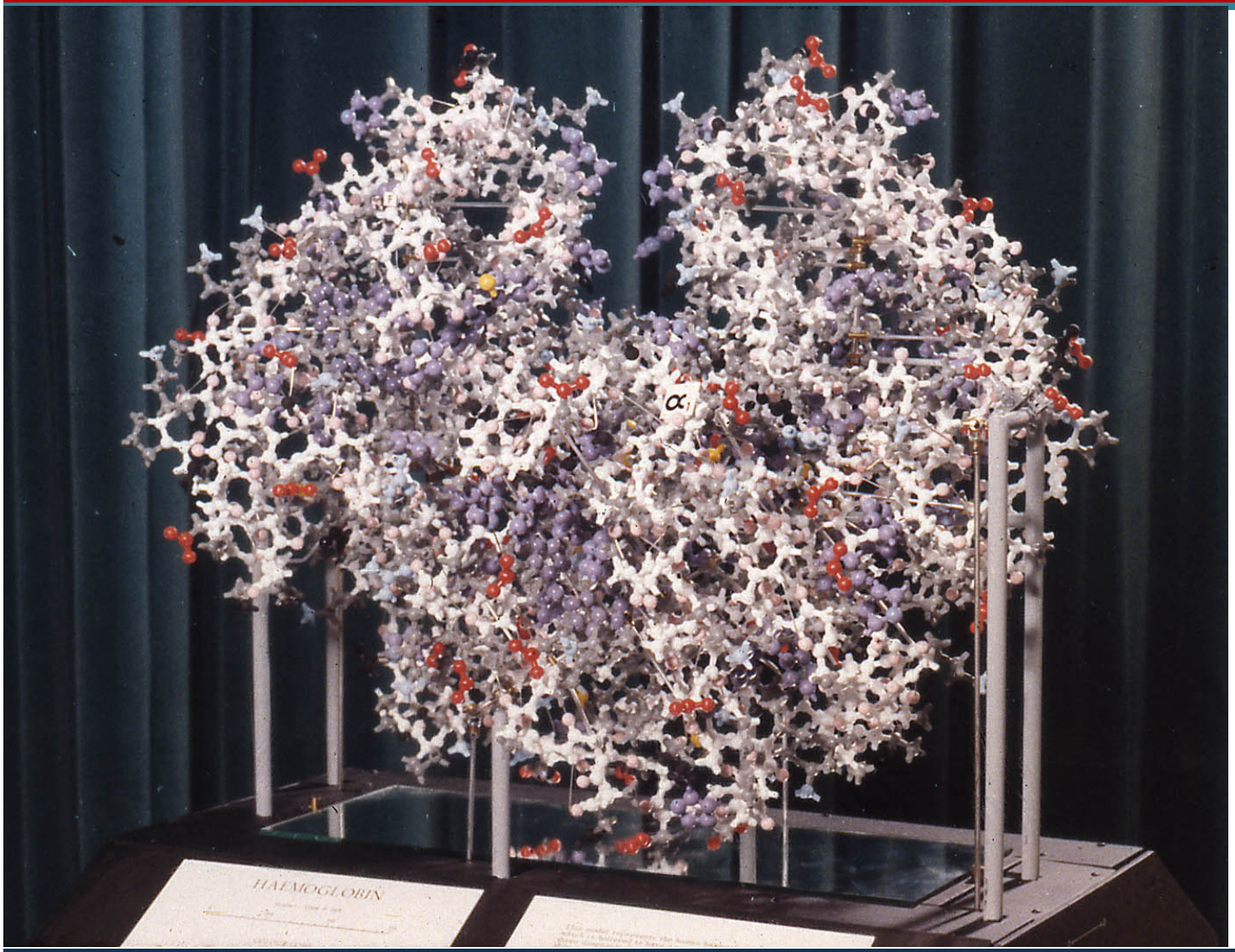



## local anesthetic

# CompChem (RS)

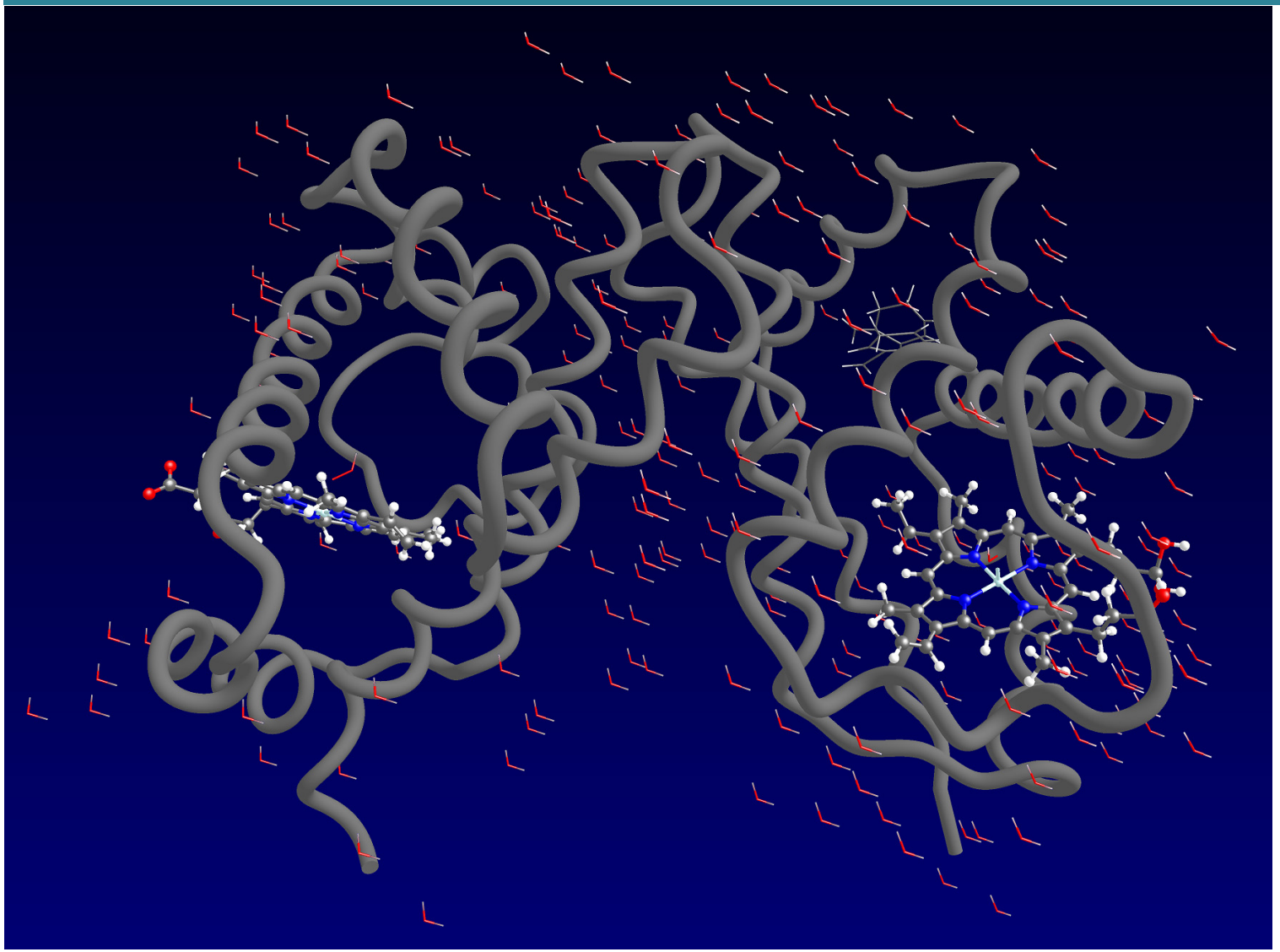



HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

**Proteins**

**- once upon a time…**

# CompChem (RS)



HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

**…and now**

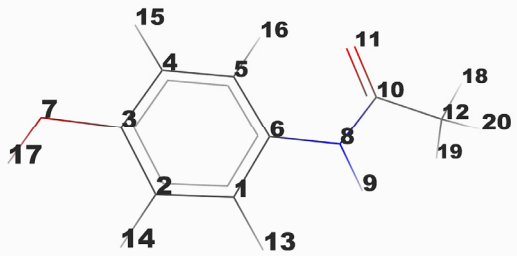
**http://pdbeta.rcsb.org/pdb**

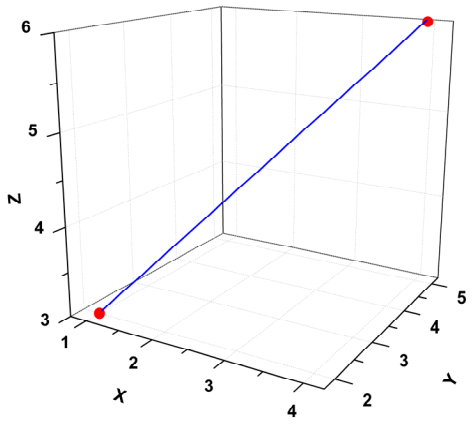
# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

## '3D molecular record'

```
@<TRIPOS>ATOM
      1 C          0.0000    0.0000    0.0000 C.ar     1 UNK1      -0.1092
      2 C          1.3927    0.0000    0.0000 C.ar     1 UNK1      -0.0518
      3 C          2.0727    1.2204    0.0000 C.ar     1 UNK1      -0.1064
      4 C          1.3639    2.4134    0.0005 C.ar     1 UNK1      -0.0690
      5 C         -0.0319    2.4068    0.0026 C.ar     1 UNK1      -0.1573
      6 C         -0.7267    1.1934   -0.0016 C.ar     1 UNK1      -0.0269
      7 C         -0.7394    3.7233    0.0011 C.2      1 UNK1       0.3421
      8 O         -0.9618    4.3391   -1.0216 O.2      1 UNK1      -0.3029
      9 C         -1.0895    4.2358    1.3874 C.3      1 UNK1      -0.1485
     10 C         -2.1828    5.2984    1.4208 C.3      1 UNK1      -0.1511
     11 C         -1.9515    6.2891    2.5458 C.2      1 UNK1       0.3841
     12 O         -1.2927    6.1731    3.5620 O.2      1 UNK1      -0.3747
     13 O         -2.5587    7.4901    2.3728 O.3      1 UNK1      -0.3074
     14 S         -3.7989    4.4289    1.6334 S.3      1 UNK1       0.0321
     15 C         -5.0031    5.5284    0.8297 C.3      1 UNK1      -0.1896
     16 C         -5.0248    5.4475   -0.6758 C.2      1 UNK1       0.3805
     17 O         -5.7473    4.7750   -1.3865 O.2      1 UNK1      -0.3712
     18 O         -4.1411    6.2609   -1.3020 O.3      1 UNK1      -0.3020
     19 C         -2.2323    1.1377   -0.0348 C.3      1 UNK1      -0.0511
     20 C         -2.7633    0.0507    0.8882 C.3      1 UNK1      -0.1162
     21 C         -2.6987    0.8996   -1.4646 C.3      1 UNK1      -0.1132
     22 C          2.1713   -1.2874   -0.0005 C.3      1 UNK1      -0.0311
     23 C          1.8754   -2.0967   -1.2542 C.3      1 UNK1      -0.1119
     24 C          1.8751   -2.0995    1.2514 C.3      1 UNK1      -0.1119
     25 H         -0.5634   -0.9488    0.0128 H        1 UNK1       0.1242
     26 H          3.1689    1.2392   -0.0022 H        1 UNK1       0.1099
     27 H          1.9034    3.3673   -0.0051 H        1 UNK1       0.1058
     28 H         -0.1476    4.6265    1.8285 H        1 UNK1       0.1005
     29 H         -1.3765    3.3734    2.0289 H        1 UNK1       0.1034
     30 H         -2.1965    5.8660    0.4547 H        1 UNK1       0.1423
     31 H         -2.3744    8.0555    3.1167 H        1 UNK1       0.2329
     32 H         -4.8568    6.5791    1.1498 H        1 UNK1       0.1125
     33 H         -5.9898    5.2145    1.2267 H        1 UNK1       0.1213
     34 H         -4.1736    6.1229   -2.2434 H        1 UNK1       0.2352
     35 H         -2.6465    2.1201    0.3164 H        1 UNK1       0.0819
```



# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

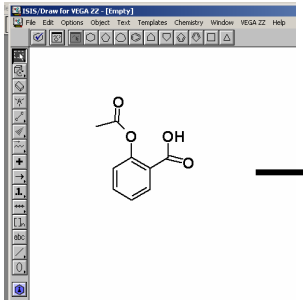
**Input**

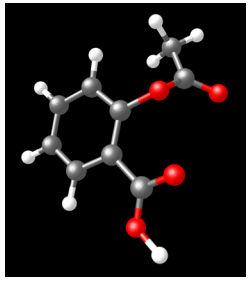
- atomic positions and connectivity
- instructions, in a manner 'what program should do'
- direction to external data
  - additional directives
  - additional datasets (force fields, parameters)

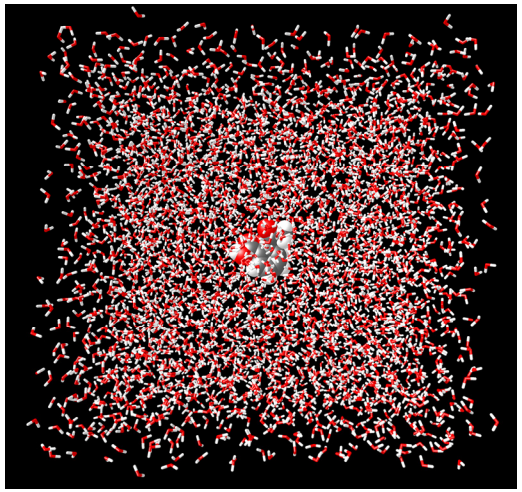
# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

## *Molecular dynamics*

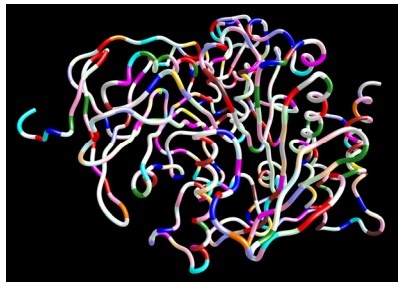


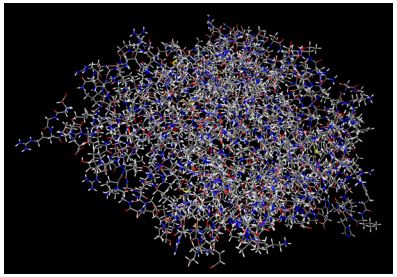


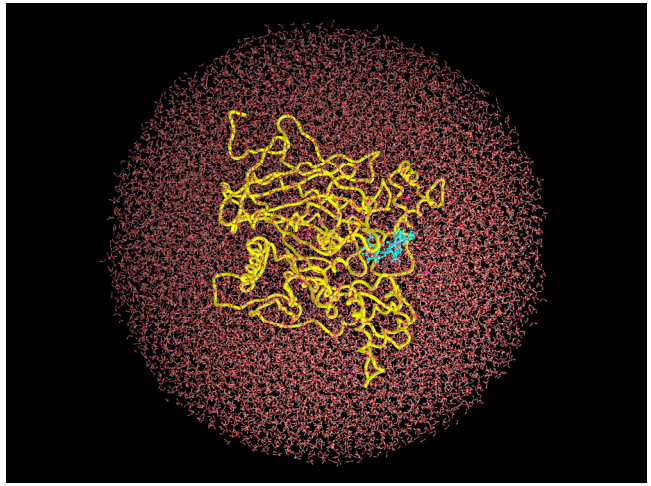


ready for simulation

structure
+
topology

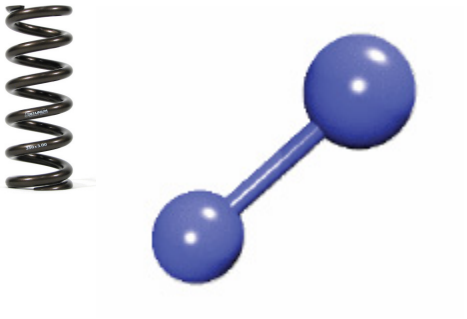




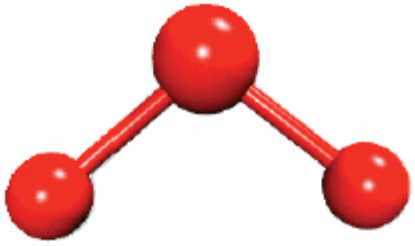
# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

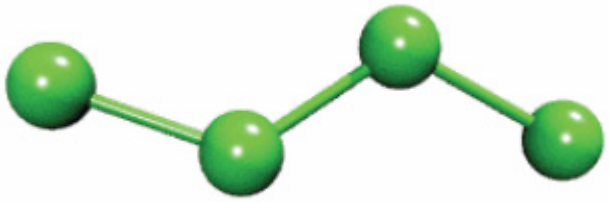
## Molecular dynamics, energy terms



Bond



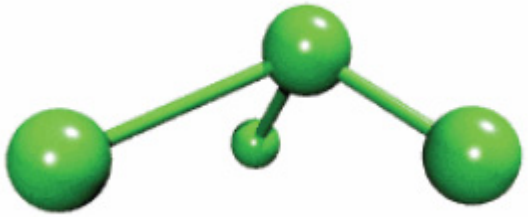
Angle



Dihedral

Nonbonded

- Electrostatic
- Lenard – Jones (LJ - vdW)



Improper

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

## Molecular dynamics, energy function

$$U(\vec{R}) = \underbrace{\sum_{bonds} k_i^{bond}(r_i - r_0)^2}_{U_{bond}} + \underbrace{\sum_{angles} k_i^{angle}(\theta_i - \theta_0)^2}_{U_{angle}} +$$

$$\underbrace{\sum_{dihedrals} k_i^{dihe}[1 + \cos(n_i\phi_i + \delta_i)]}_{U_{dihedral}} +$$

$$\underbrace{\sum_i \sum_{j \neq i} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \sum_i \sum_{j \neq i} \frac{q_i q_j}{\epsilon r_{ij}}}_{U_{nonbond}}$$

**$U_{bond}$**

**oscillation about equil. bond length**

**$U_{angle}$**

**oscillation about equil. bond angle**

**$U_{dihedral}$**

**torsion about central bond**

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

## Molecular dynamics, energy function

### nonbonded terms

$$\sum_{nonbonded} \frac{q_i q_j}{4\pi D r_{ij}} + \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right]$$

$q_i$ – partial atomic charge

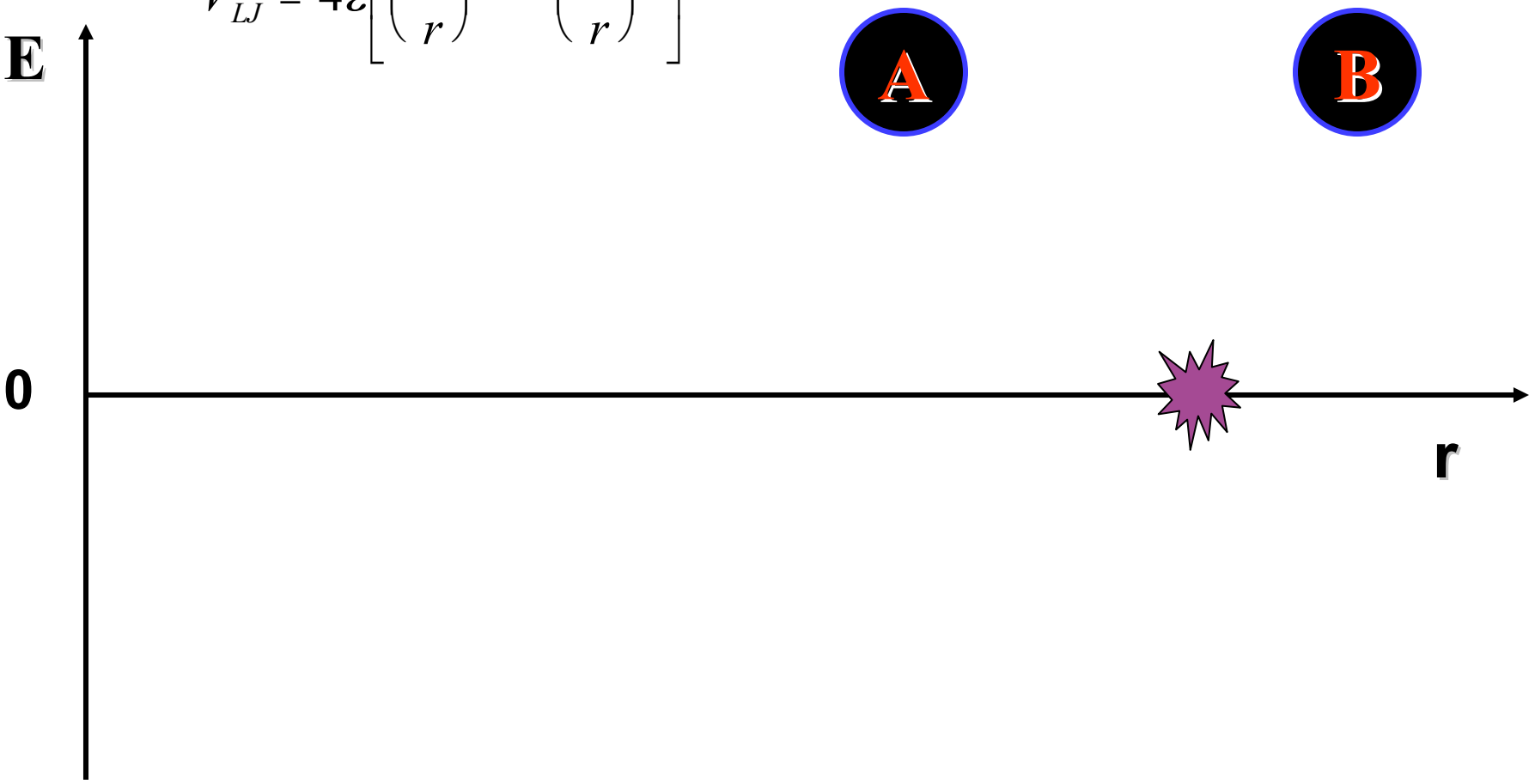
D – dielectric constant

$\varepsilon$ - Lenard-Jones well-depth

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

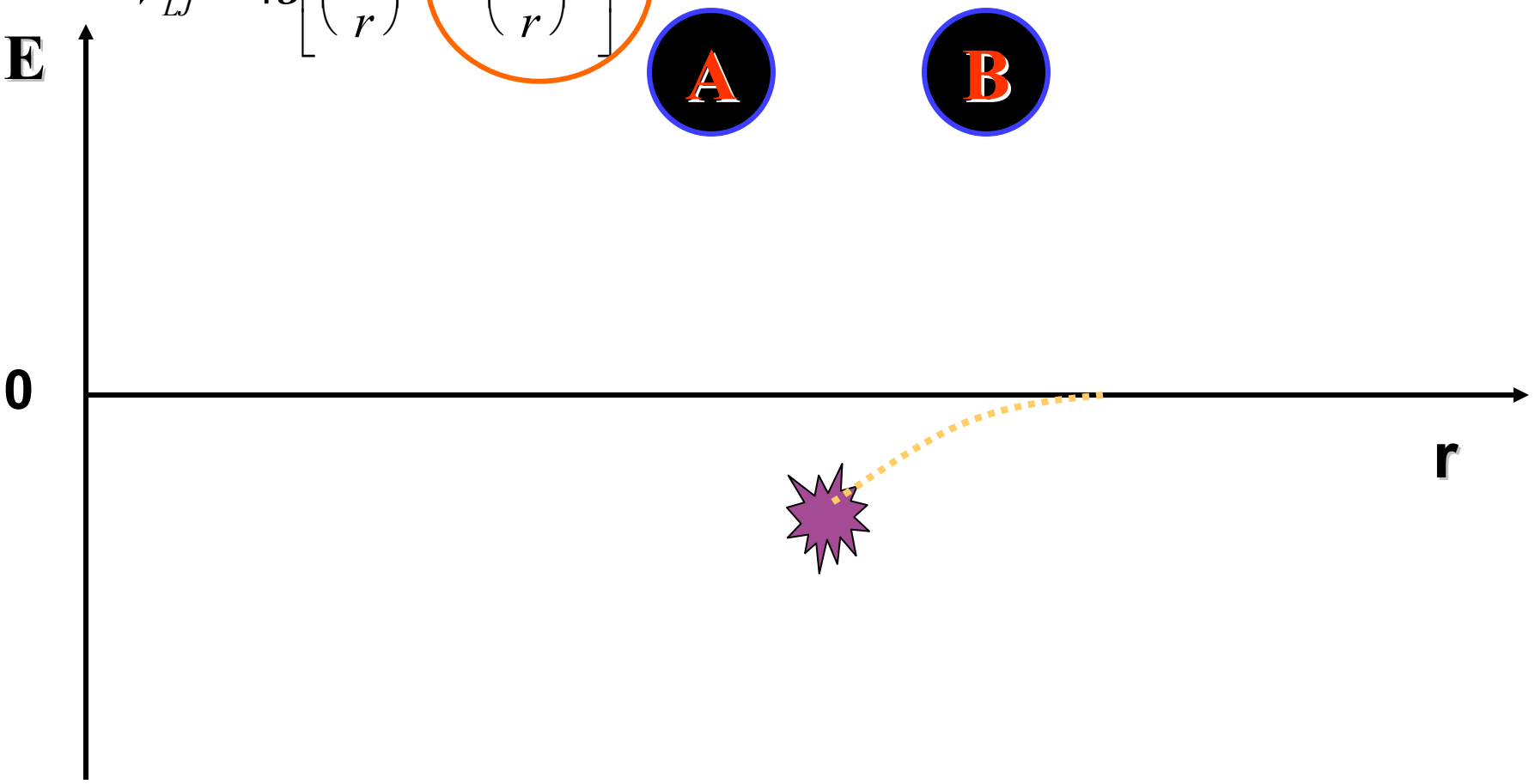
$$V_{LJ} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$



# CompChem (RS)



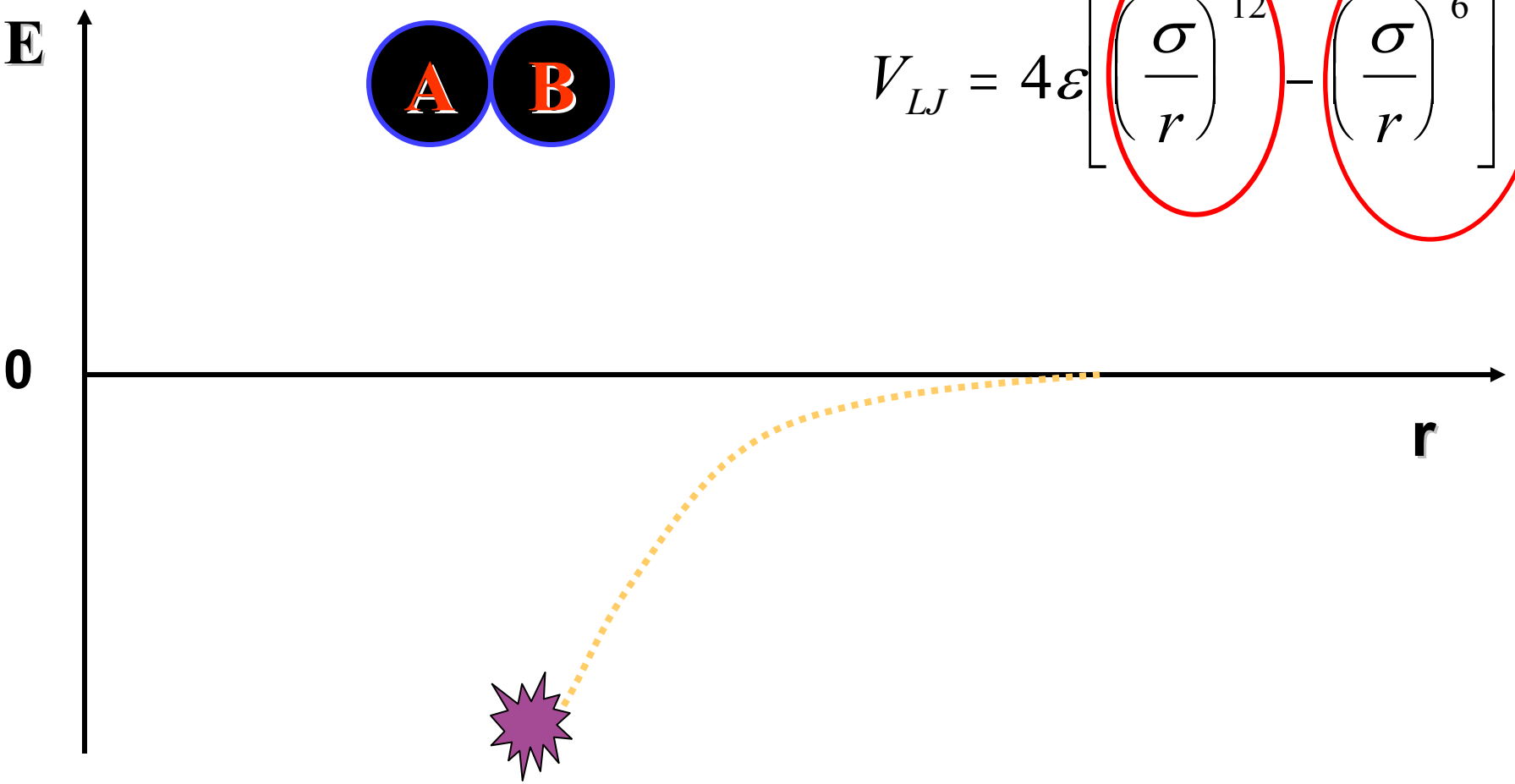

$$V_{LJ} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$



# CompChem (RS)

HP-SEE

High-Performance Computing Infrastructure for South East Europe's Research Communities

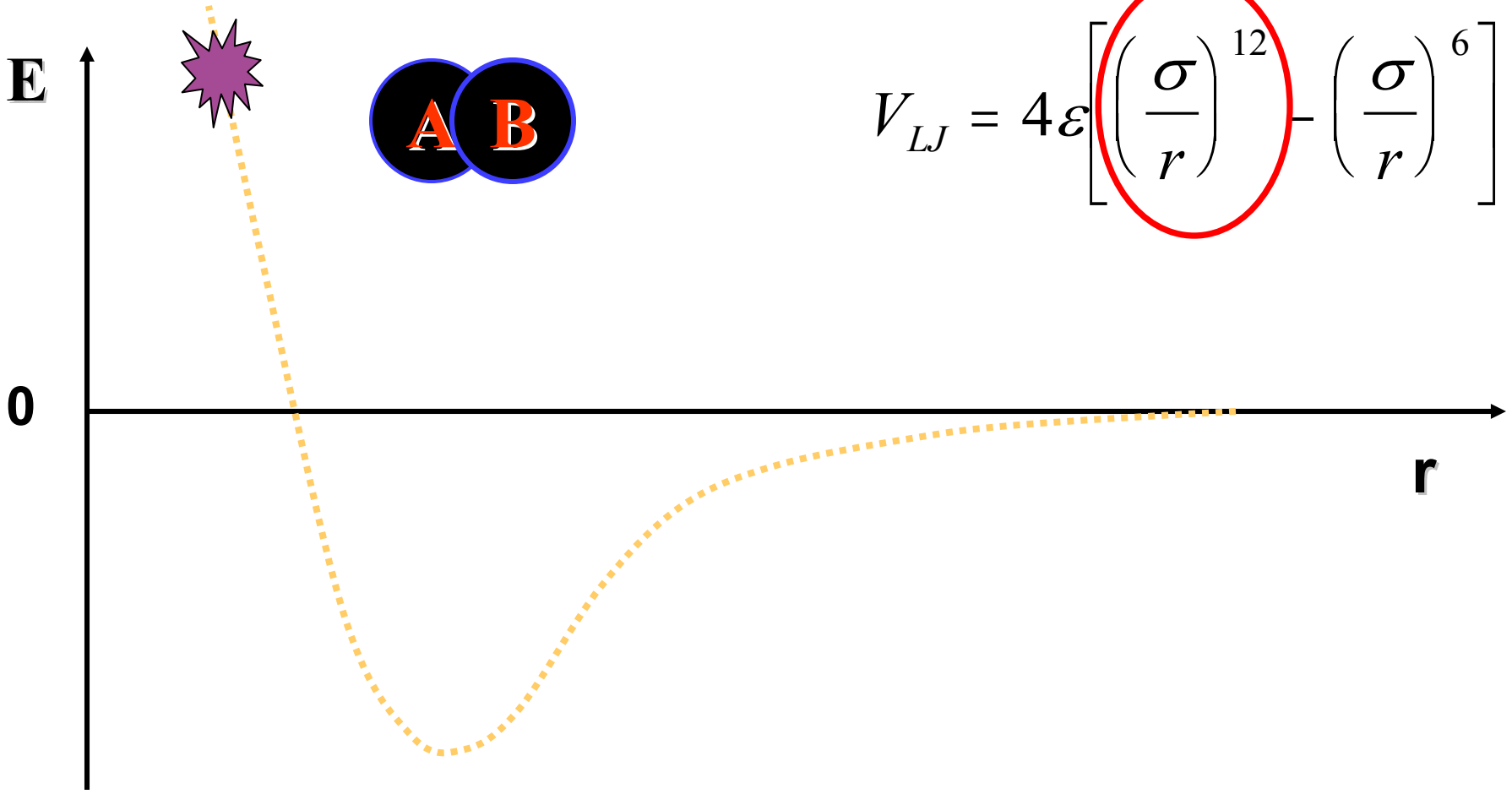



$$V_{LJ} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

# CompChem (RS)

HP-SEE

High-Performance Computing Infrastructure for South East Europe's Research Communities




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

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

**Classical dynamics**

**Energy function** $U(\vec{r}_1, \vec{r}_2, \vec{r}_3 \ldots \vec{r}_N) = U(\vec{R})$

**used to determine the force on each atom**

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i = -\vec{\nabla} U(\vec{R})$$

**- result with the set of 3N coupled 2nd order differ. equat. that can be propagated forward or backward in time**

**- initial atomic position from experiment or precalculated**

**- randomly taken velocities from Boltzman distribution**

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

## Classical dynamics

### Discretization in time for computing

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i = -\vec{\nabla} U(\vec{R})$$

$$\mathbf{r}(t+\delta t) \approx \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2$$

+

$$\mathbf{r}(t-\delta t) \approx \mathbf{r}(t) - \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2$$

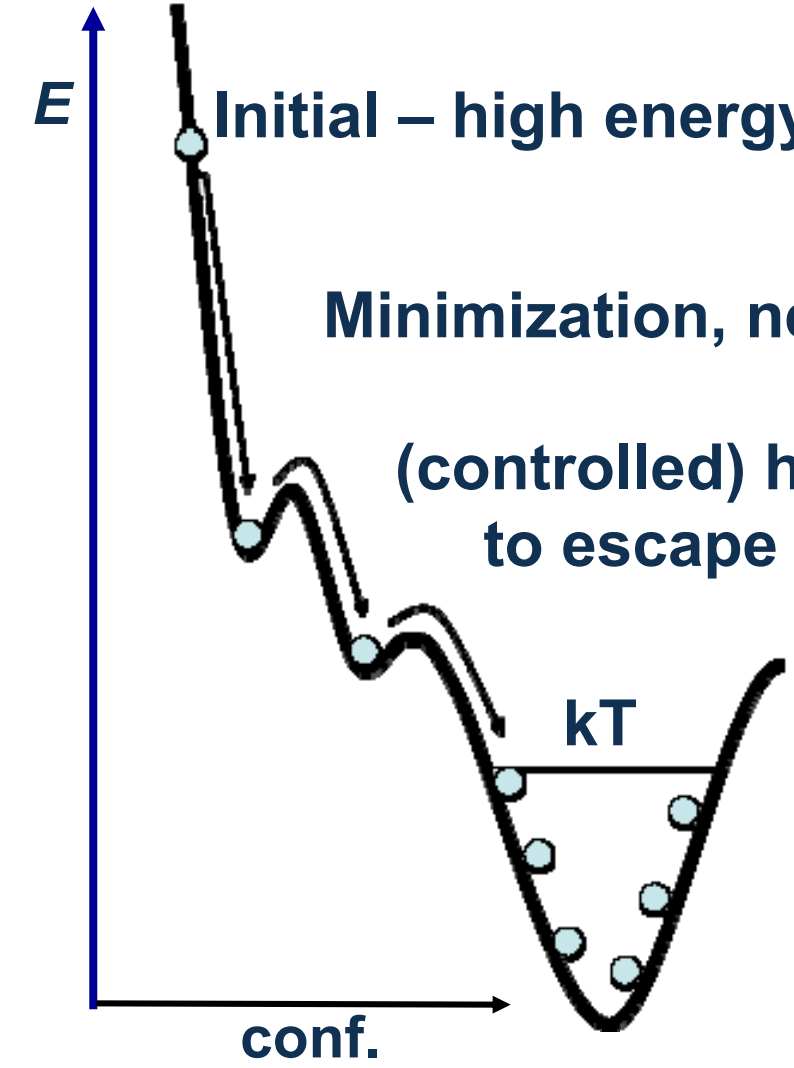
Use (atomic) positions and acceleration in time *t* and position in time *t-δt* to calculate position at time *t+δt*

Verlet algorithm

$$\mathbf{r}(t+\delta t) \approx 2\mathbf{r}(t) - \mathbf{r}(t-\delta t) + \mathbf{a}(t)\delta t^2$$

$$\mathbf{a} = -\vec{\nabla} U(\vec{R}) / m_i$$

# CompChem (RS)



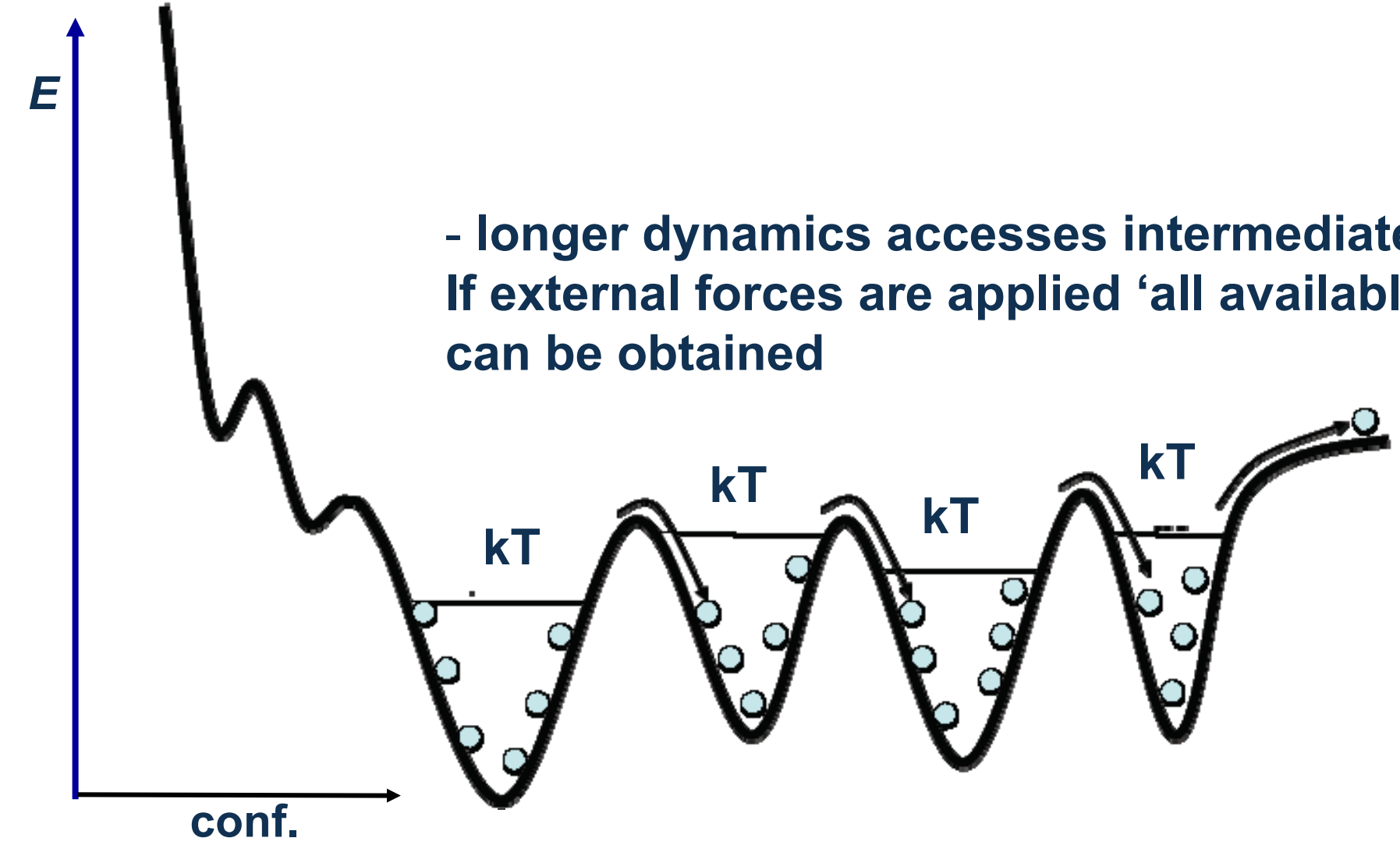

Initial – high energy

Minimization, nearby minima

(controlled) heating, cooling or equilibration
to escape local minima

Initial dynamics to sample thermally accessible states

# CompChem (RS)



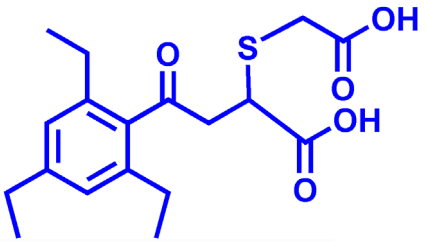

- longer dynamics accesses intermediate states. If external forces are applied 'all available' states can be obtained

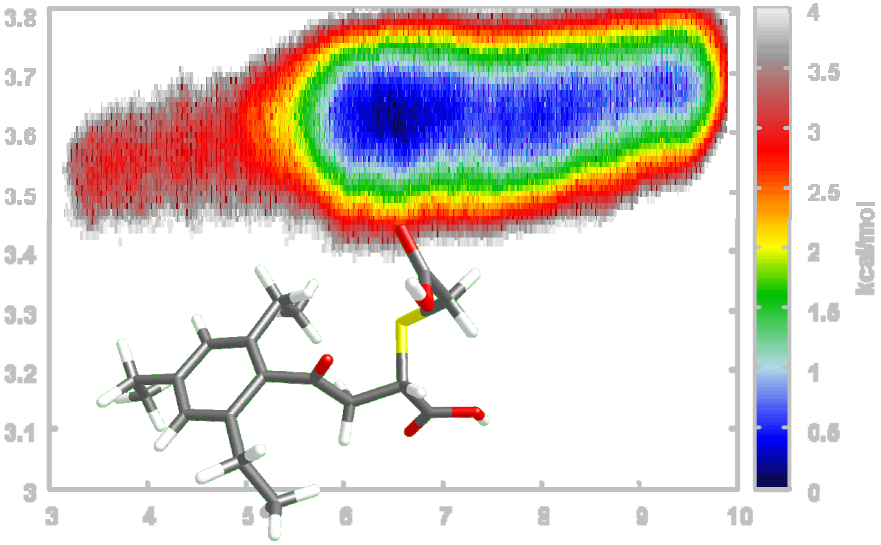
# CompChem (RS)

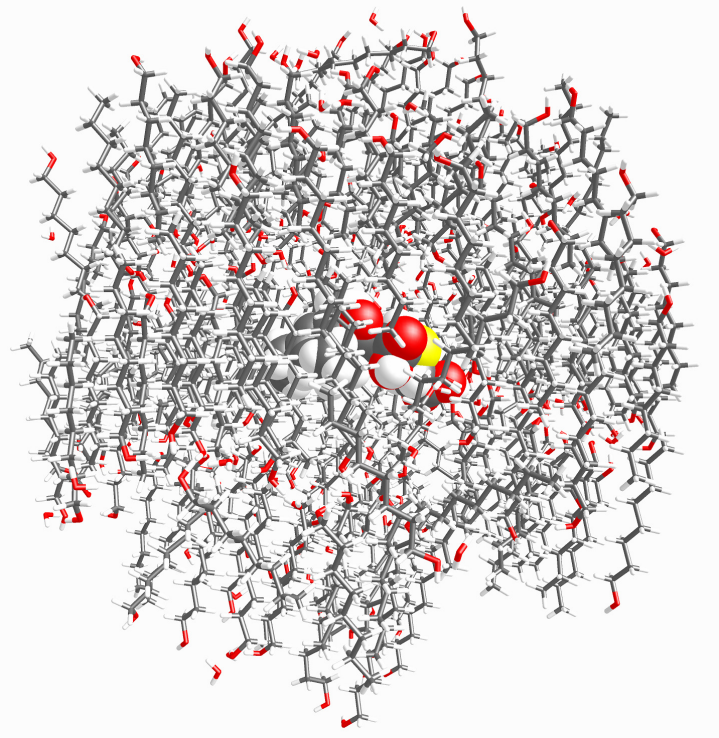
HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

## Example I

- conformatinal space of small molecule
- biasing force applied to speed-up sampling
- free-energy surface obtained



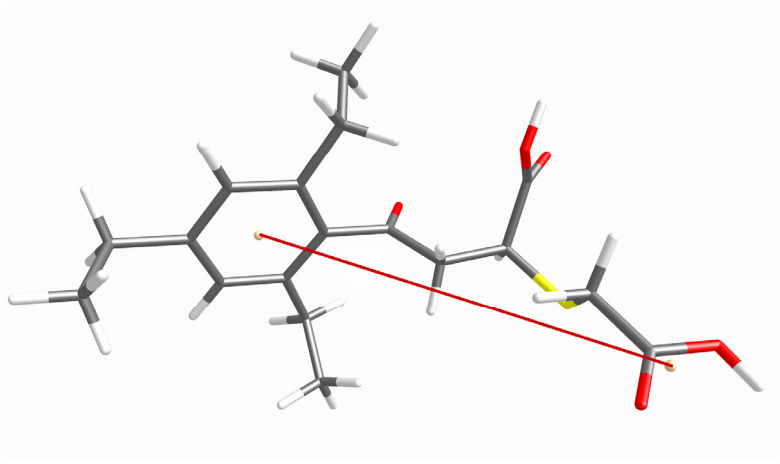



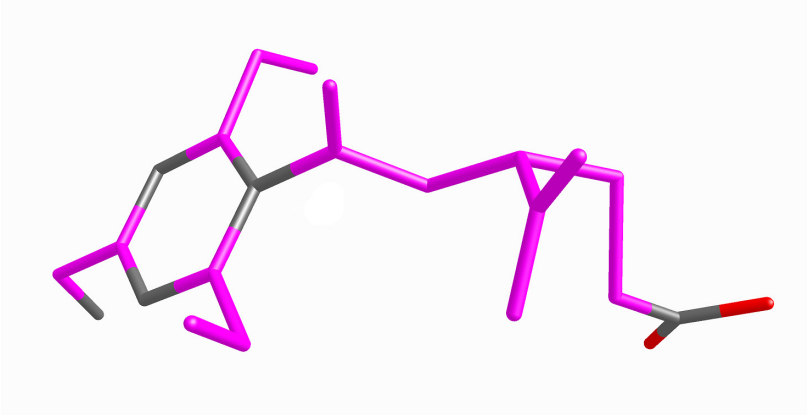
# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- biasing forces



- distance



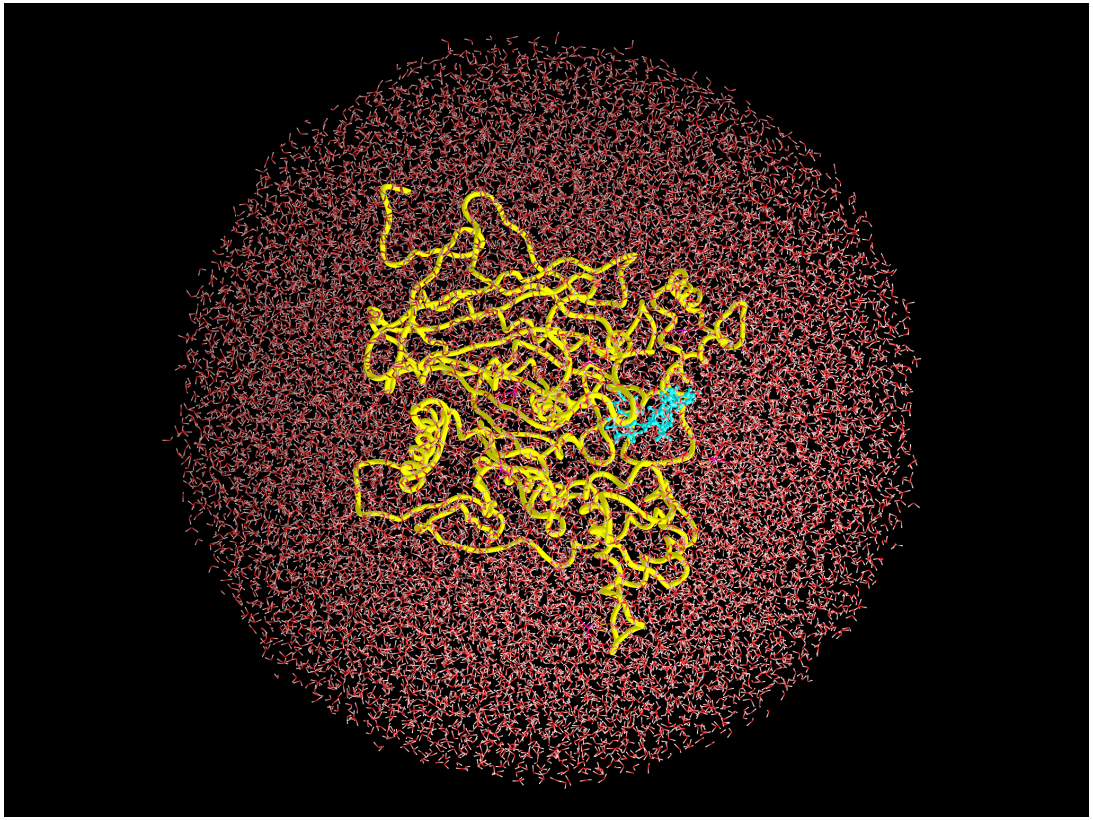
- radius of gyration

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

## Example II

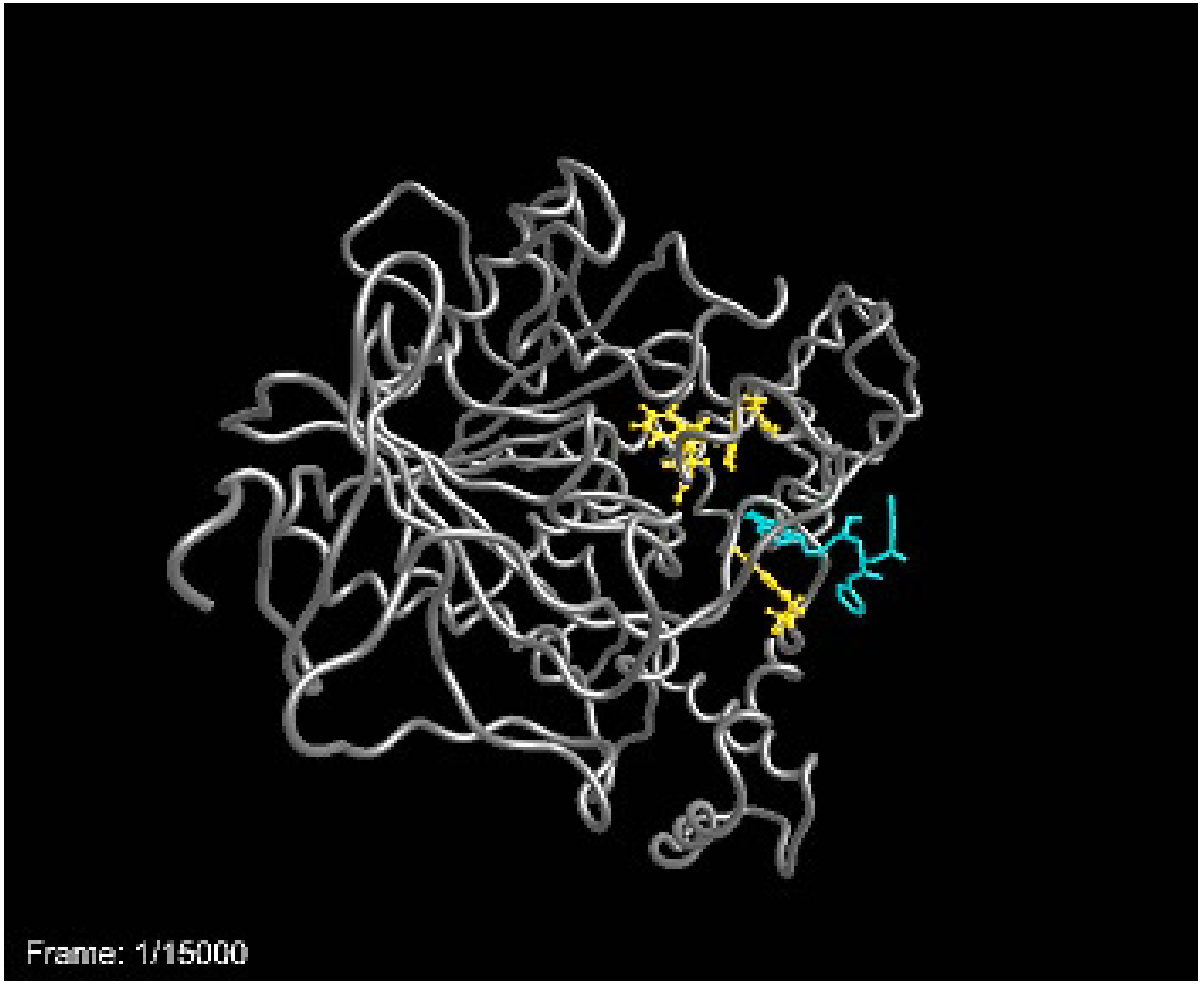
- interaction of the ligand with a protein
- acethylcholinesterase inhibitor



# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

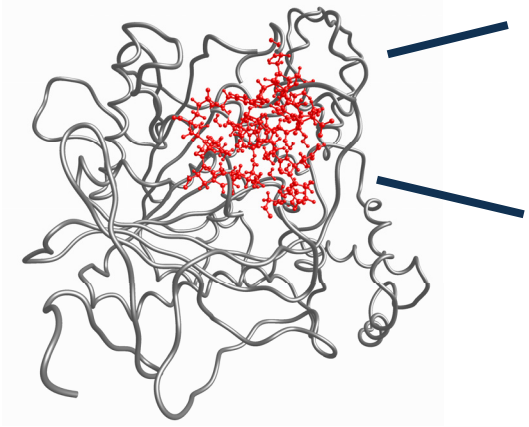
- interaction of the ligand with a protein

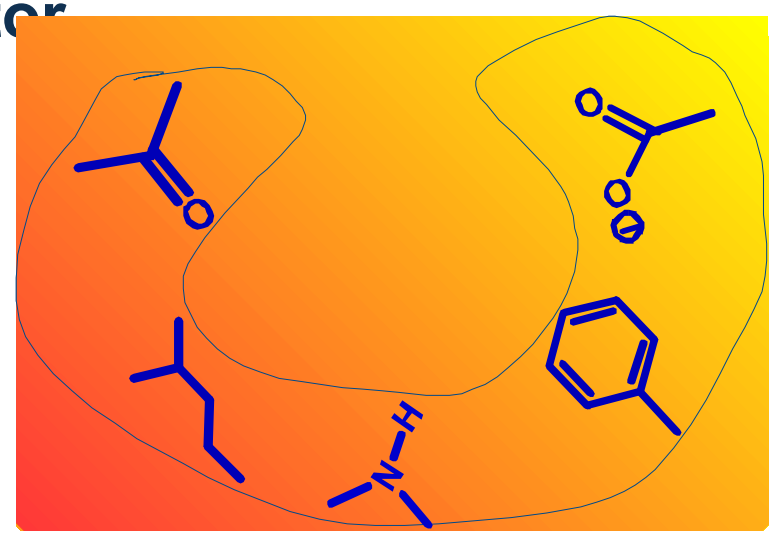


# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

## Ligand and receptor



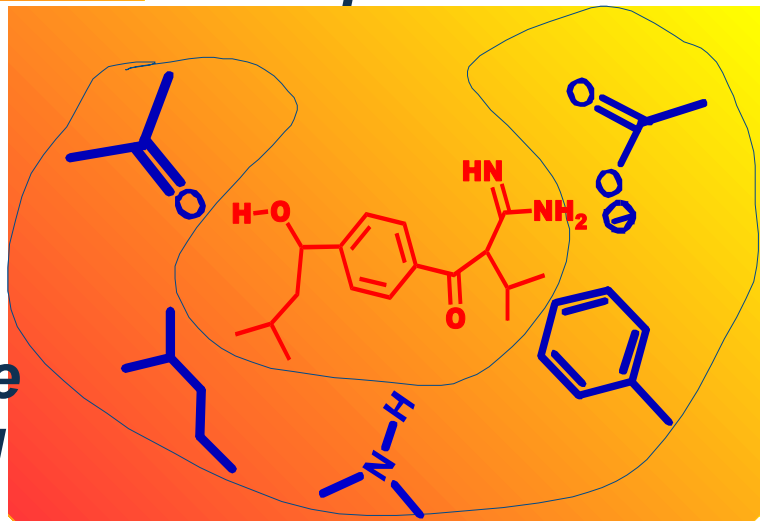


***part of protein able to accommodate natural substrate or synthetic small molecule is a receptor***

**- if there is a many ligands**

**-method should classify ones that could fit**

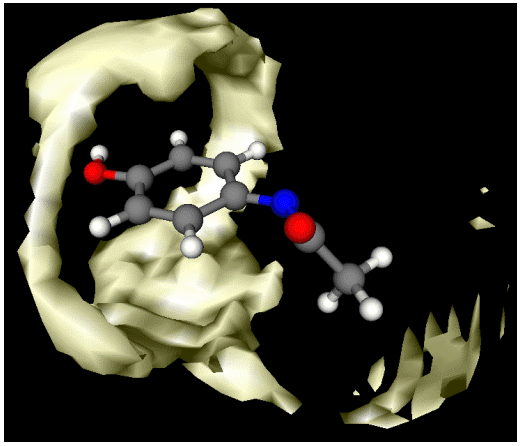
***molecule that fit the receptor is a ligand***

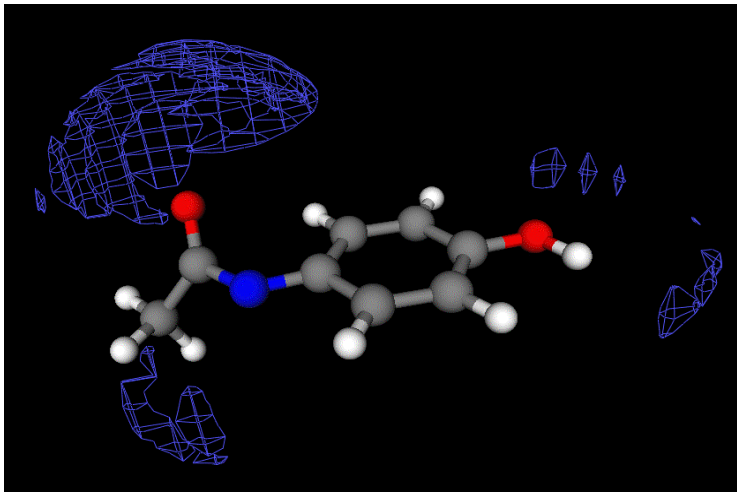


# CompChem (RS)

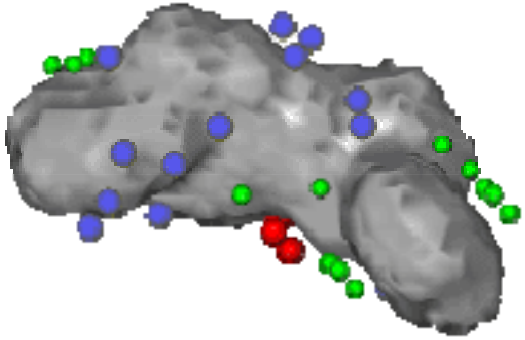
HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

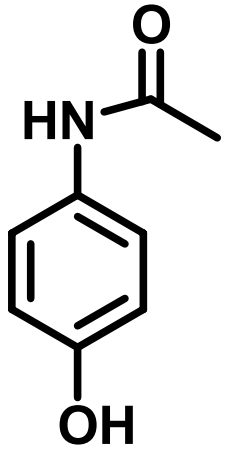
## Variation on theme

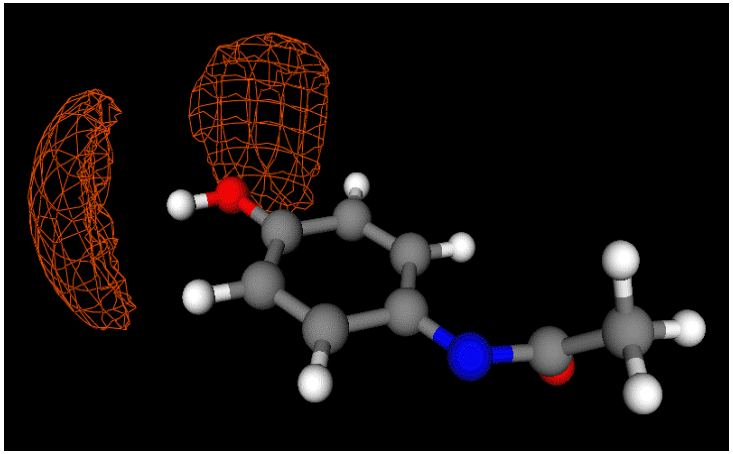




**combined = pharmacophore**





**many small molecules, so which ones fit to pharmacophore**

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

*Virtual screening*

- molecular mechanics in spirit
- fitting of ligand to target receptor, or
- fitting of ligand to target pharmacophore pattern
- often few thousands of compounds should be screened
- high computational demand
- there are online academic and commercial servers for such purposes

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

**Size *vs*. accuracy, quest of computational power**

- **DFT and *ab initio* methods**
  - **for geometry optimization**
  - **determination of reaction paths (probable course of (often simple) chemical reactions)**
  - **calculation of excited states of molecules (UV spectra, fluorescence etc.)**
  - **vibrational frequencies (bond stretching constants, IR spectra)**
  - **rotational barriers…**

**Used for parameterization of the methods on the lower level of theory.**

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

*Pre/post processing*

- pre processing
  - most often GUI (graphical user interface)
  - able to visualize molecules, show data needed, and change/adjust some associate data (molecular position in the space, charges, atom types, bond orders, different type of visualization…)
  - there are many software of this type which are almost exclusively free to academia
  - scripting is often available to speed-up, or make automatic specific user need
  - input data almost exclusively recorded as ASCII files

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- some of high quality

VMD – by University of Illinois @ Urbana
http://www.ks.uiuc.edu/Research/vmd/

*‘ takes advantage of inexpensive game technologies, graphics boards and stereo glasses to provide capabilities for PC users which were once only available in $40,000 workstations. For … a few hundred dollars, it is now possible to equip most desktop PC's with stereoscopic display capabilities, hardware accelerated 3-D rendering, and six-degree-of-freedom motion control capability ’*

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- **Maestro** by Schroedinger
  http://www.schrodinger.com/products/14/12/

- **VegaZZ** by Drug Design laboratory, Faculty of Pharmacy, University of Milan
  http://www.ddl.unimi.it/

- **PyMol** formerly by DeLano Scientific LLC, now user supported
  http://www.pymol.org/

- **CHIMERA** by UCSF (University of California, San Francisco)
  http://plato.cgl.ucsf.edu/chimera/

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

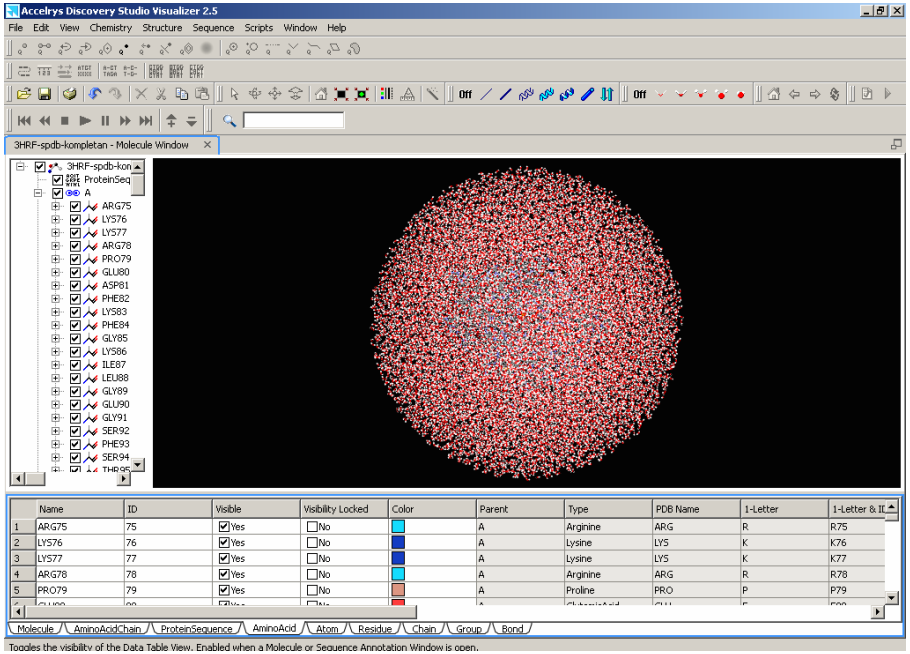
- Post processing
  - size of output is often huge, quest of storage on HPC facility
    - output must be visually inspected, still
    - additional data could be presented as tables and graphs
    - often graph editors, simple text editors, scripts for analysis, and similar are built-in to GUI
    - there are user-ready scripts which allow data reduction (tlc, python, C…)
    - common sparedsheets could be used (OpenOffice Calc, Excel)

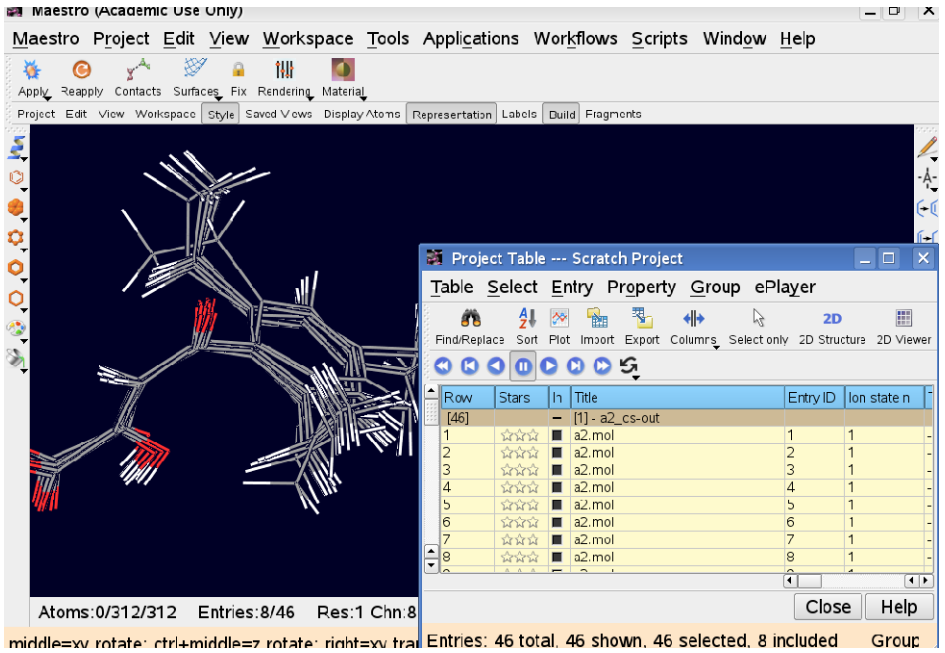
# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- graph plotting by GNU plot, Xmgrace or similar

- project tables are often used, supplied by GUI




-so, should we develop novel pre/post processing applications….?

# CompChem (RS)

HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

Thanks for your attention and time