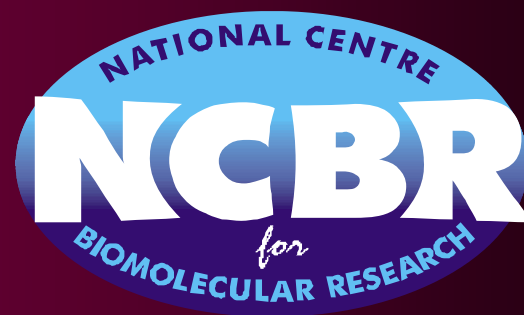


Úvod do molekulové dynamiky – simulace proteinů

Eva Fadrná

evaf@chemi.muni.cz



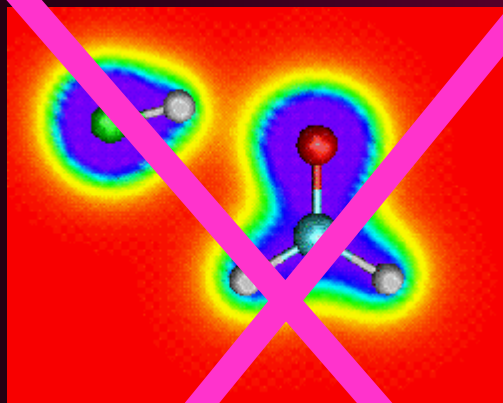
Molekulová mechanika

=

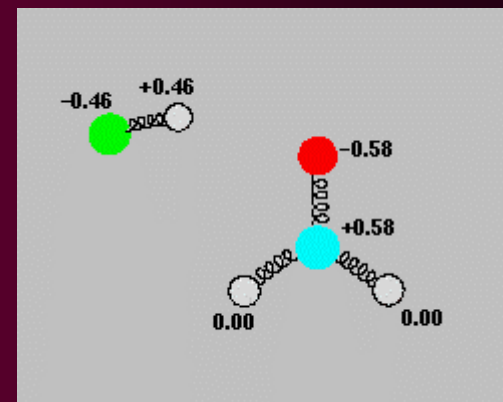
metoda silového pole

=

„force field“



Energie vypočtená
řešením Schrodingerovy
rovnice



Energie vypočtená
z empirických
konstant pnutí pružin a
z atomových nábojů

AMBER software

Assisted Model Building with Energy Refinement

<http://amber.scripps.edu/>

50 modulů – stavba modelu, MD, free energy, TI, analýza, ...

sander, gibbs, XleaP, mmpbsa, nmode, ptraj,...

Tutoriály:

simulace DNA, simulace HIV-I, streptavidin-biotin komplex,...

Fees: Academic/non-profit/government: \$400

Amber 9

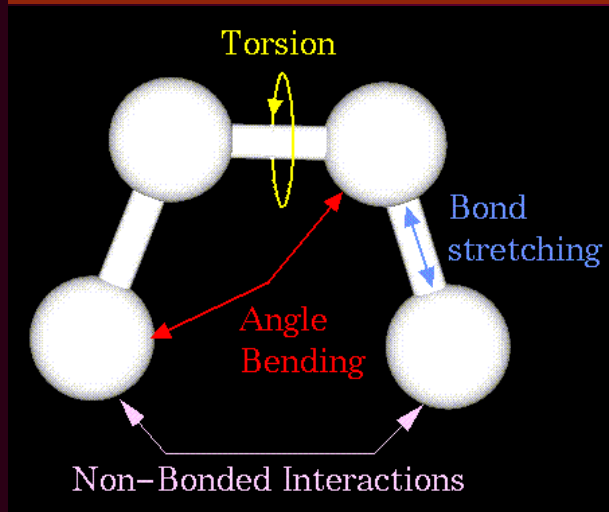


Rovnice silového pole

energie pnutí vazeb
energie rotace vazeb (torze)

energie deformace vazebných úhlů

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



vdW energie

elektrostatická energie

nevazebné interakce

Parametry silového pole

AMBER - parm94.dat (W. Cornell et al. 1995)

```
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)
```

definice atomových typů

vazby

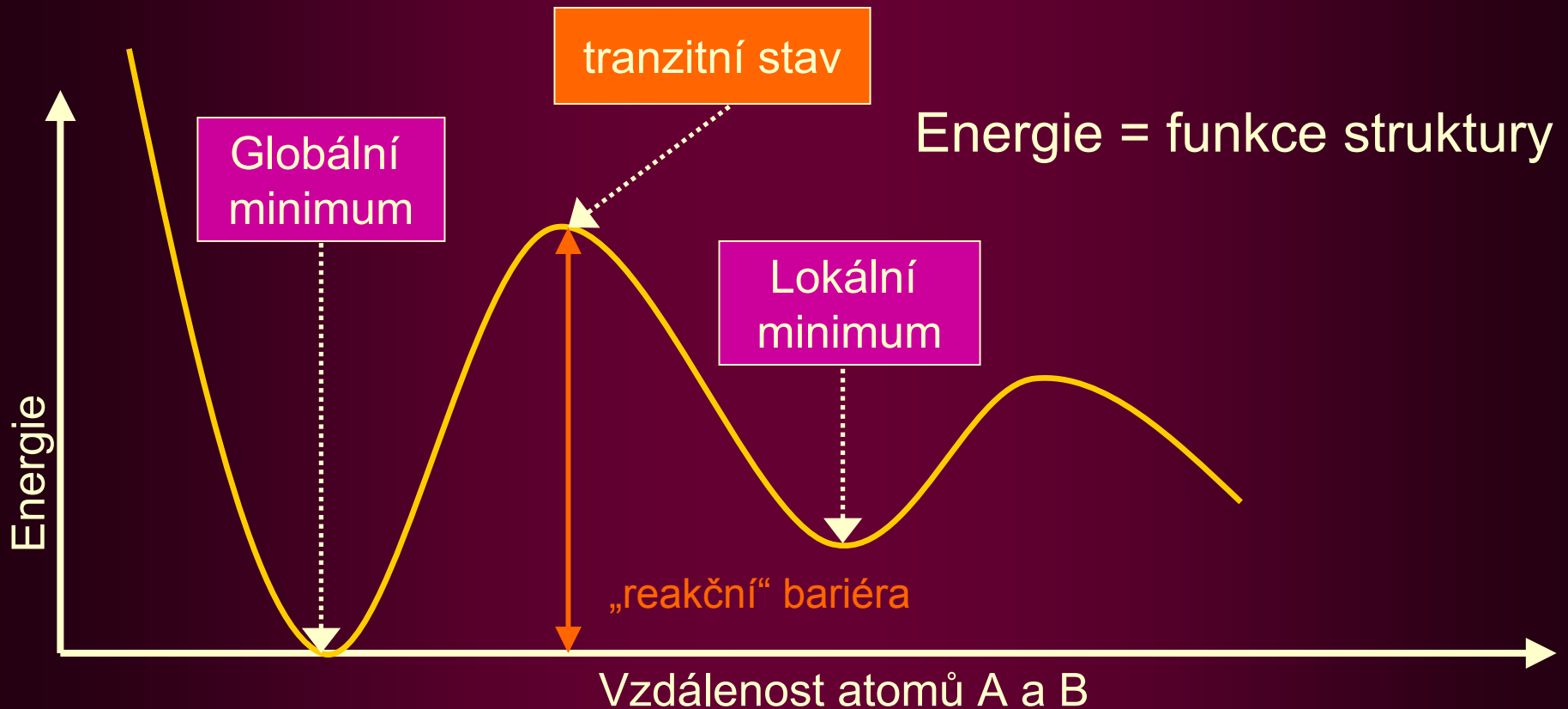
úhly

torze

vow proměry

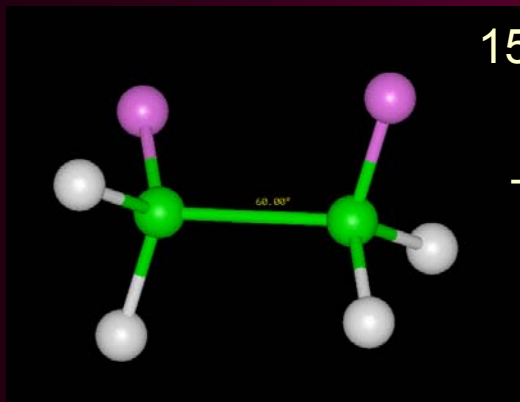
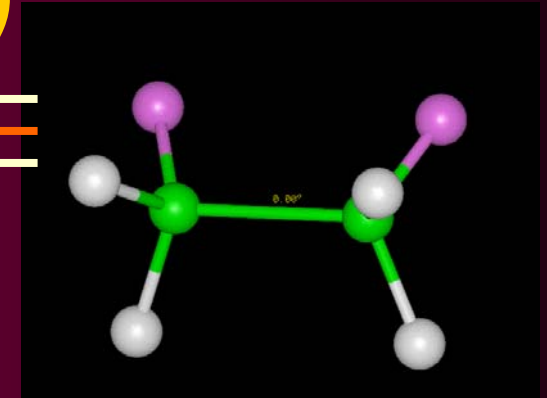
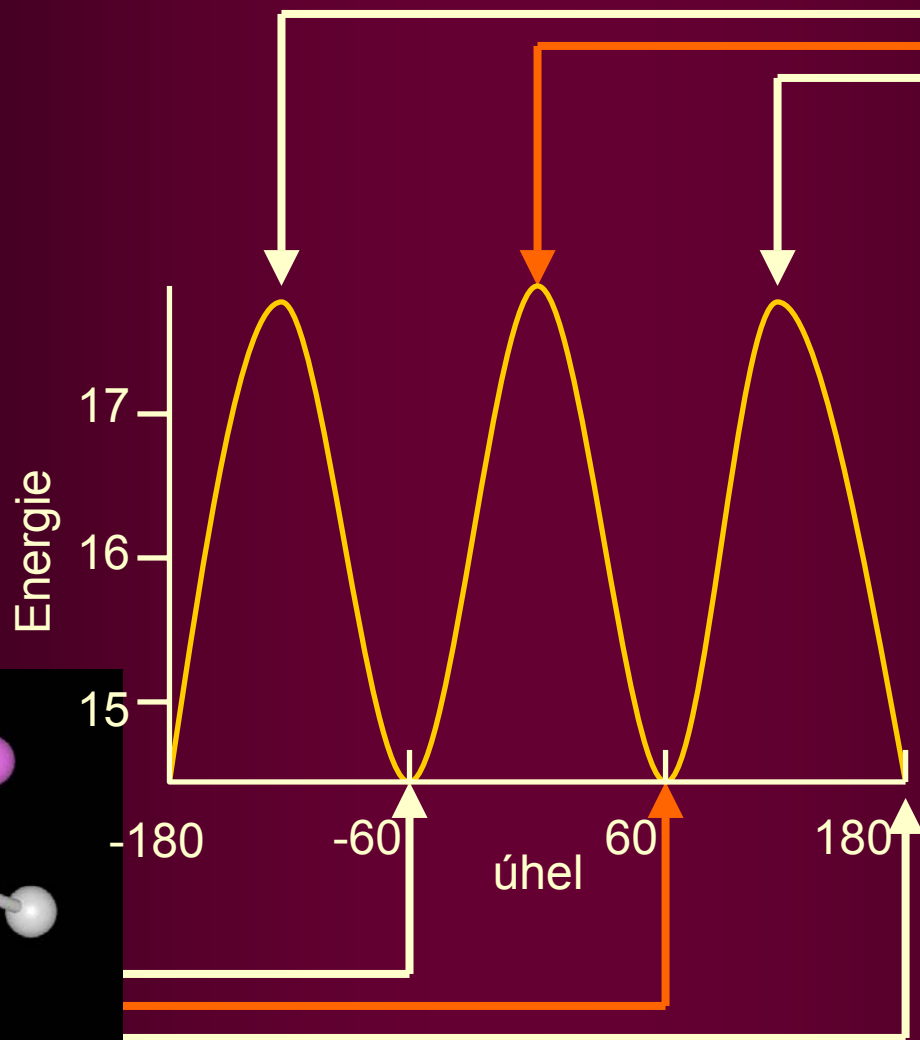
Hyperplocha potenciální energie

= Potential Energy (hyper)Surface (PES)

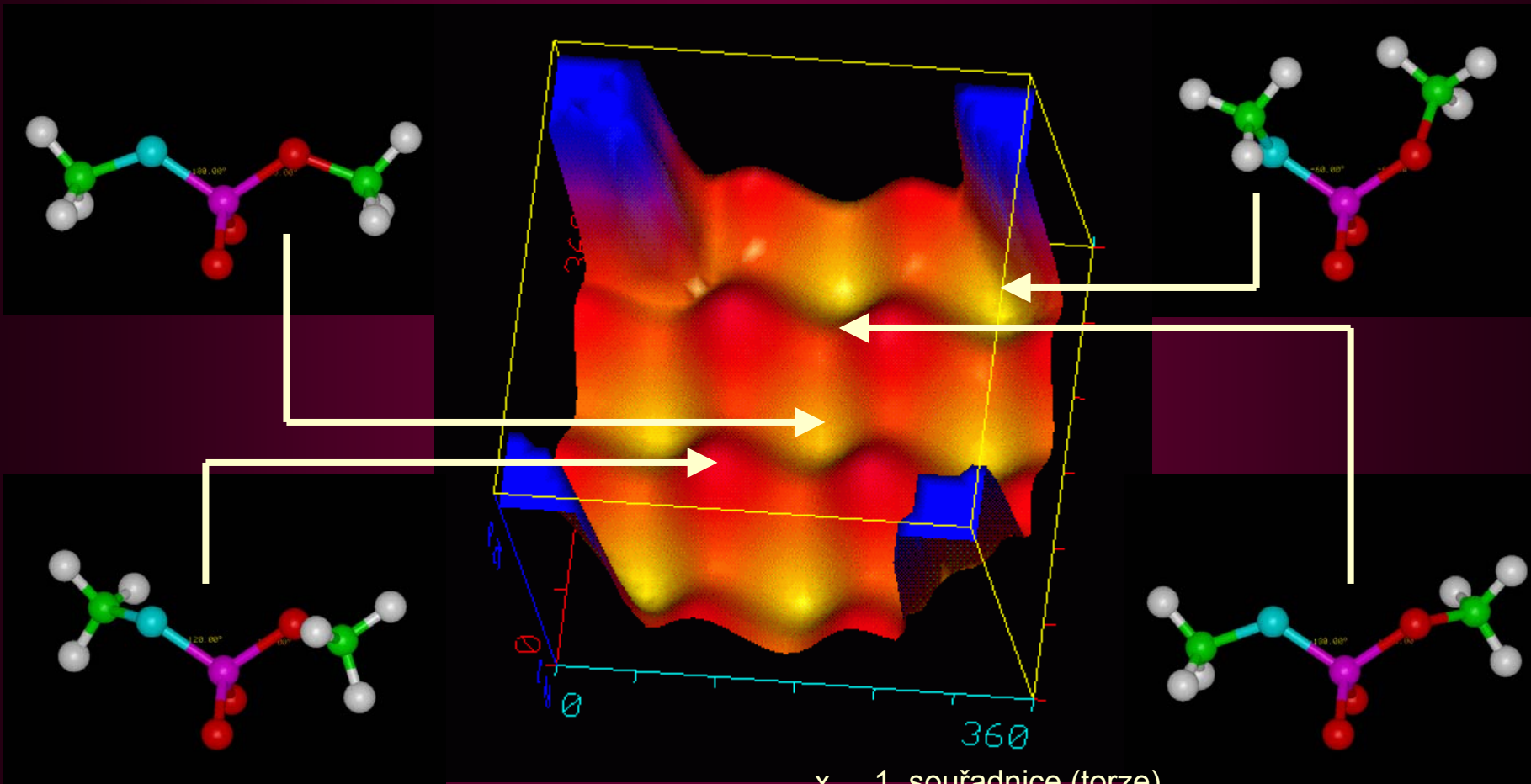


Hyperplocha potenciální energie (PES)

1D plocha:
 $E = f(\text{torze})$



Dimethylfosfát



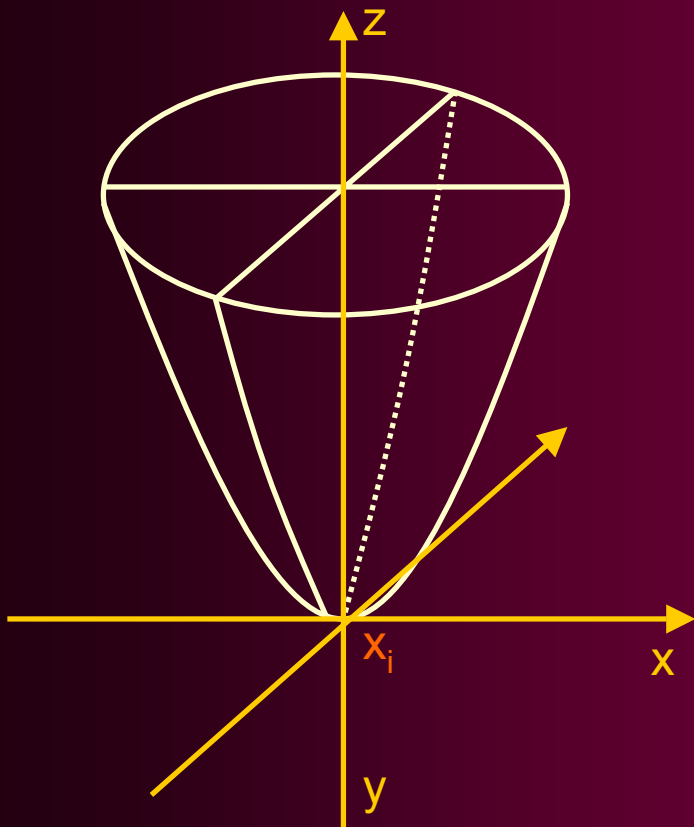
Energie = f (souřadnice)

x ... 1. souřadnice (torze)

y ... 2. souřadnice (torze)

z ... energie

Minimum funkce



Funkce 2 proměnných:

1. derivace v bodě $x_i = 0$
2. derivace v bodě $x_i > 0$

Funkce více proměnných:

gradient v bodě x_i ... vektor
1. derivací

Hessian v bodě x_i ... matice
2. parc. derivací

Minimalizace na PES

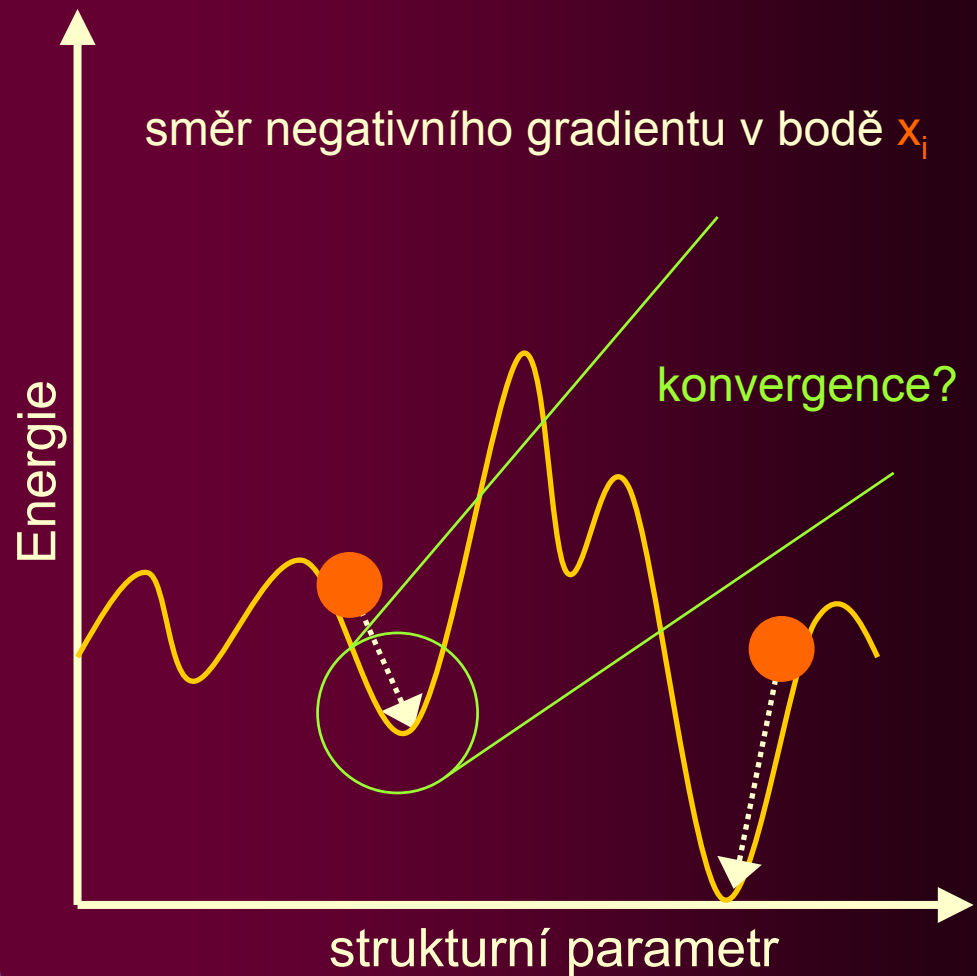
Spádové metody:

metoda největšího spádu
(*steepest descent*)

iterativní proces:

- startovní bod
- směr pohybu?
- prohledání regionu
- konvergenční kritéria?
- nový bod

minimum



Konvergenční kritéria

Rozdíl mezi 2 po sobě následujícími kroky:



Metody minimalizace

derivační:

- ⇒ metoda největšího spádu
- ⇒ konjugované gradienty
- ⇒ Newton-Raphsonova metoda

daleko od minima

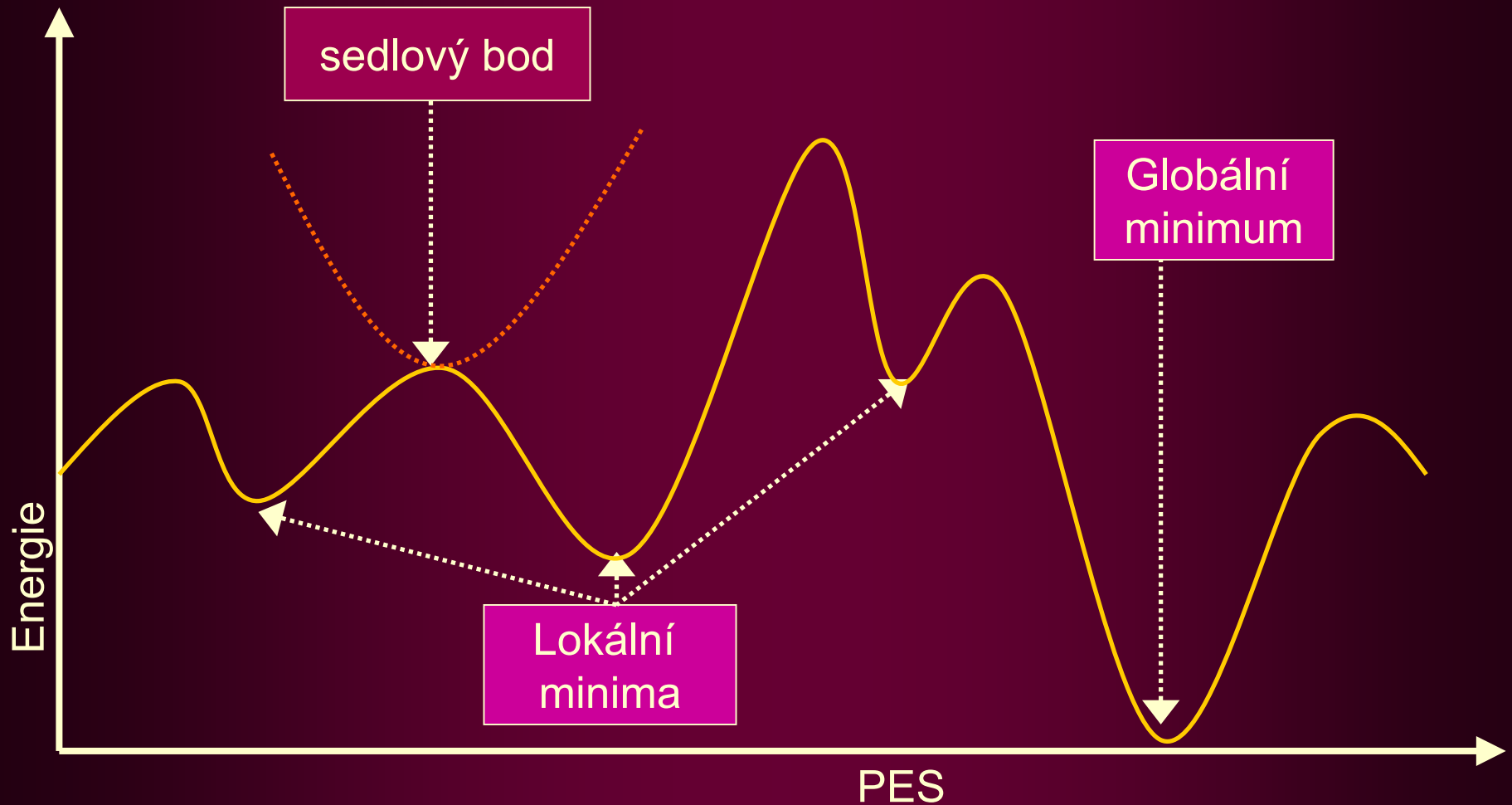
blízko minima

pro malé molekuly

nederivační:

- ⇒ simplexová metoda

Problém lokálních minim



Hledání minim na PES

⇒ **lokální metody** = minimalizace (hledání lokálních minim)

⇒ **globální metody** = hledání všech lokálních minim + globální minimum

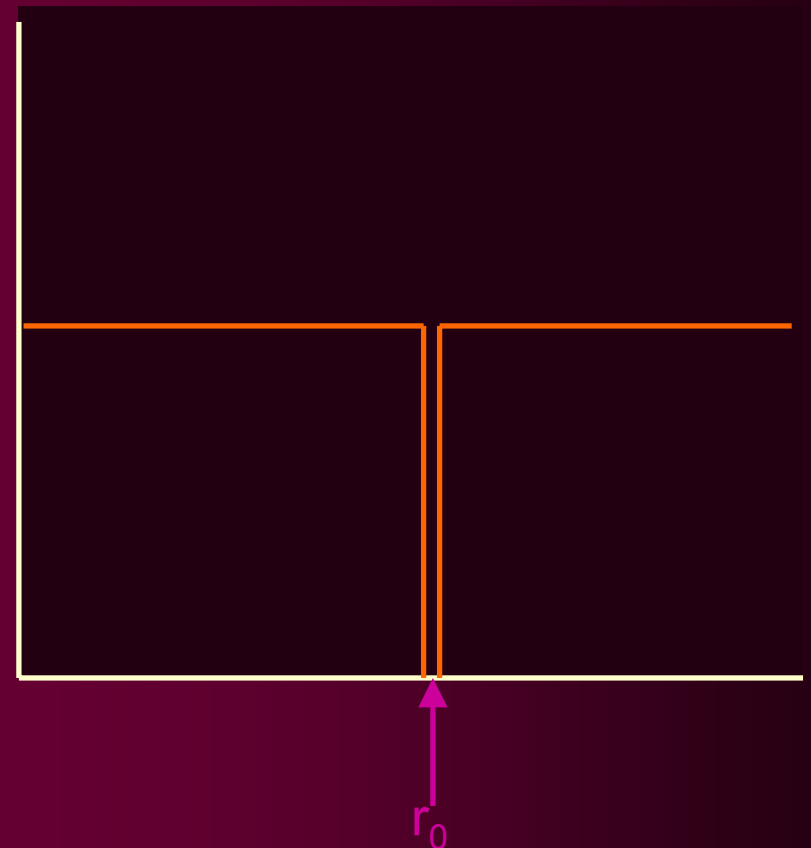
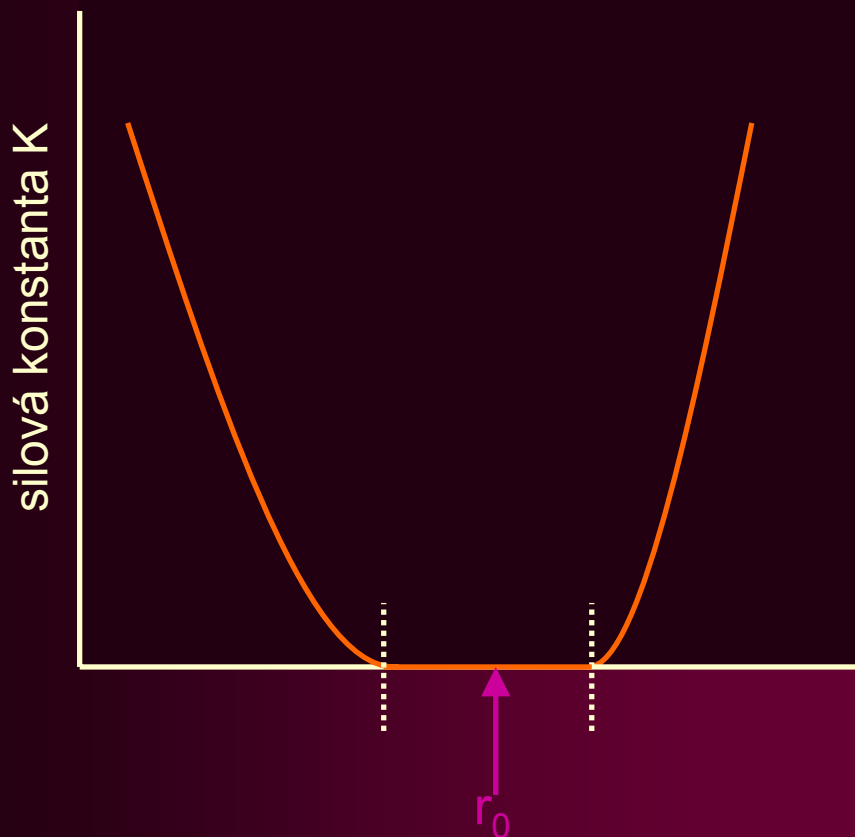
- systematické prohledávání (grid)
- MD a odvozené techniky (simulované žíhání, LES, FEP)
- Monte Carlo
- CICADA

Restraints X Constraints

(*restraint* = omezení, nátlak; *constraint* = donucení, přesné vymezení)

zvýšení silové konstanty K

⇒ restraints = měkké vazné podmínky ⇒ constraints = tvrdé vazné podmínky



Dlouhodosahové elektrostatické interakce

System N-částic:

nejnáročnější část minimalizace = výpočet
dlouhodosahových interakcí (Coulomb síly)

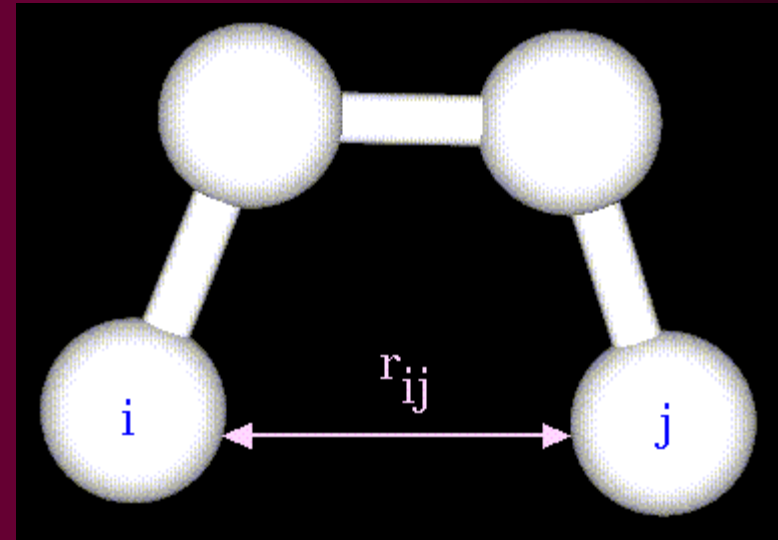
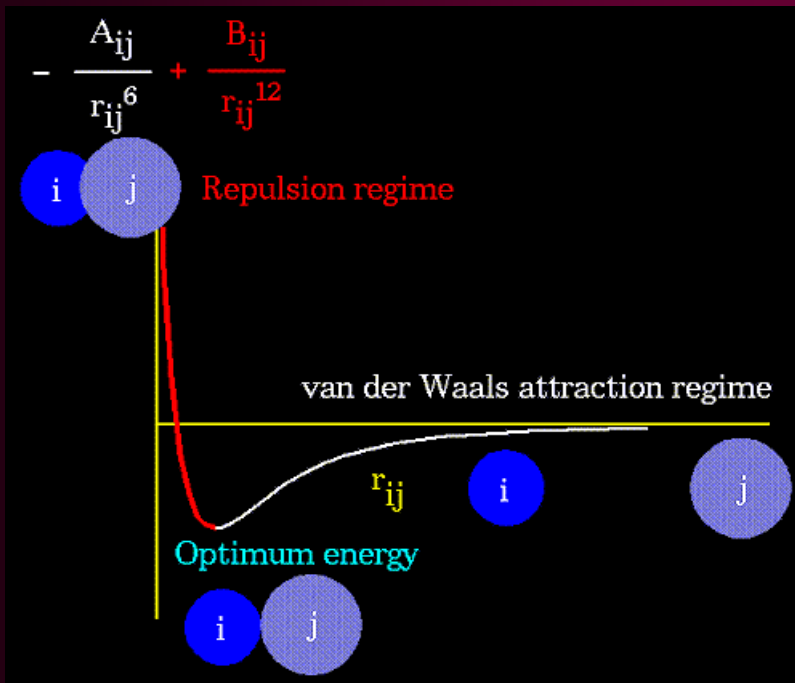
N^2 výpočtů

elektrostatická energie

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

Nevazebné interakce

vdW



elektrostat


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

Electrostatic term

Dielektrický model solventu

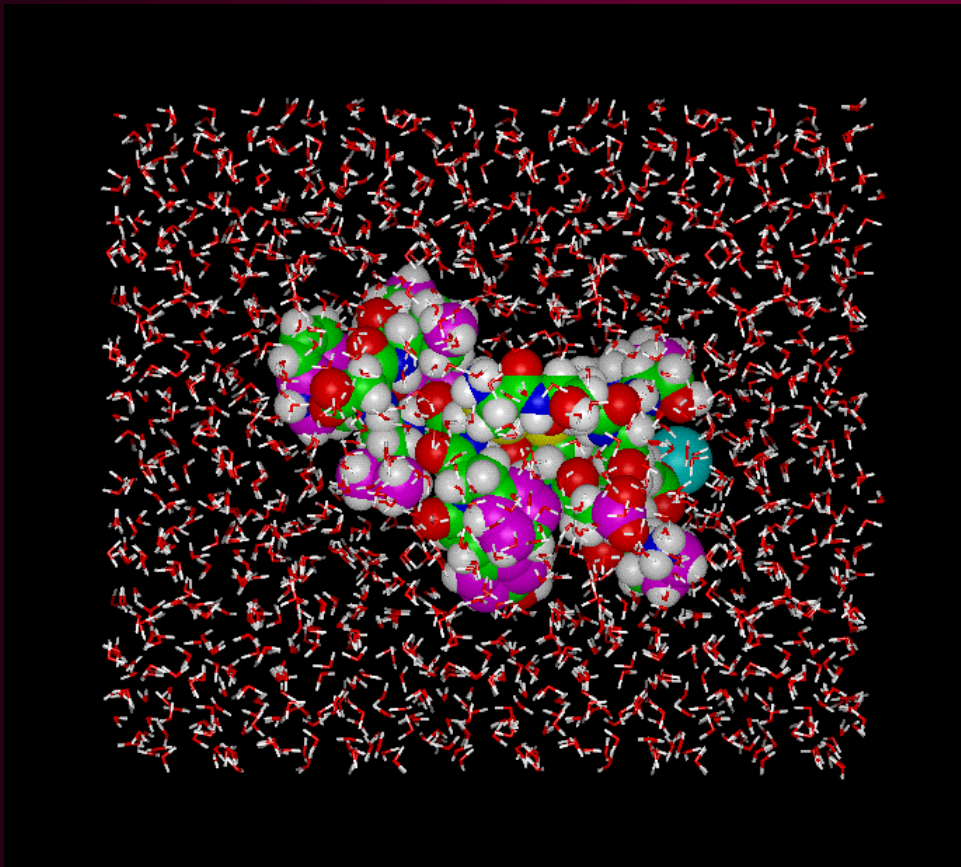
$$E(r) = \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r}$$

vakuum


$$\varepsilon = \varepsilon_0 \varepsilon_r$$

prostředí o relativní permitivitě ε_r

Explicitní solvent



☺ sledování specifických interakcí s molekulami H_2O

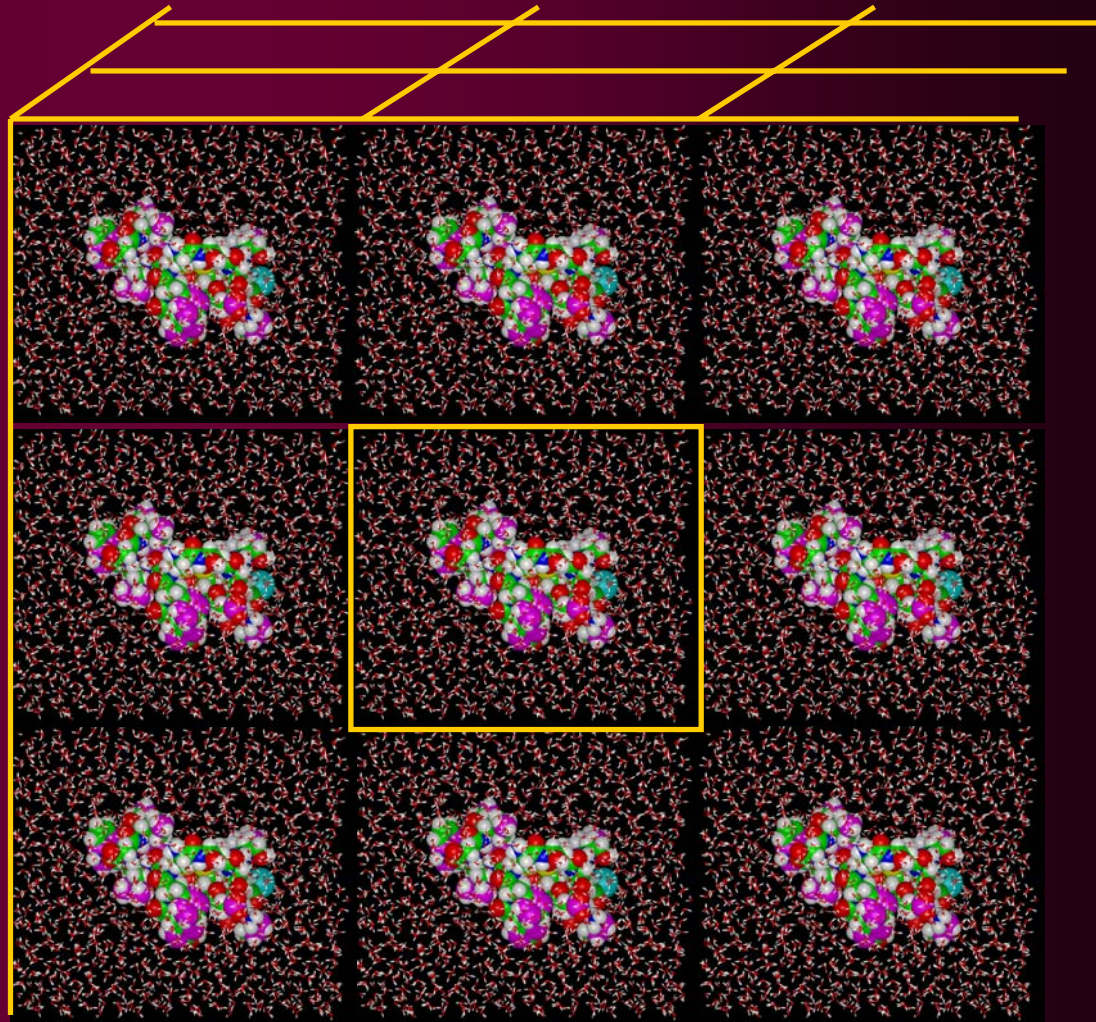
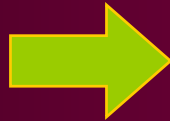
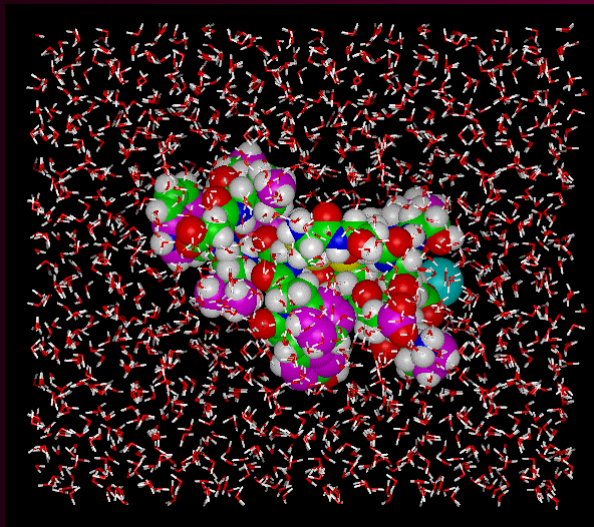
☹ značná výpočetní náročnost

peptid (330 atomů)

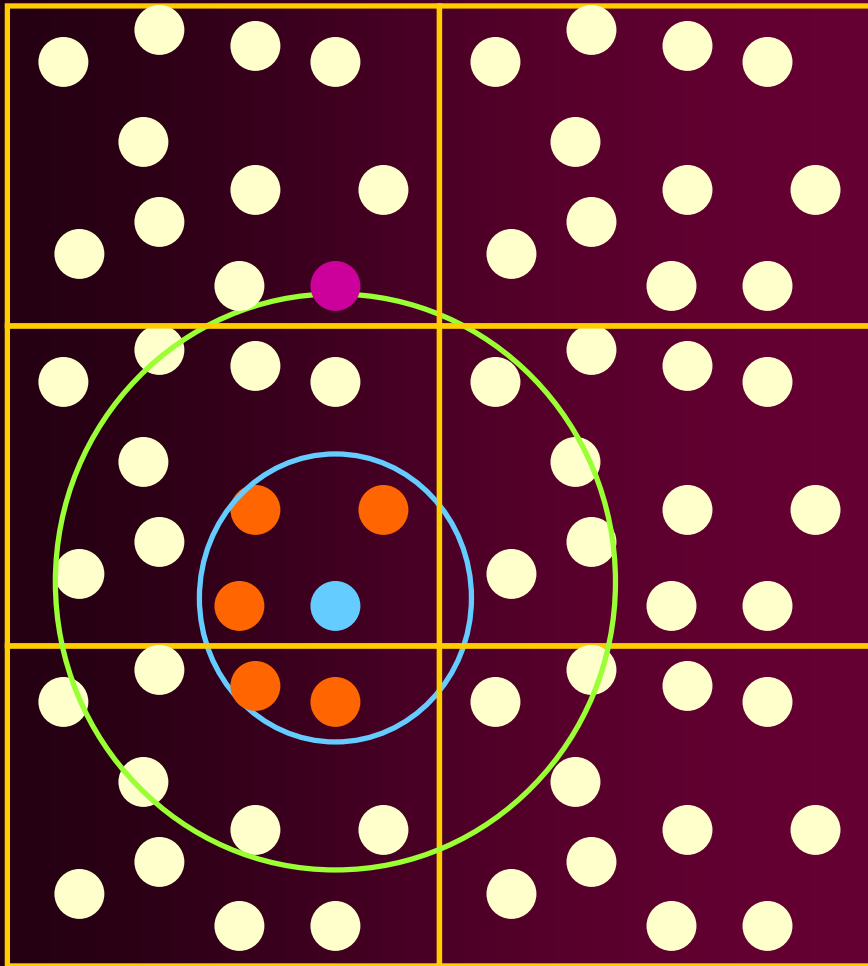
+ 2400 molekul H_2O

Hraniční podmínky

Periodic Boundary
Conditions (PBC)

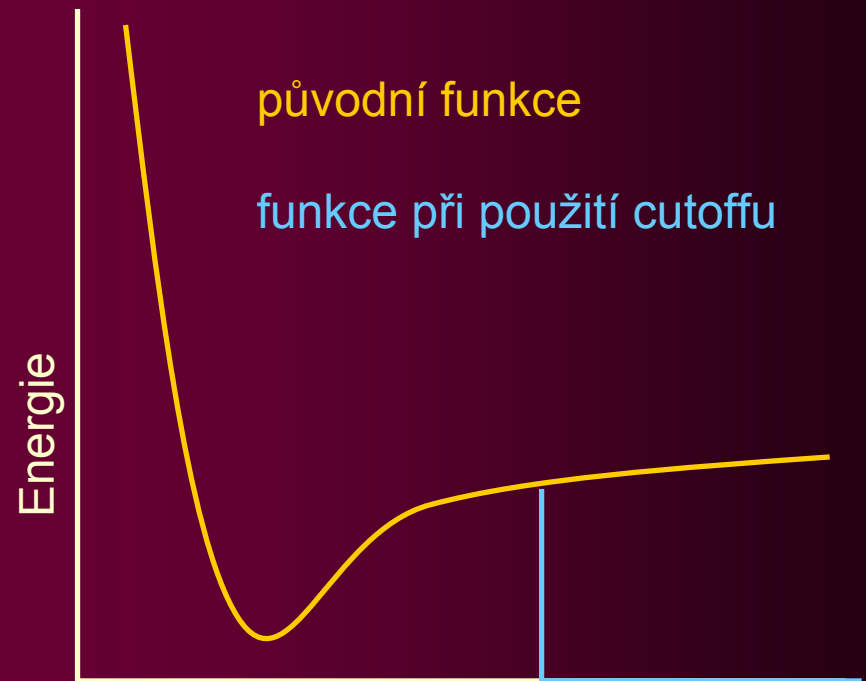


Výpočet elektrostatických interakcí v PBC



problém vyšších cutoffů

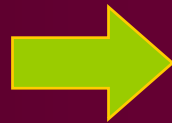
„odříznutí“ interakcí o vzdálenosti $>$ cutoff



vzdálenost

Ewaldova sumace

$$E(r) = \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r}$$



Periodický box

$$E(r) = \sum_{n=0}^{\infty} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r + n}$$

n ... násobek délky hrany boxu L

Ewald (1921):

reálný prostor



reciproký prostor

$$E = f(L)$$

$$E = f(1/L)$$

Particle Mesh

bodové náboje → model kontinuálního solventu

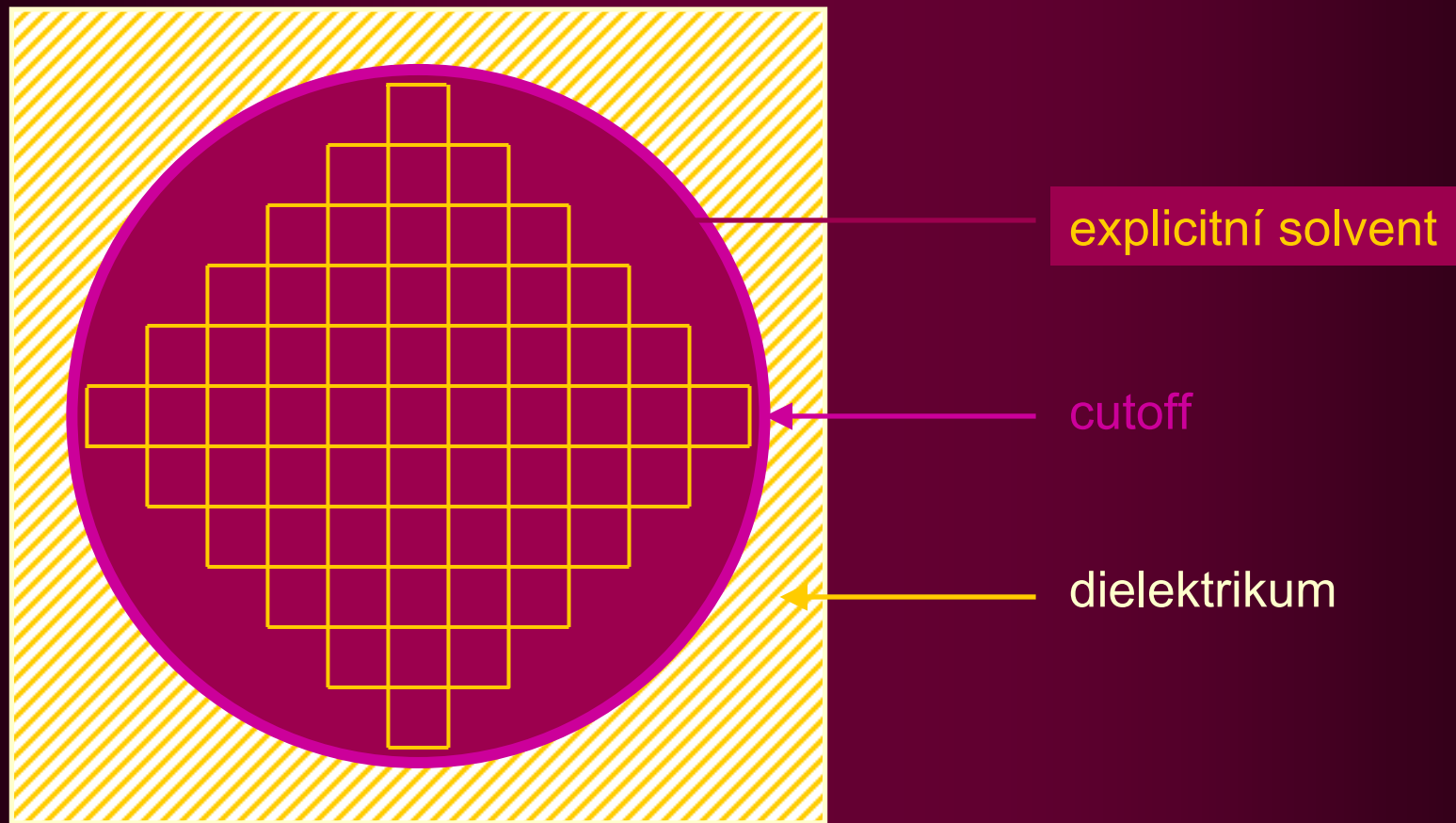
Poissonova rovnice:

$$\nabla^2 \phi = \rho(\mathbf{x})$$

elektrický potenciál nábojová hustota

- ⇒ umístění nábojů na mřížku
- ⇒ výpočet E pomocí P. rovnice
- ⇒ interpolace E mezi body mřížky

Particle Mesh Ewald (PME)

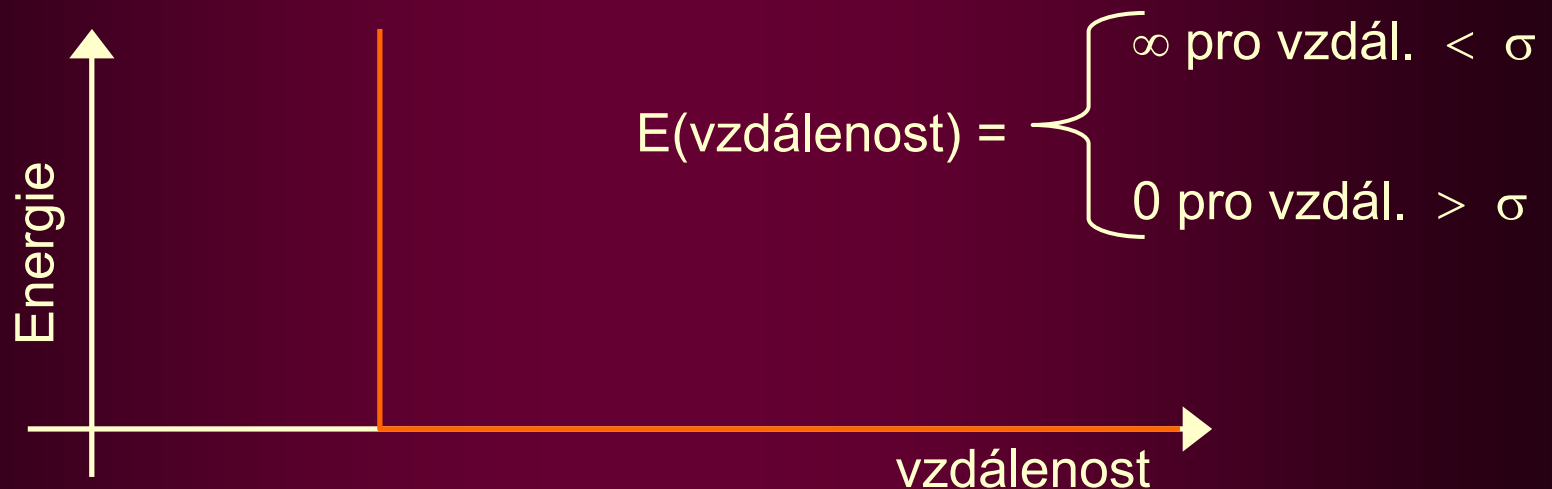


Molekulová dynamika

Historie molekulové dynamiky

1. MD simulace: 1957 „hard-sphere model“

kolize pevných částic pohybujících se konstantní rychlostí,
sledování srážek částic



Historie MD

1976 - „dark ages“

před. r. 1995:

MD simulace nukleových kyselin (vysoce nabité systémy, nutno správně pracovat s **dlohodosahovými elektrostatickými interakcemi**) - nestabilní (zkroucení duplexu, zlomení párů bazí).

Nedostatek výpočetní síly:

⇒ Pro výpočet dlouhodosahových elektrostatických interakcí je užíván cutoff - useknutí interakcí - nestabilita. Neužívá se PME (známo od r. 1993)

⇒ Pomalý vývoj silových polí - nemožnost kvalitních QM výpočtů

⇒ Krátké simulace $< 200\text{ps}$

Historie MD

1995 - 1998 - zvýšená počítačová síla (superpočítače)

⇒ korektní zacházení s dlouhodosahovými nevázebnými interakcemi (elektrostatika, vdW) - **PME**

⇒ kvalitní QM výpočty na vyšší úrovni ⇒ silové pole 2. generace - např. AMBER **parm94.dat** (W. Cornell) ⇒ stabilní trajektorie

⇒ **přesná interpretace krystalových struktur**

Př.:

MD simulace nukleových kyselin - DNA (duplex, ..., kvadruplex), RNA (hairpin, ribozymy), modifikovaná DNA, RNA, protein/NA interakce

Historie MD

1998 - přesná interpretace struktur + ΔG

Př.:

MD simulace nukleových kyselin + výpočty volné energie komplexů molekul.

Omezení:

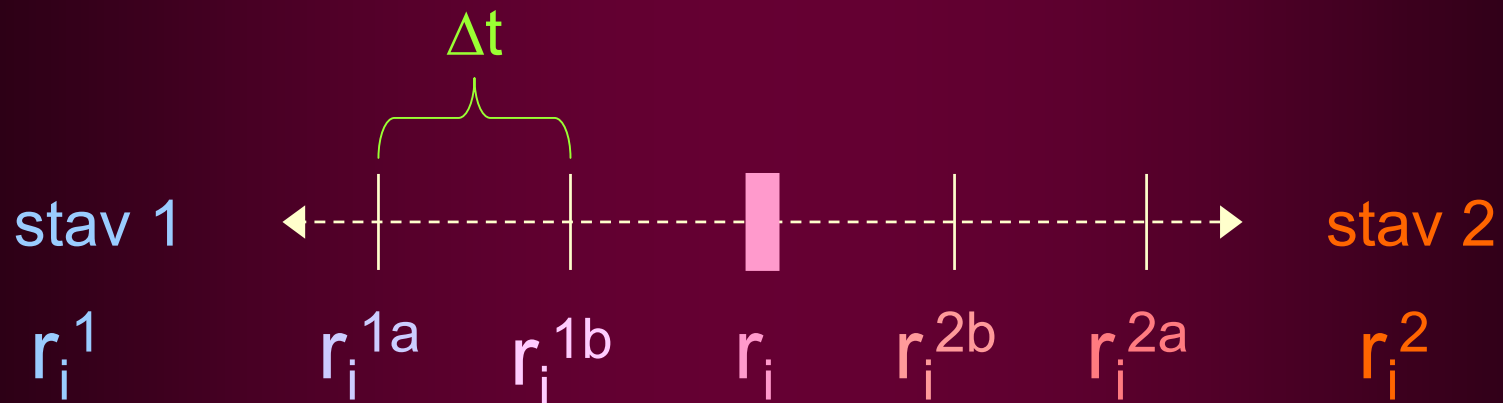
⇒ nedostatky silových polí

⇒ nezahrnutí polarizace

Molekulová dynamika

integrace Newtonových pohybových rovnic

$$F_i = m_i a_i$$



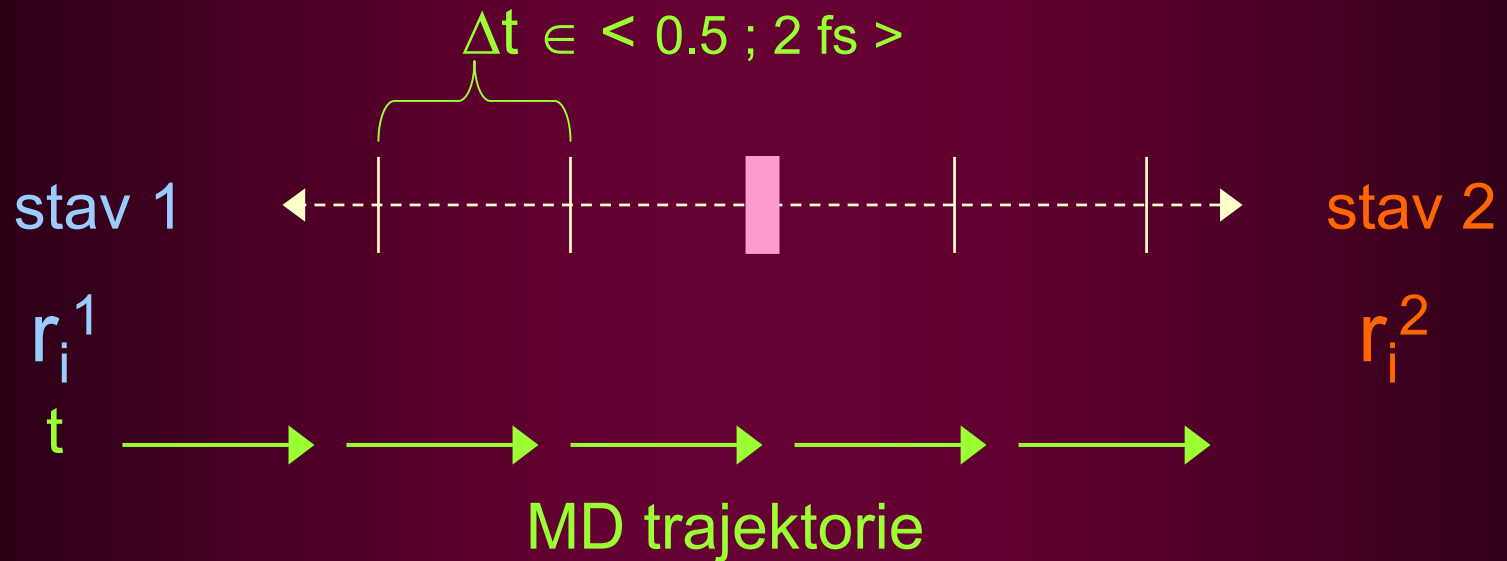
Termodynamika: jaké stavy jsou možné?

Kinetika: jak (a jak rychle) stavy interkonvertují?

Molekulová dynamika

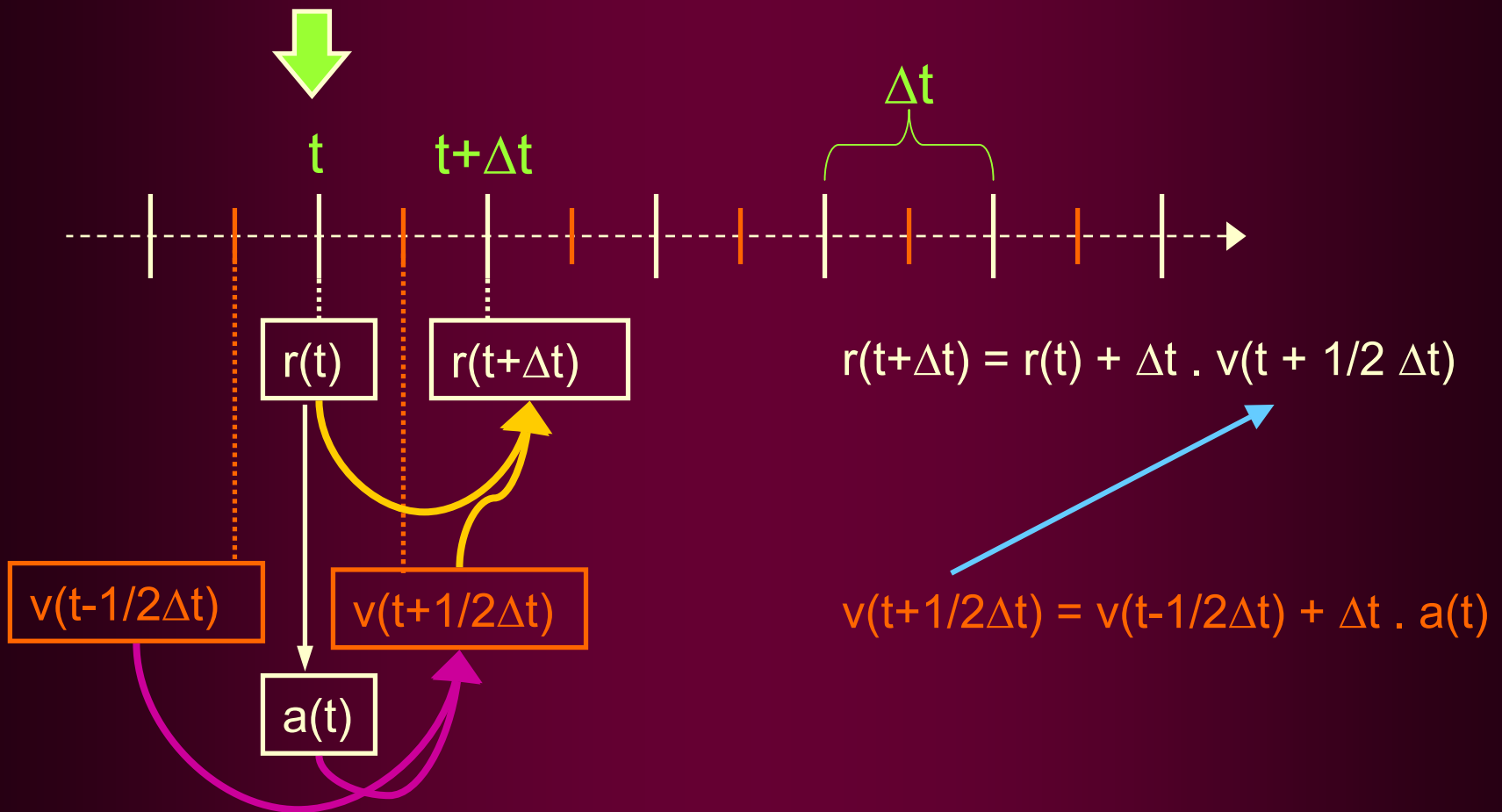
$$-\frac{dE}{dr_i} \quad \xrightarrow{\quad} \quad F_i = m_i a_i \quad \xrightarrow{\quad} \quad \frac{dv_i}{dt} \quad \xleftarrow{\quad} \quad \frac{dr_i}{dt}$$
$$-\frac{dE}{dr_i} = m_i \frac{d^2 r_i}{dt^2}$$

Molekulová dynamika



Molekulová dynamika

leap-frog algoritmus



Časový krok v MD

⇒ příliš malý - MD pokrývá jen omezenou část konformačního prostoru

⇒ příliš velký - nestabilita trajektorie, vysoká E

Správný časový krok Δt :

1/10 nejkratšího vibračního pohybu

Př.:

C–H vazba vibruje s periodou 10fs \Rightarrow integrační krok $\Delta t = 1$ fs

Př.:

constraints na vazby obsahující H (SHAKE algoritmus) \Rightarrow
integrační krok $\Delta t = 2$ fs

Počáteční konfigurace pro MD

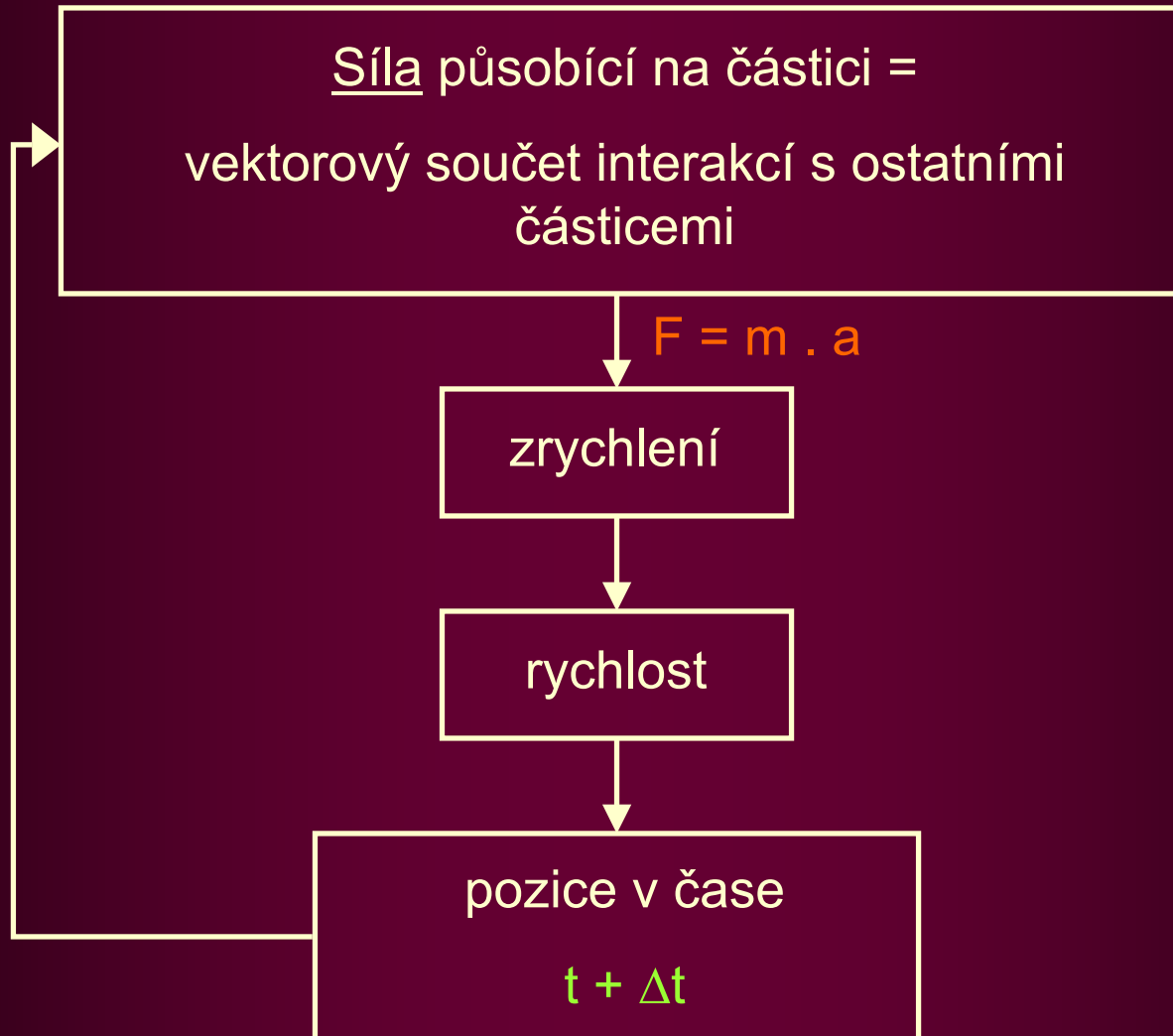
- ⇒ z experimentu (strukturní databáze...)
- ⇒ úprava (dostavění) struktury
- ⇒ přiřazení počátečních rychlostí - náhodně podle **Maxwell - Boltzman distribuce** (Gaussova distribuce)
- ⇒ relaxace dostavěných částí

Generátor (pseudo)náhodných čísel:

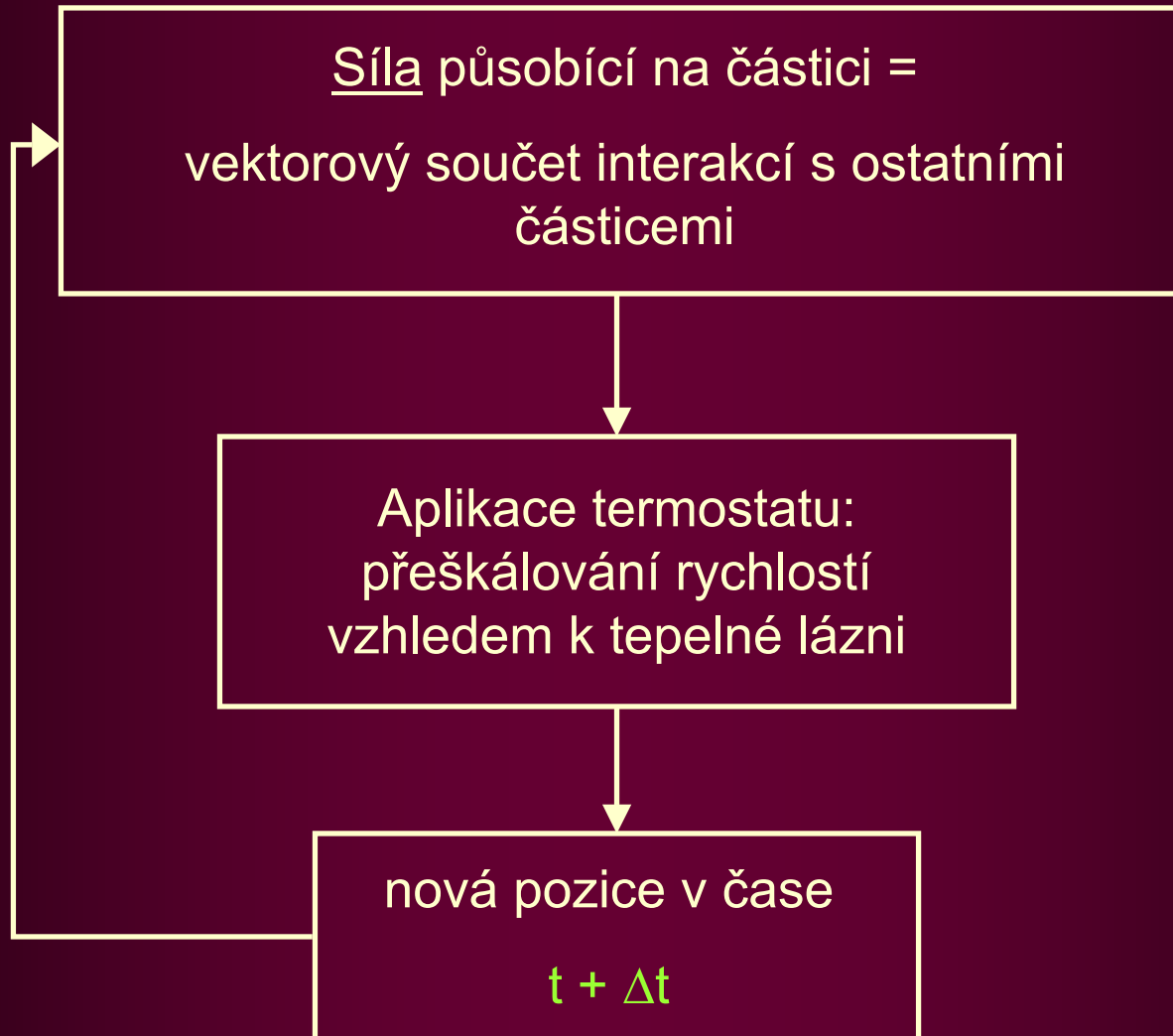
produkuje čísla $\in \langle 0, 1 \rangle$ rovnoměrně

přepočítá na čísla odpovídající gaussovské distribuci

Algoritmus MD



Algoritmus MD

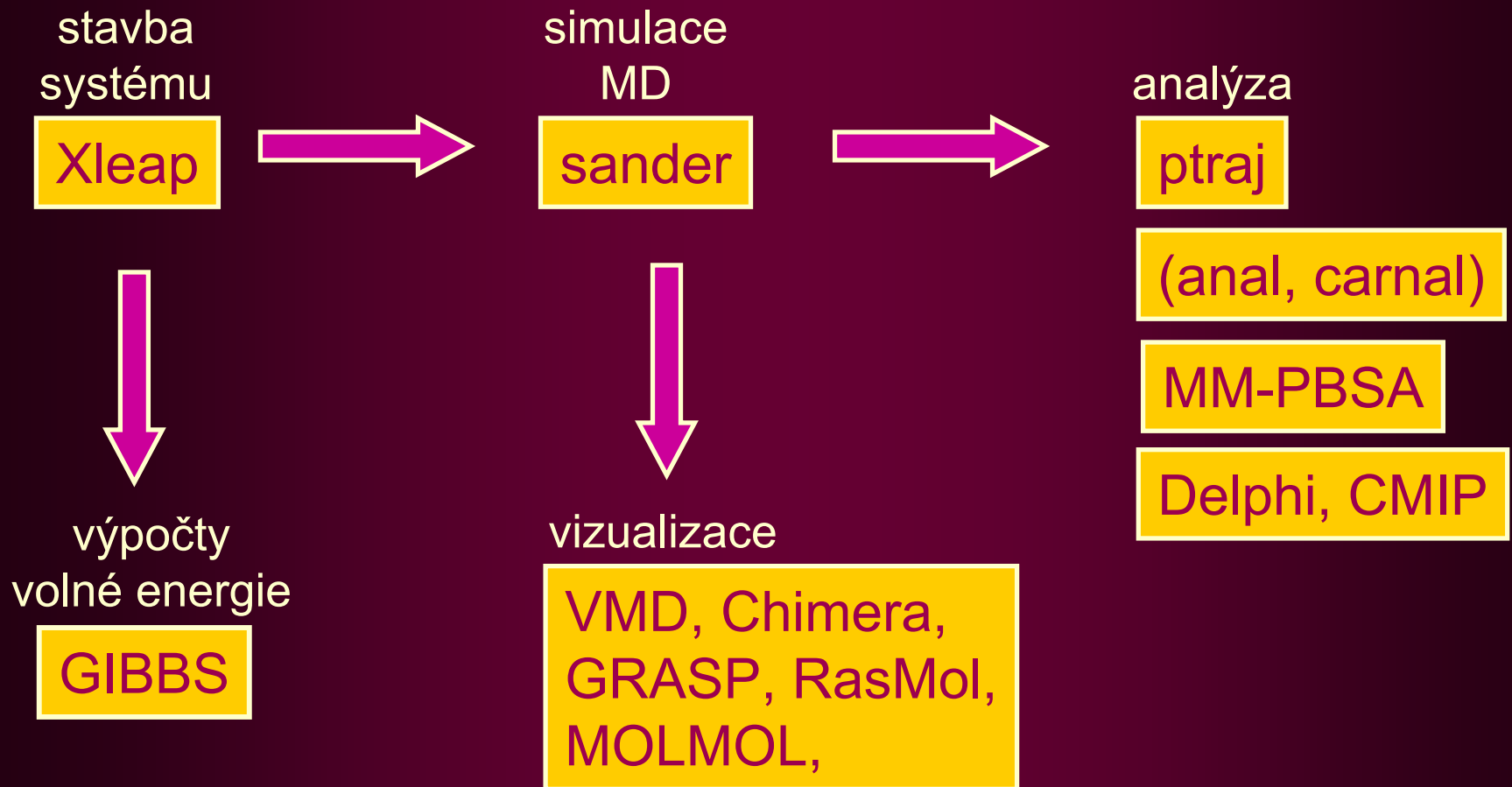


Termostat

napodobuje působení teploty na systém - působení makroskopických vibračních modů na dynamiku jednotlivých atomů

= násobení každé rychlosti atomu faktorem, který tvoří požadovanou teplotu (každý krok nebo periodicky každých X kroků)

Moduly AMBERu



Příprava systému - XLeap

*PDB souřadnice
struktury
- soubor.pdb*

ATOM	1	H5T	DG5	1	2.681	-24.540	4.934
ATOM	2	O5'	DG5	1	2.553	-24.602	5.052
ATOM	3	C5'	DG5	1	2.065	-24.636	5.468
ATOM	4	1H5'	DG5	1	1.667	-24.702	5.262
ATOM	5	2H5'	DG5	1	2.122	-25.015	5.496
ATOM	6	C4'	DG5	1	1.977	-24.087	6.164
ATOM	7	H4'	DG5	1	1.502	-24.376	6.678
ATOM	8	O4'	DG5	1	3.161	-23.900	6.288
ATOM	9	C1'	DG5	1	3.252	-23.078	6.426
ATOM	10	H1'	DG5	1	2.941	-22.964	7.057

Přidání solventu
Přidání iontů
...



databáze
parametrů

souřadnice -

topologie + PARAMETRY

soubor.crd

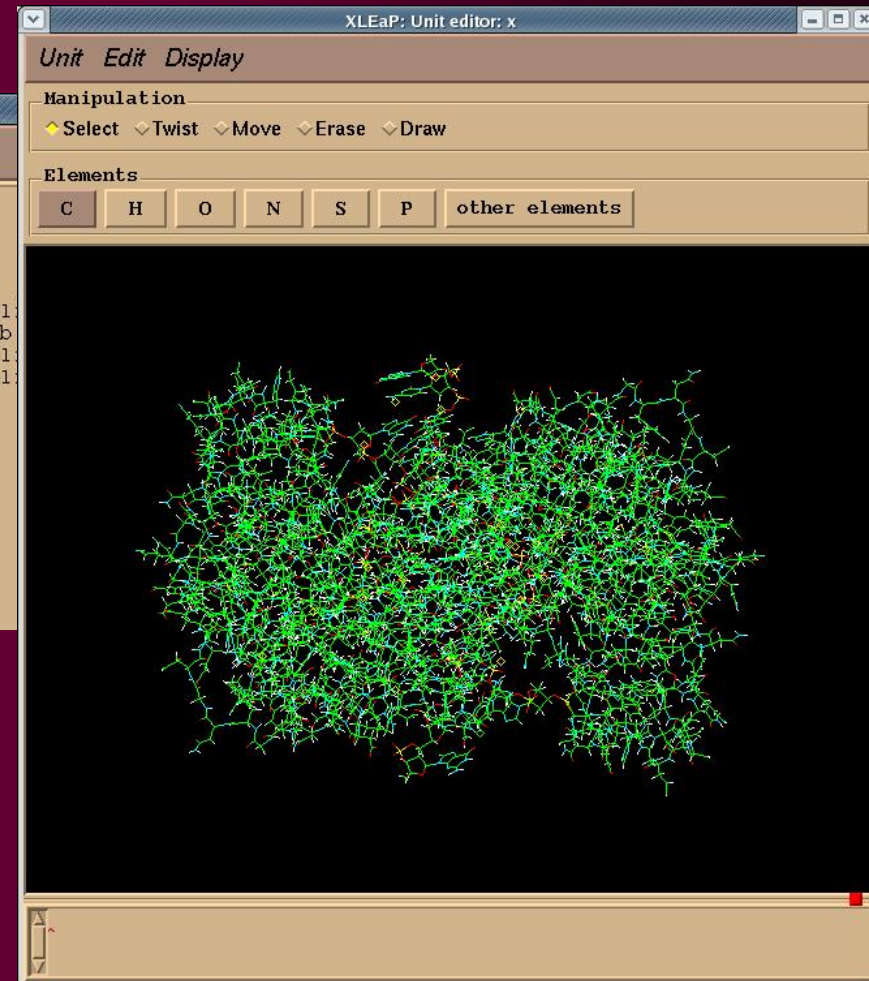
soubor.top

sander

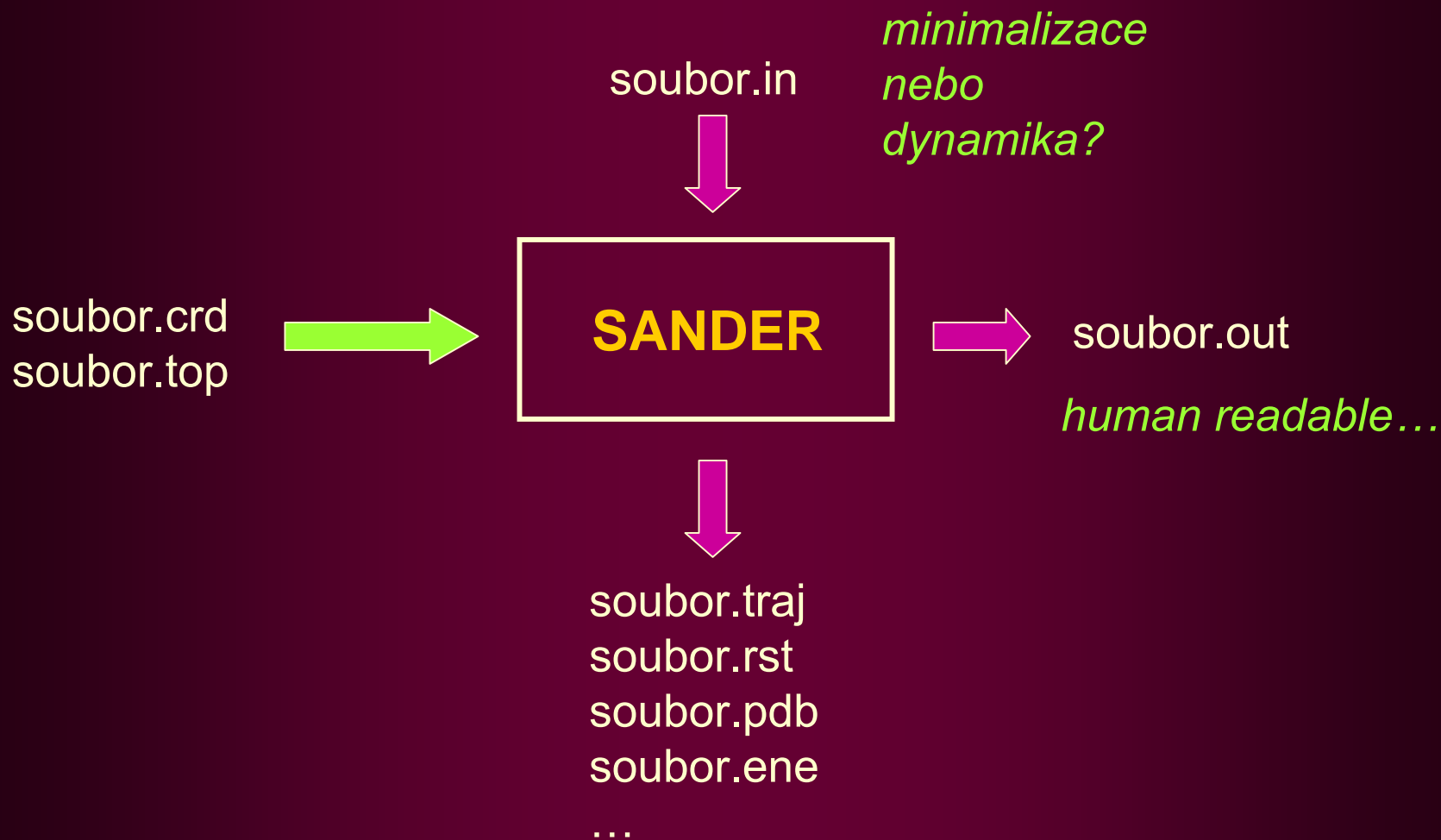
Příprava systému - XLeap

```
XLEaP: Universe Editor
File Edit Verbosity

Welcome to LEaP!
Sourcing leaprc: /home/evaf/soft/amber-7.0//dat/leap/cmd/leaprc
Log file: ./leap.log
Loading parameters: /home/evaf/soft/amber-7.0//dat/leap/parm/parm94.dat
Loading library: /home/evaf/soft/amber-7.0//dat/leap/lib/all_nucleic94.lib
Loading library: /home/evaf/soft/amber-7.0//dat/leap/lib/all_amino94.lib
Loading library: /home/evaf/soft/amber-7.0//dat/leap/lib/all_amnobct94.lib
Loading library: /home/evaf/soft/amber-7.0//dat/leap/lib/all_aminont94.lib
Loading library: /home/evaf/soft/amber-7.0//dat/leap/lib/ions94.lib
Loading library: /home/evaf/soft/amber-7.0//dat/leap/lib/solvents.lib
> x = loadpdb average.pdb
Loading PDB file: ./average.pdb
Unknown residue: MG number: 539 type: Nonterminal
Unknown residue: MG number: 540 type: Nonterminal
Unknown residue: MG number: 541 type: Nonterminal
One sided connection. Residue: missing connect1 atom.
Created a new atom named: H1 within residue: .R<SER 27>
Created a new atom named: H2 within residue: .R<SER 27>
Created a new atom named: H3 within residue: .R<SER 27>
```



SANDER modul



Vstup/výstupní soubory

```
# 3ps restrained MD on water and non-crystal CIO
&cntrl  imin=0, ntx=1, irect=0, ntrx=1,ntxo=1, ntpr=100,
        ntwx=500, ntwv=0, ntwe=0,
        ioutfm=0, ntwprt=0, ntf=2, ntb=2, igb=0, dielc=1.0, cut=8.0,
        nsnb=10, scnb=2.0, scee=1.2,
        ibelly=0, ntr=0, nstlim=1500, nscm=0,
        t=0.0, dt=0.002, temp0=298.16, tempi= 298.16, ig=555333,
        heat=0.0001, ntt=1, dtemp=10.0, tautp=0.2,
        vlimit=20.0, ntp=1, pres0=1.0,
        comp=44.6, taup=0.2, ntc=2,  tol=0.00001,|
        jfastw=0,
        lastrst=4000000, lastist=4000000,

&end
&ewald
  ew_type = 0, skinnb = 1.0, nbflag = 1, use_pme = 1, vdwmeth = 1,
&end
Restraining on protein + CRYSTAL Na+
1000.0
RES 1 544
END
END
```

vstupni_soubor.in

vystupni_soubor.out

```
NSTEP =      0 TIME(PS) =      0.000  TEMP(K) =      0.00  PRESS = -4849.7
Etot   = -172156.5898  Ektot   =      0.0006  Eptot   = -172156.5903
BOND   =      217.7758  ANGLE   =     1348.2458  DIHED   =     5226.0102
1-4 NB =     2918.8378  1-4 EEL =    21766.8592  VDWAALS =     9145.0052
EELEC  = -212779.3244  EHBOND  =      0.0000  RESTRAINT =      0.0000
EKCMT  =      0.0002  VIRIAL  =     80090.1467  VOLUME  =    764872.0157
                                           Density  =      0.8404

Ewald error estimate:  0.7674E-03
```

```
NSTEP =     100 TIME(PS) =      0.200  TEMP(K) =    253.81  PRESS =  -143.4
Etot   = -161992.8872  Ektot   =    32659.3466  Eptot   = -194652.2338
BOND   =     855.4771  ANGLE   =     3002.3136  DIHED   =     5794.1733
1-4 NB =     2230.5528  1-4 EEL =    20583.4261  VDWAALS =    23912.6425
EELEC  = -251030.8193  EHBOND  =      0.0000  RESTRAINT =      0.0000
EKCMT  =     15720.5808  VIRIAL  =    17945.3438  VOLUME  =   718468.6039
                                           Density  =      0.8947

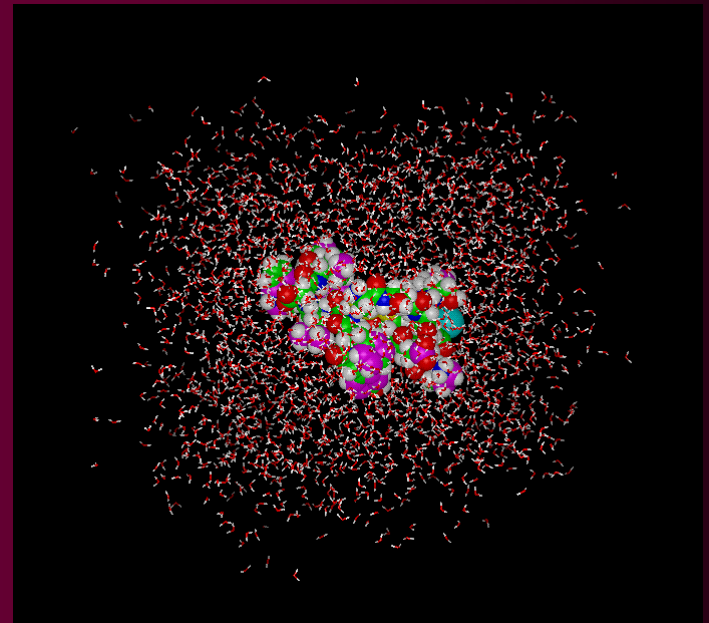
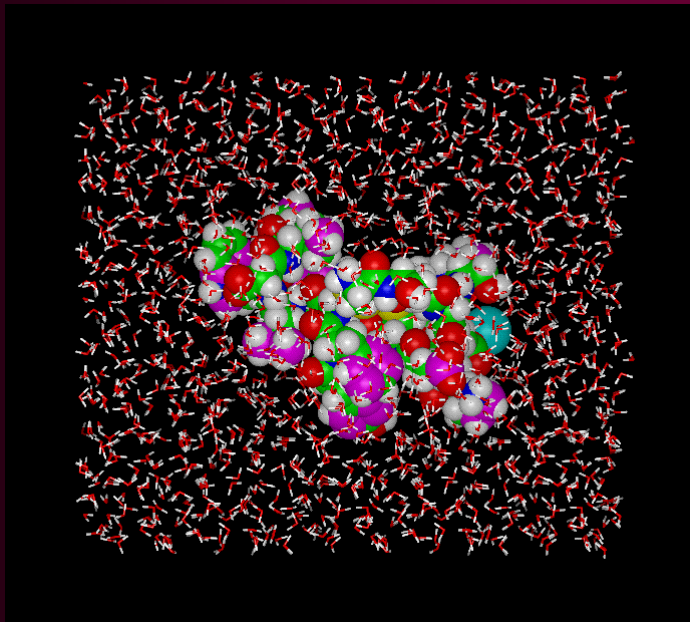
Ewald error estimate:  0.2293E-04
```

Postup MD

- volba vstupní struktury (např. z RTG), úprava (doplnění chybějících residuí, vodíků) - **Xleap**
- optimalizace doplněných částí - **sander**
- přidání iontů (neutralizace systému), solvatace - přidání molekul vody (solvatační box) - **Xleap**
- ekvilibrace systému - minimalizace vody a iontů - **sander**
- krátká MD vody a iontů - **sander**
- postupné uvolňování proteinu (= snižování restrains) - **sander**
- zahřívání na simulační teplotu - **sander**
- produkční MD - **sander**

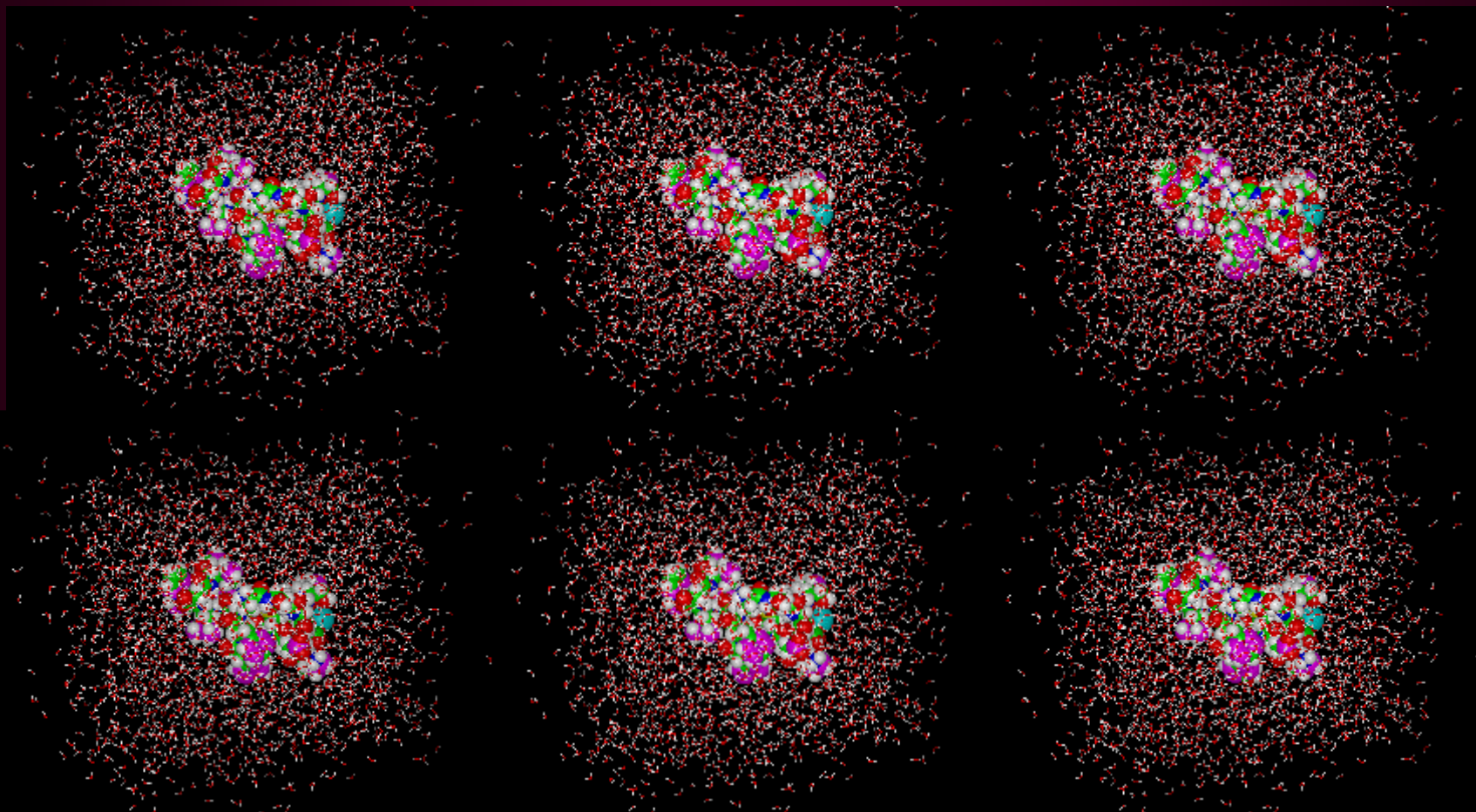
Ekvilibrace systému

⇒ Minimalizace a dynamika solventu a iontů
constraints na molekulu peptidu



⇒ Několikastupňová minimalizace s postupným uvolňováním peptidu

restraints na molekulu peptidu (25 kcal/mol/Å, 20, 15, 10, 5, 0)

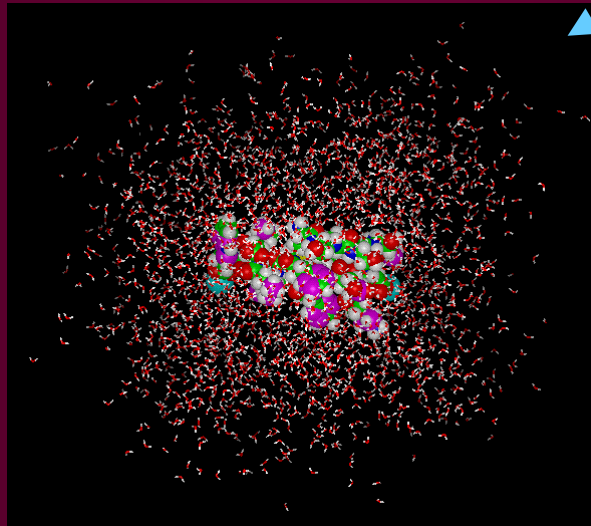


Zahřívání systému

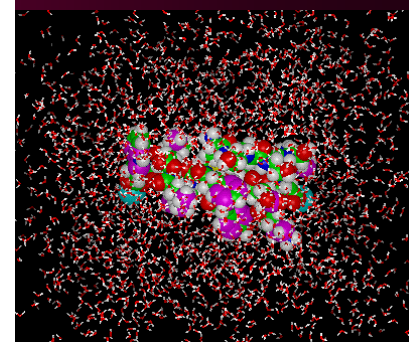
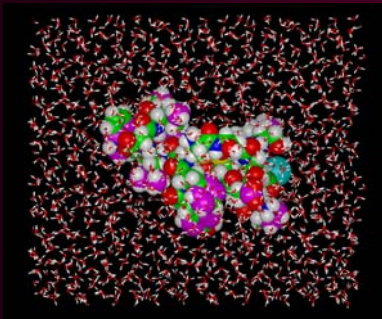
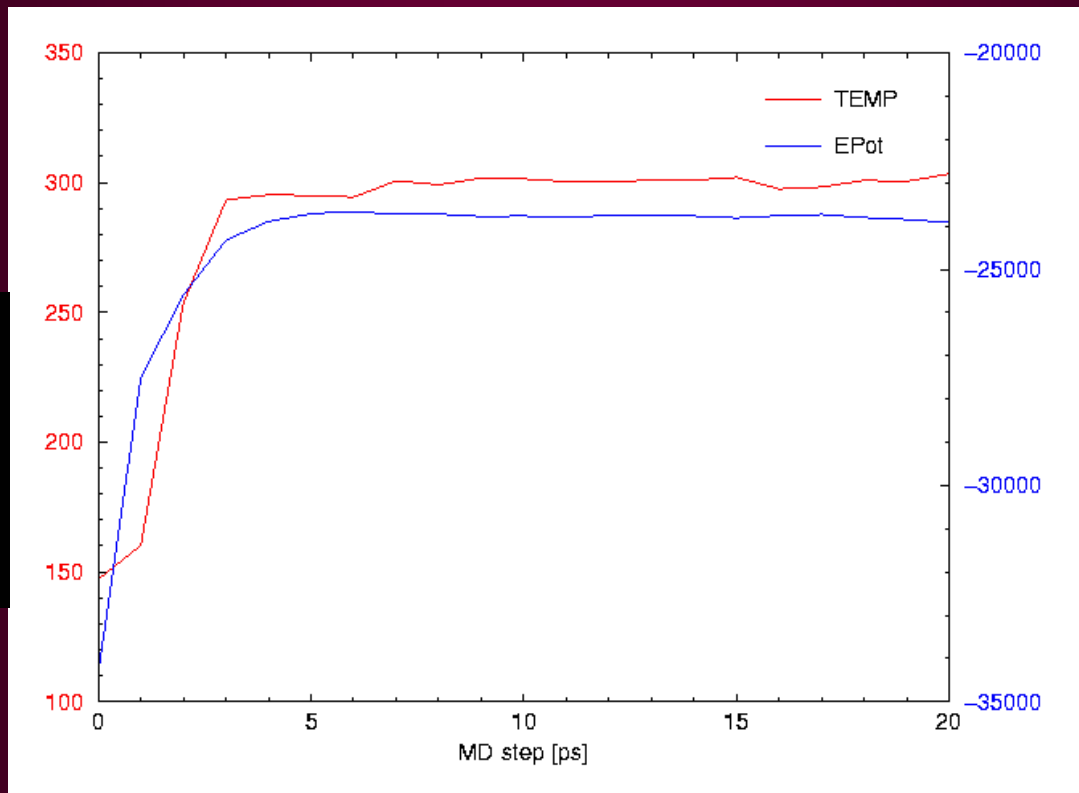


Zahřívání

100 K → 300 K
během 20 ps



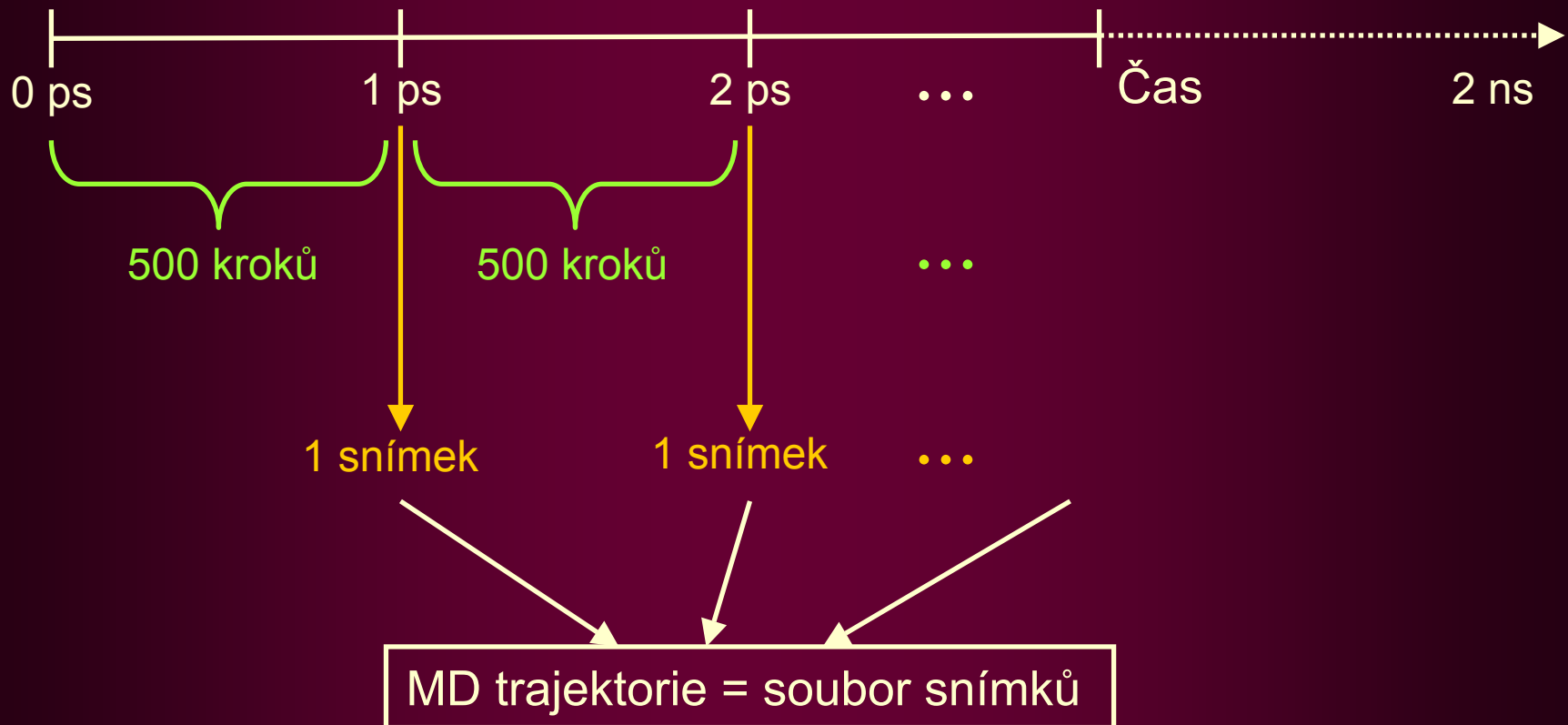
Kontrola ekvilibrace



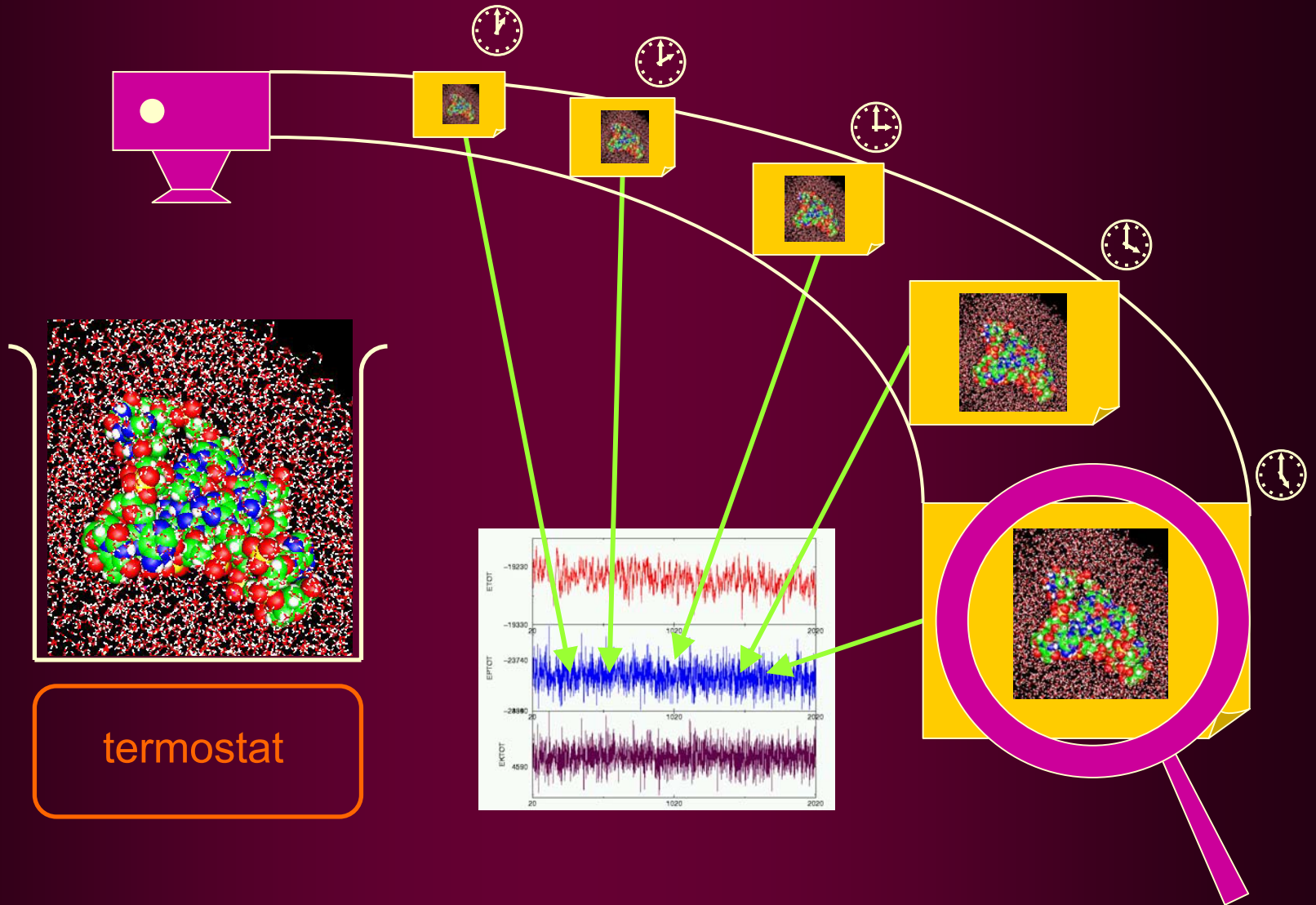
Produkční fáze MD

1 krok MD = 0.002 ps = 2 fs

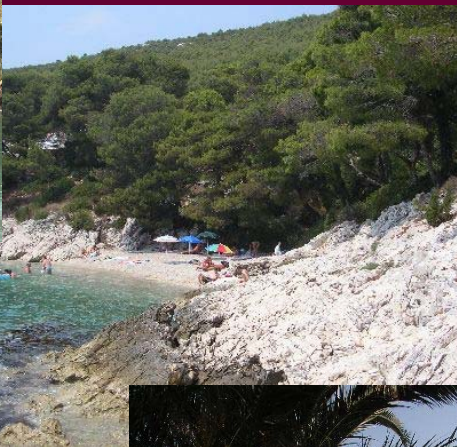
Frekvence ukládání snímků
= 500 kroků



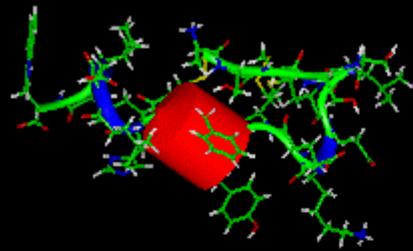
Molekulová dynamika



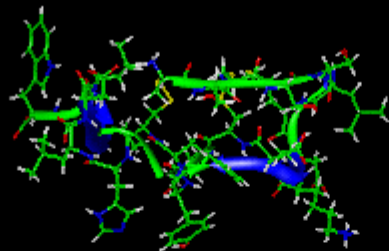
Výpočet MD



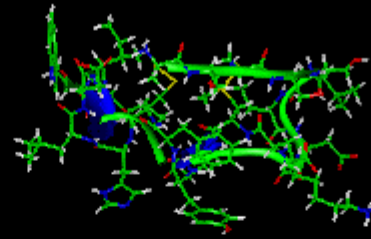
Trajektorie - snímky



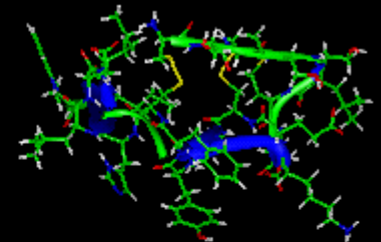
0 ps



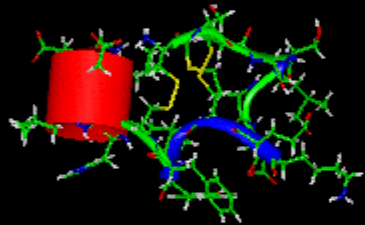
200 ps



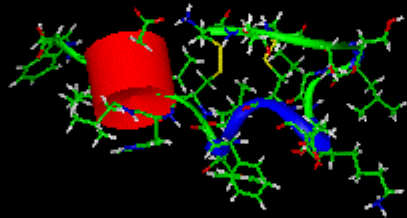
400 ps



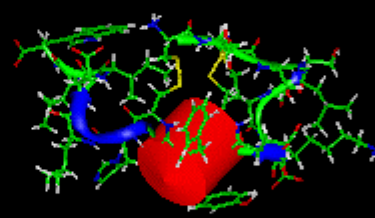
600 ps



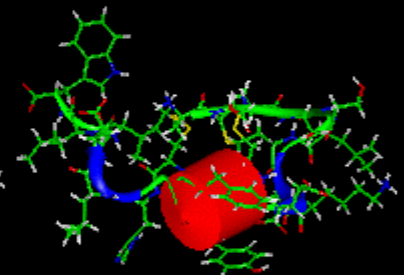
800 ps



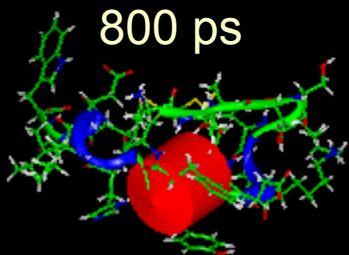
1000 ps



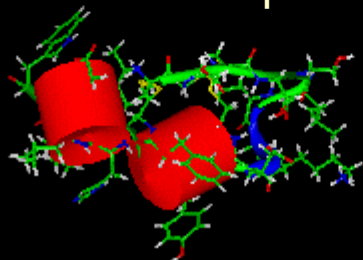
1200 ps



1400 ps

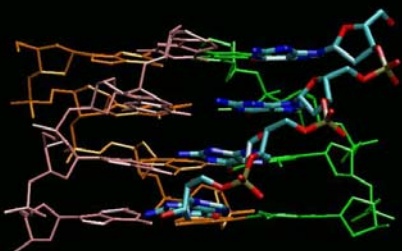
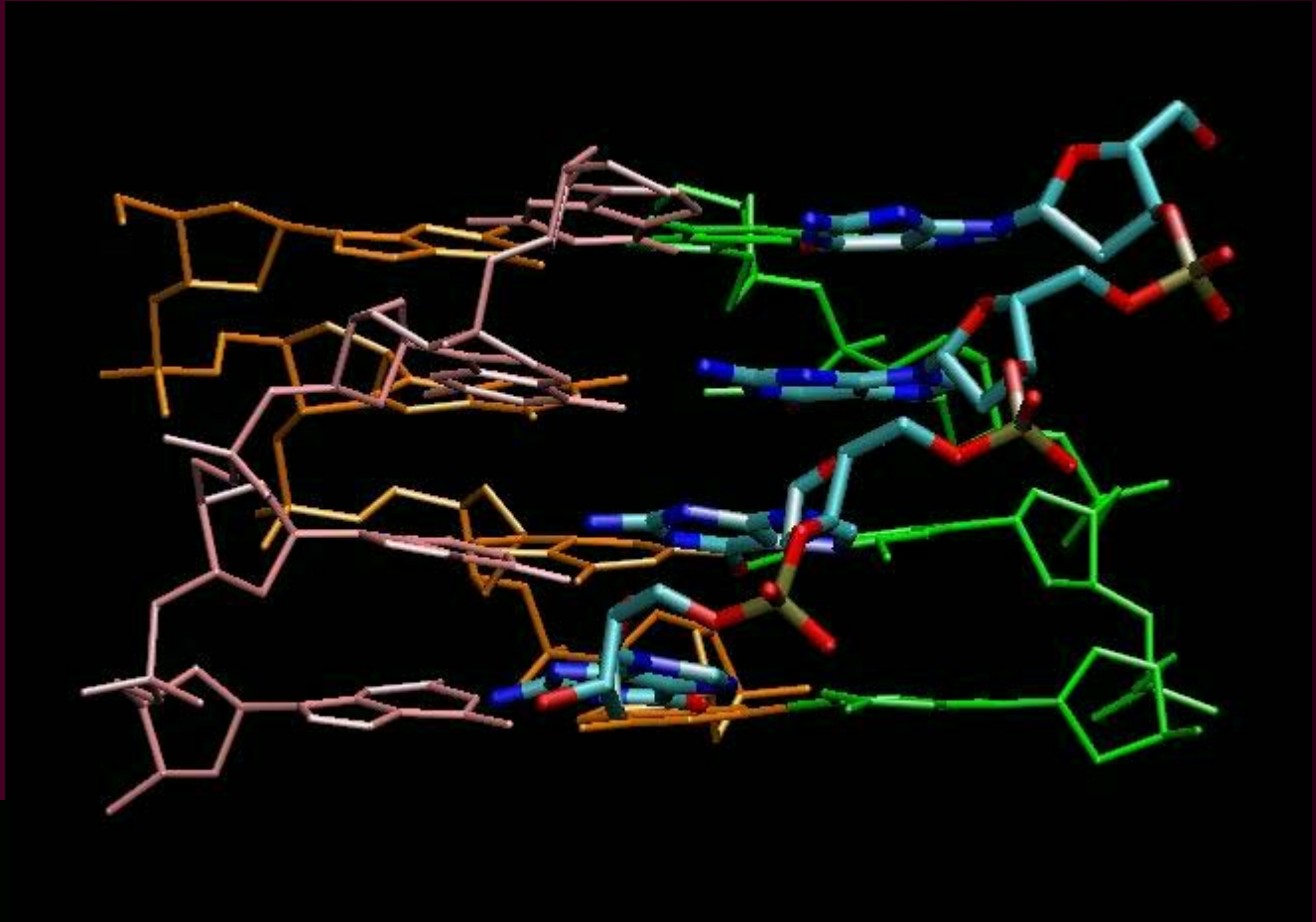


1600 ps



1800 ps

MD trajektorie



Analýza - modul PTRAJ

autor T. E. Cheatham III.

soubor.in

```
trajin strip_ns1-ns10.traj
trajout strip_ns1-ns10_centered.traj nobox
center :1-542 mass origin
image origin center
rms first mass out rms_ns1-10_centered :1-542
go
```

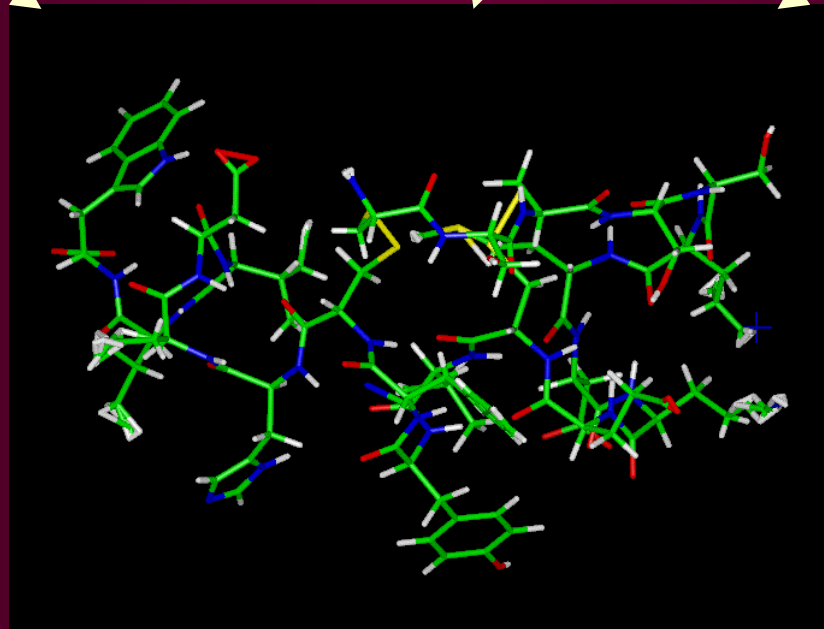
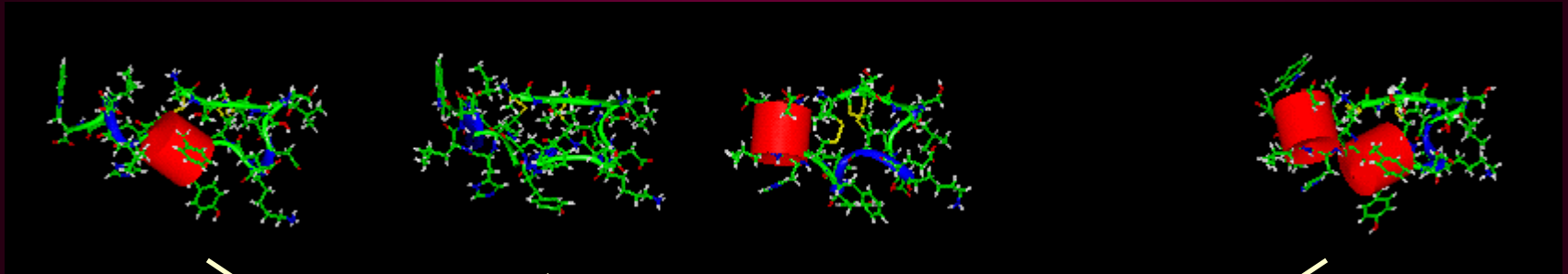
trajektorie = soubor snímků
soubor.top

PTRAJ

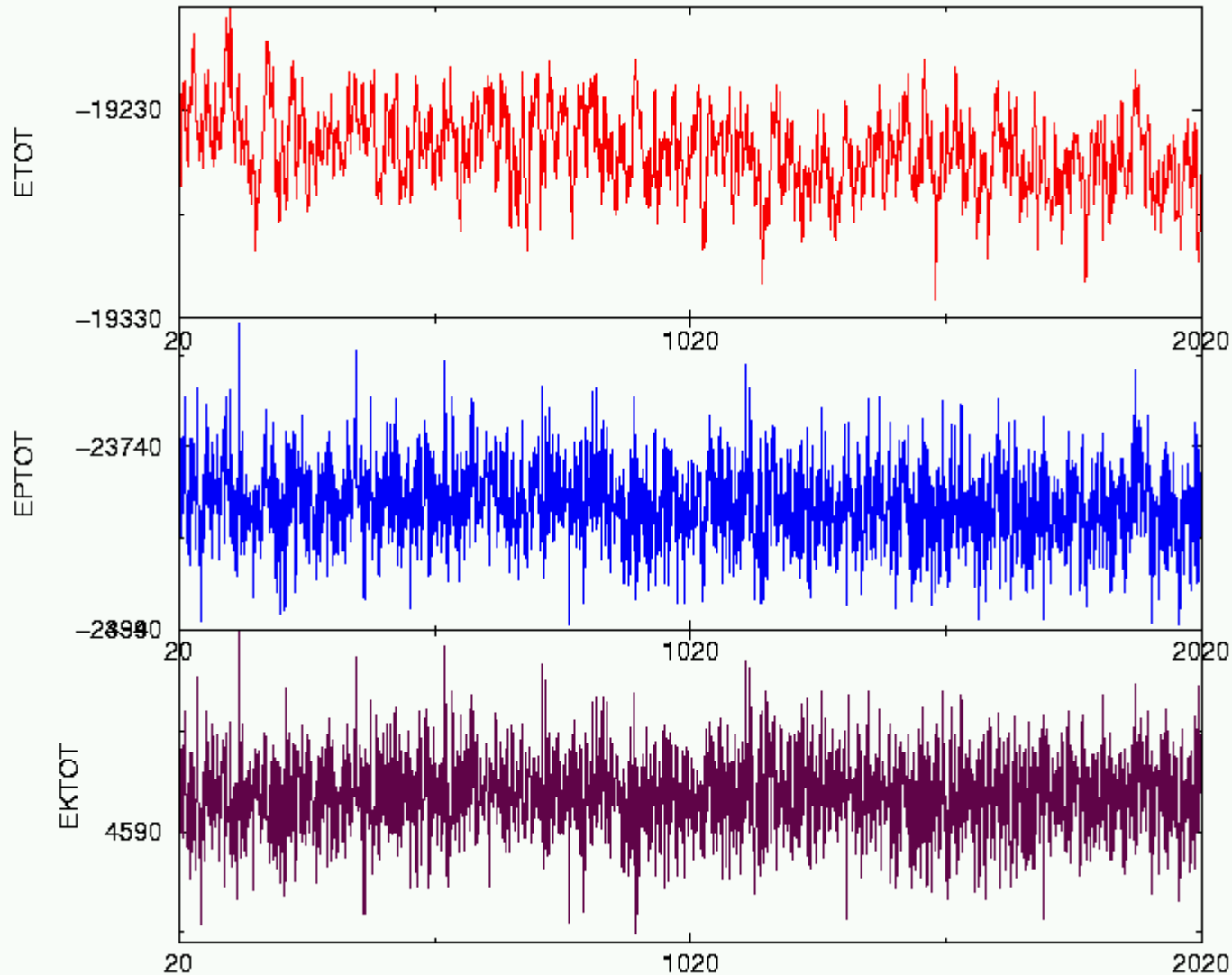
- modifikovaná trajektorie
- (imaging, center, strip, closest,...)
- jednotl. snímky
- vzdálenosti, úhly, torze, rmsd
- gridy, b-faktory
- ...

Průměrná struktura

z poslední fáze MD (např. 0.5 ns)



Analýza trajektorie

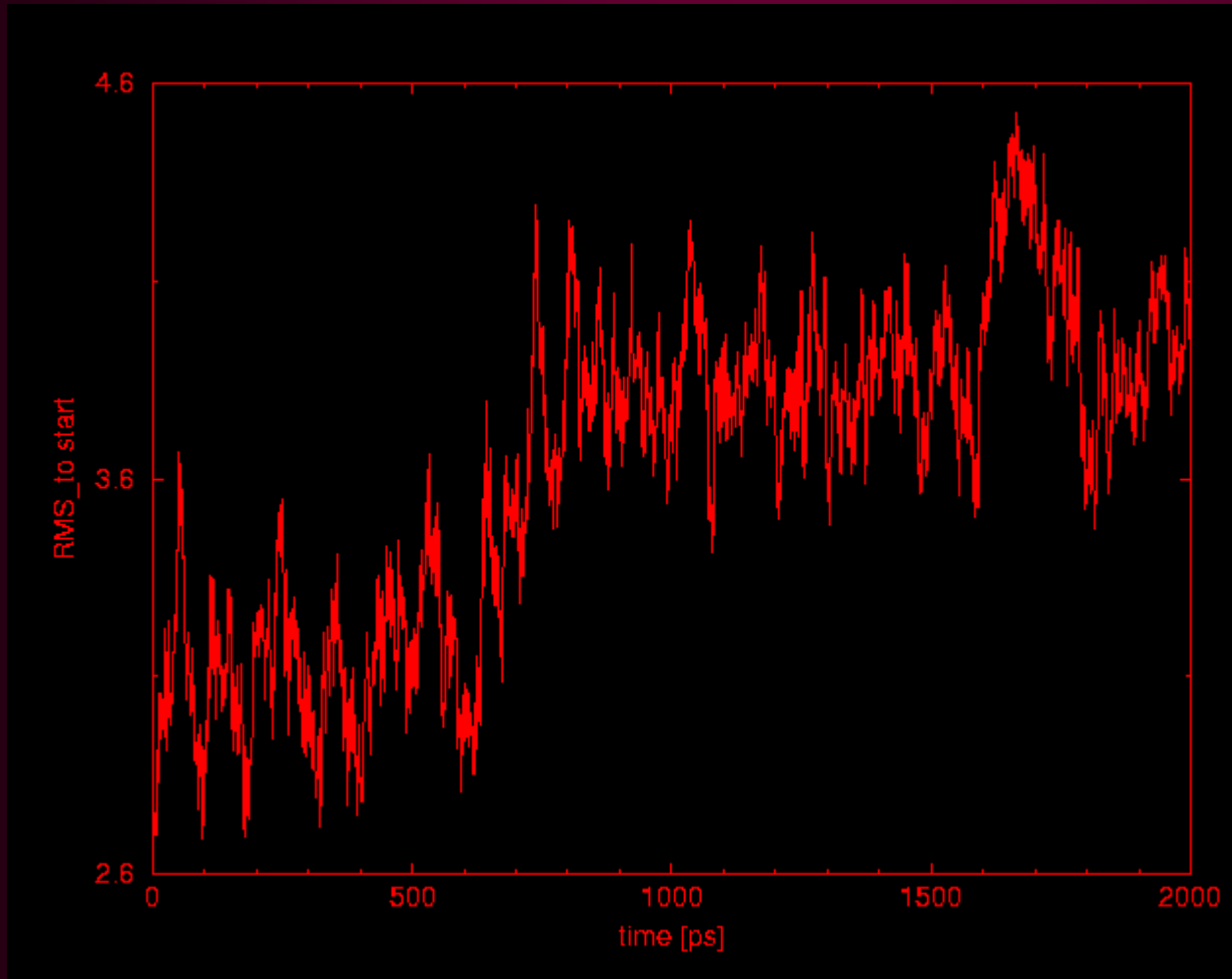


celková
energie

potenciální
energie

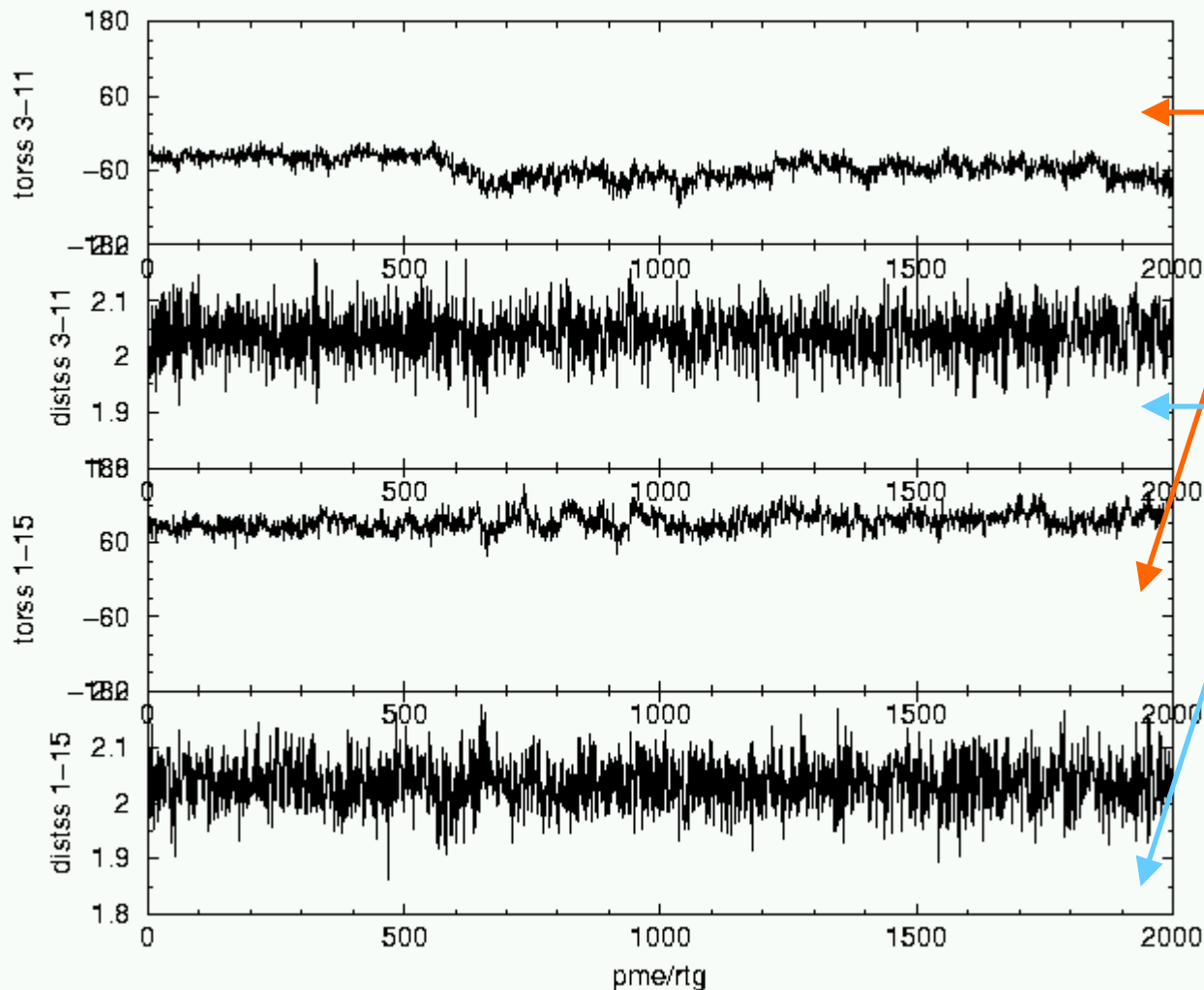
kinetická
energie

RMS



RMS odchyłka
(root-mean-square
deviation)

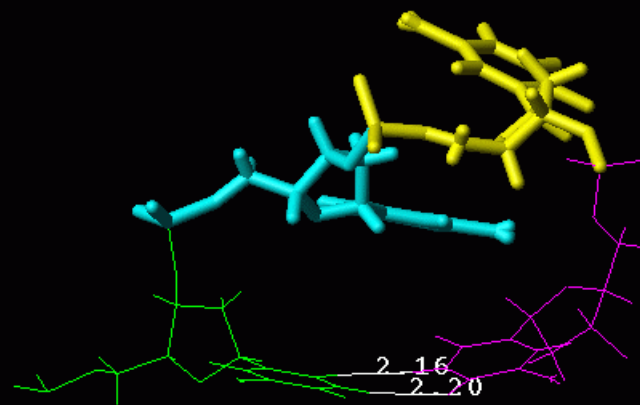
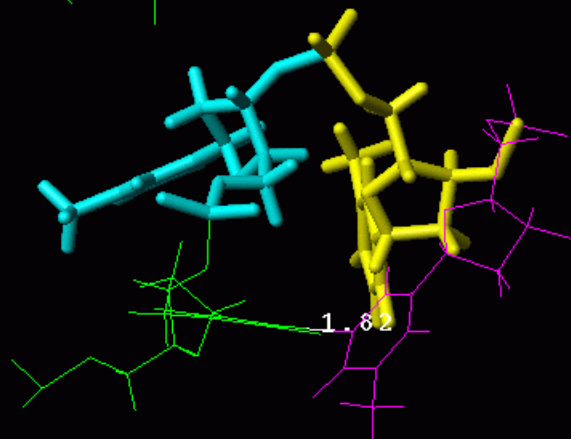
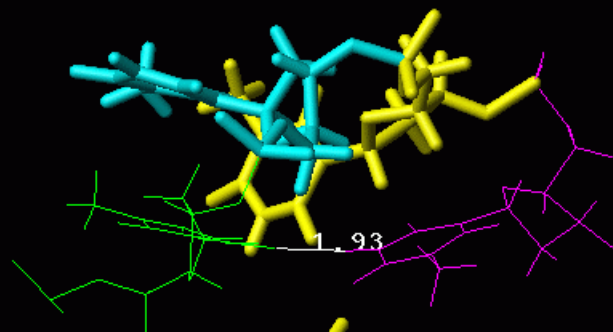
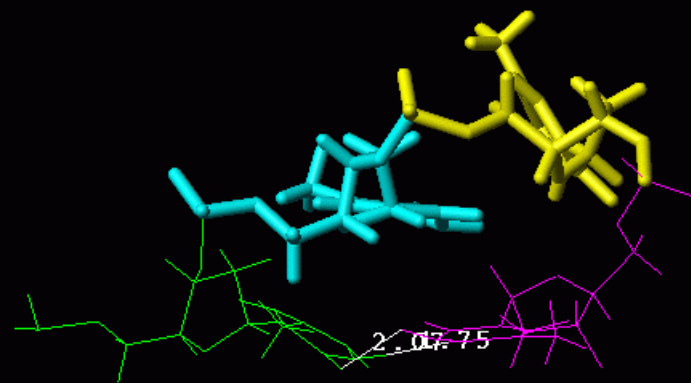
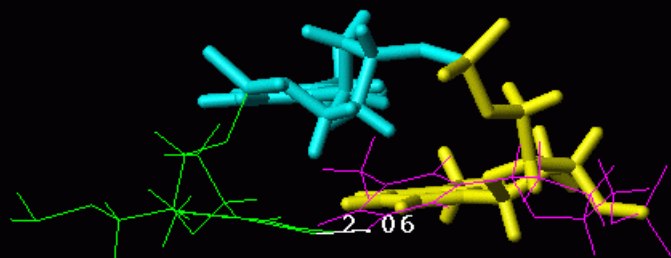
Geometrické parametry



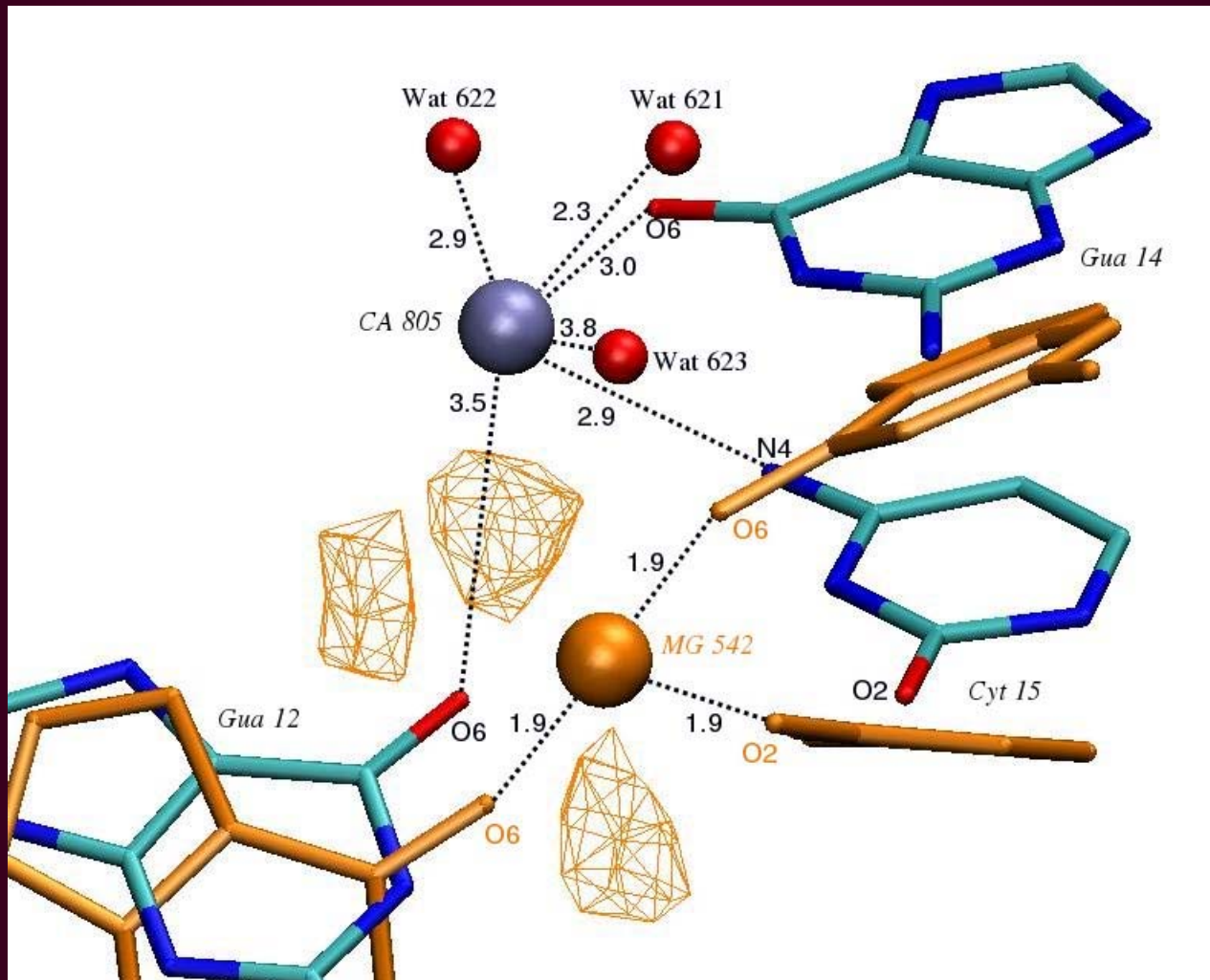
vzdálenosti

torzní úhly

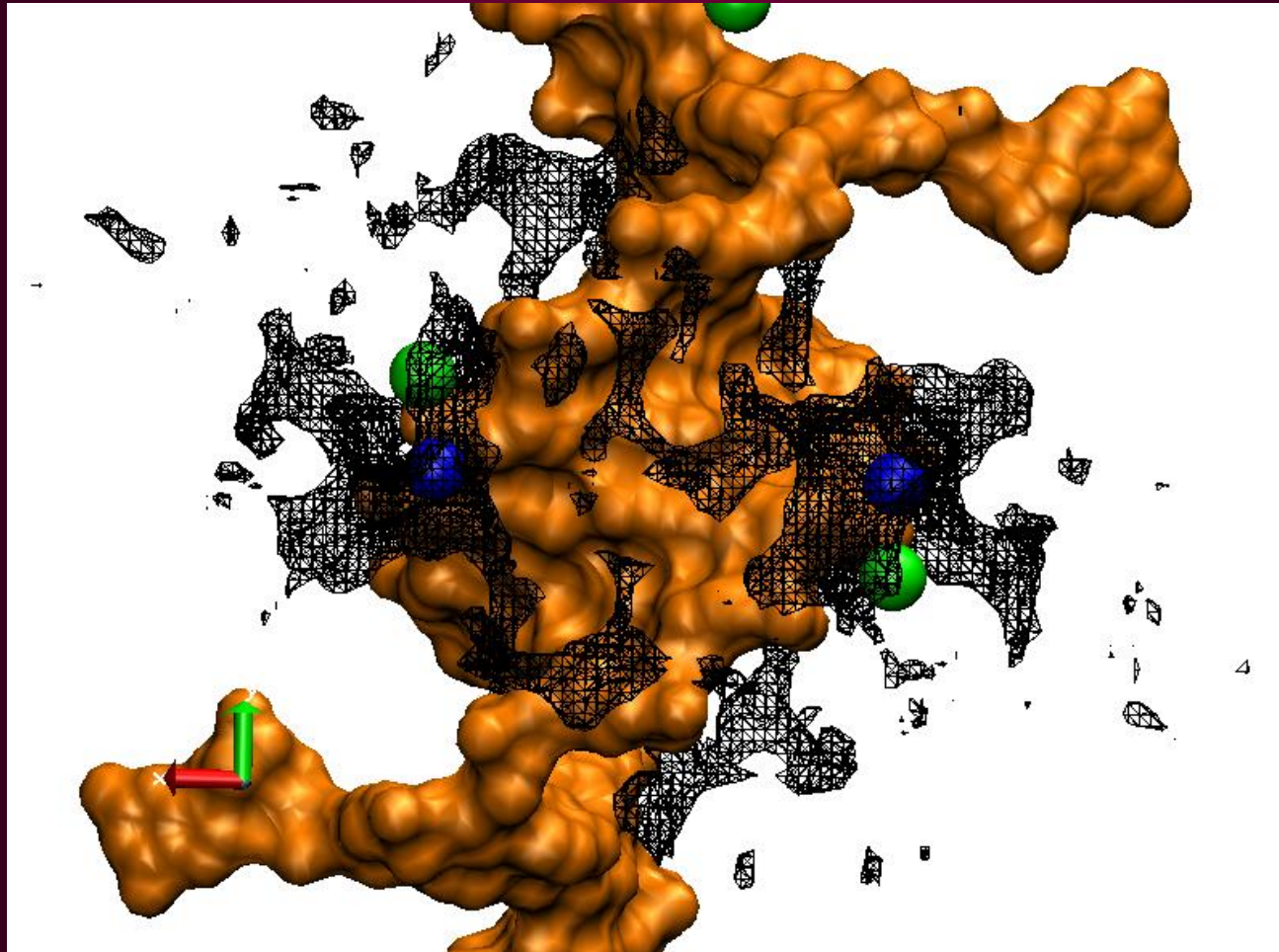
Změny ve struktuře



Vody a ionty



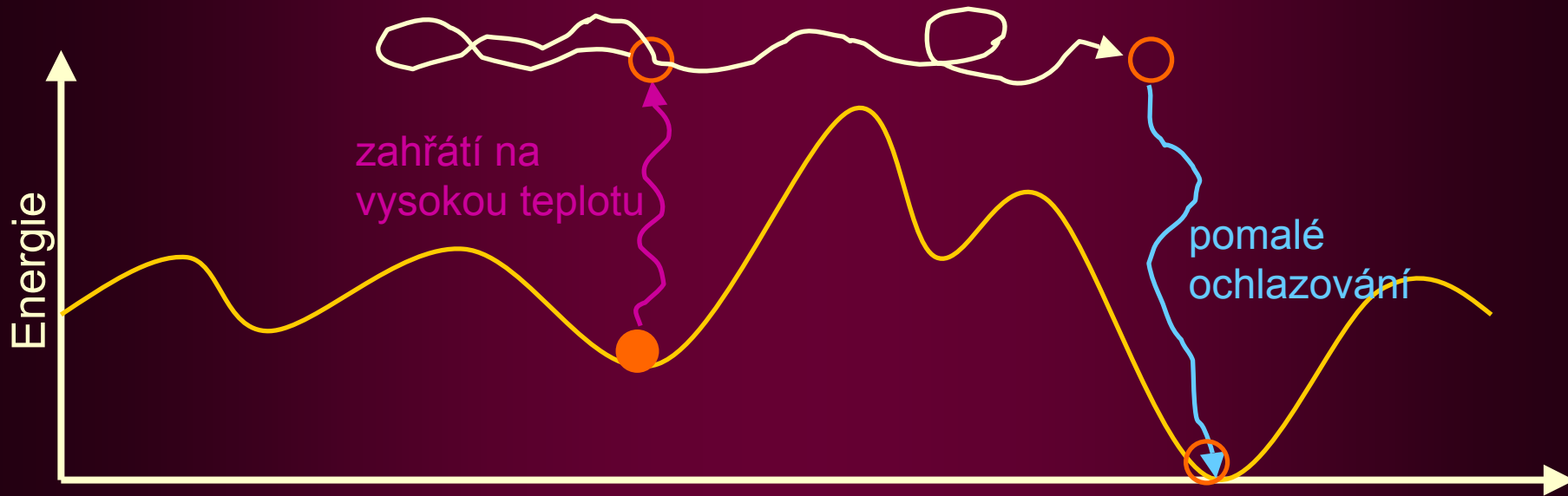
Elektrostat. potenciál



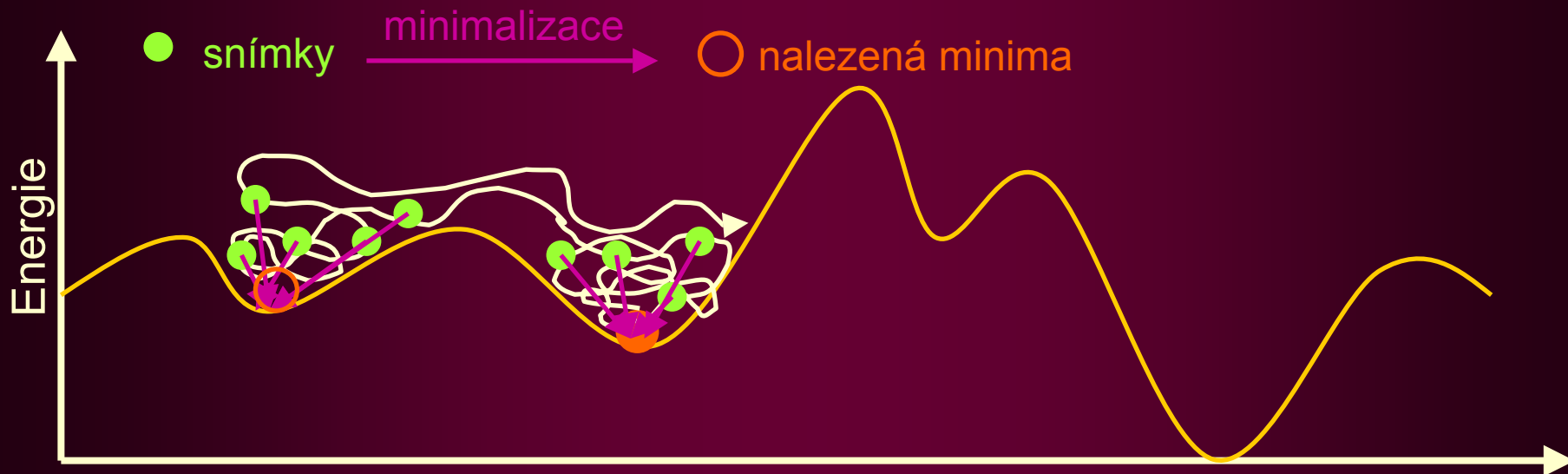
Vyhodnocení dat

- srovnání s experimentem
- mutace klíčových míst ve struktuře a následná MD
- termodynamická analýza
- vývoj lepšího modelu
- ...

Simulované žíhání



Quenched dynamics



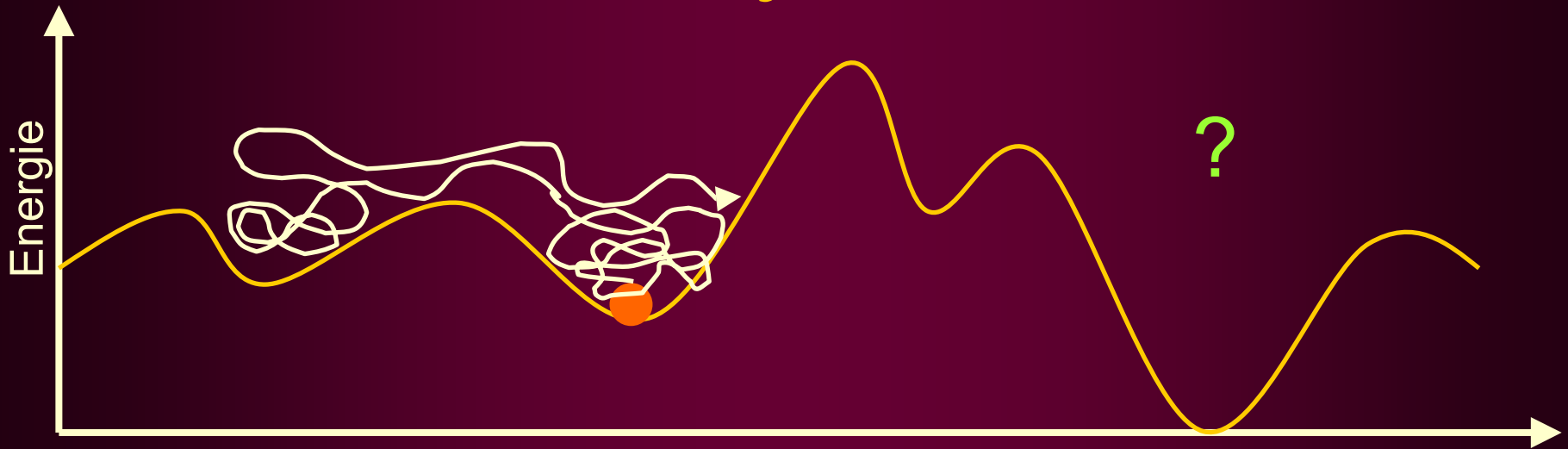
„Vázaná“ MD

= MD s užitím restraints

Restraints (vazné podmínky) ... získány z NMR



Problémy simulací



⇒ časová škála:

cíl: μs - ms

realita: ps - ns

⇒ kvalita reprezentace:

cíl: komplexní systém

realita: molekulová mechanika



SGI Power Challenge



SGI Onyx



SGI Origin2000



vt 100



PC clusters