HP-SEE

www.hp-see.eu

NAMD on PARADOX, CompChem (RS)

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

Input files: structure, and preparation of files

- Description of PDB file
- Description of PSF file
- Usage of topology files
- Usage of parameter files
    - Working example 1UBQ (NAMD native tutorial)
    - Solving of common problems
- Preparation of PDB and PSF files in Windows environment using VegaZZ
- Preparation of PDB and PSF files using VMD, native psfgen and command line

Overview

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

Job submission, monitoring and upload of results

- Transfer of prepared files to user home directory
- Basic shell commands
- Structure of .pbs and executable script
- How to do settings needed
- Job submission and monitoring
- Inspecting results on PARADOX
- Uploading results to user workstation

Software and files needed

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

(already on user terminals)

Secure file transfer - pscp

http://www.chiark.greenend.org.uk/~sgtatham/putty/download.html

GUI and console - VMD

http://www.ks.uiuc.edu/Development/Download/download.cgi?PackageName=VMD

Tutorial files

http://www.ks.uiuc.edu/Training/Tutorials/
http://www.ks.uiuc.edu/Research/namd/wiki/index.cgi?NamdAndVEGA

Input preparation, output analysis - VegaZZ

http://www.vegazz.net/

Alternative PDB files preparation - DeepView

http://spdbv.vital-it.ch/disclaim.html

NAMD

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

Molecular dynamics code NAMD (versions 2.8)
http://www.ks.uiuc.edu/Research/namd/

- compiled under Charmrun
- installed on PARADOX http://scl.ipb.ac.rs/
- 8 cores per node, 86 nodes; 64bit cluster
- many nodes can be used for parallel processing
- efficiency depend on the size of system and setting-up of the jobs
- known issue – slow communication between the nodes
- Windows version installed in your home directory

Input - PDB

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

PDB – structure of (bio)molecule

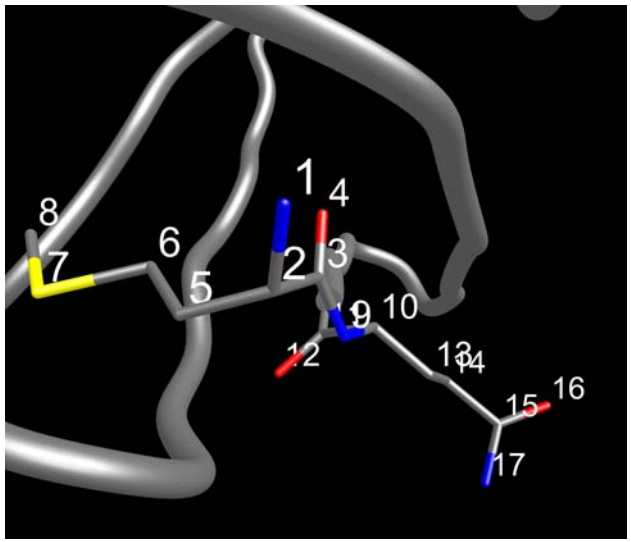
atom ID
atom name
residue name
residue ID
position of atoms (coordinates)
occupancy
temperature factor
segment name

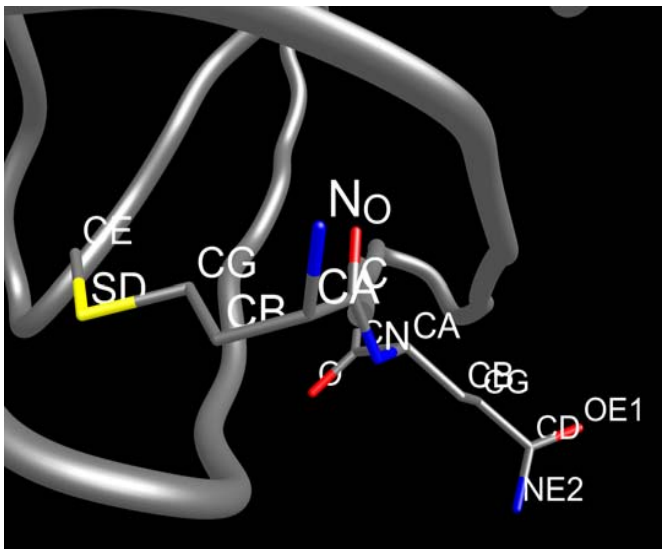
```
ATOM      1  N    MET     1      27.340  24.430   2.614  1.00  9.67      1UBQ  71
ATOM      2  CA   MET     1      26.266  25.413   2.842  1.00 10.38      1UBQ  72
ATOM      3  C    MET     1      26.913  26.639   3.531  1.00  9.62      1UBQ  73
ATOM      4  O    MET     1      27.886  26.463   4.263  1.00  9.62      1UBQ  74
ATOM      5  CB   MET     1      25.112  24.880   3.649  1.00 13.77      1UBQ  75
ATOM      6  CG   MET     1      25.353  24.860   5.134  1.00 16.29      1UBQ  76
ATOM      7  SD   MET     1      23.930  23.959   5.904  1.00 17.17      1UBQ  77
ATOM      8  CE   MET     1      24.447  23.984   7.620  1.00 16.11      1UBQ  78
ATOM      9  N    GLN     2      26.335  27.770   3.258  1.00  9.27      1UBQ  79
ATOM     10  CA   GLN     2      26.850  29.021   3.898  1.00  9.07      1UBQ  80
ATOM     11  C    GLN     2      26.100  29.253   5.202  1.00  8.72      1UBQ  81
ATOM     12  O    GLN     2      24.865  29.024   5.330  1.00  8.22      1UBQ  82
ATOM     13  CB   GLN     2      26.733  30.148   2.905  1.00 14.46      1UBQ  83
ATOM     14  CG   GLN     2      26.882  31.546   3.409  1.00 17.01      1UBQ  84
ATOM     15  CD   GLN     2      26.786  32.562   2.270  1.00 20.10      1UBQ  85
ATOM     16  OE1  GLN     2      27.783  33.160   1.870  1.00 21.89      1UBQ  86
ATOM     17  NE2  GLN     2      25.562  32.733   1.806  1.00 19.49      1UBQ  87
```

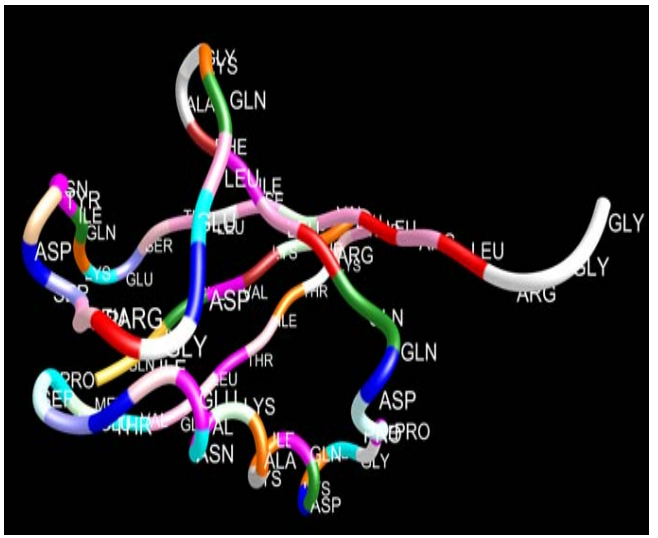
Input - PDB

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

PDB file http://www.pdb.org/pdb/home/home.do







-useful link:

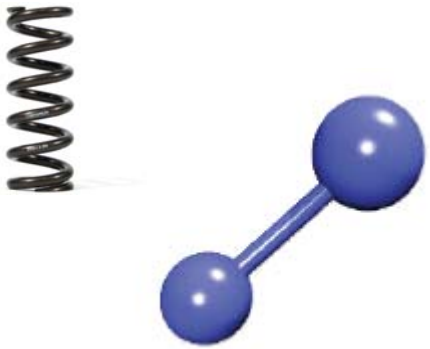
http://www.pdb.org/pdb/101/static101.do?p=education_discussion/Looking-at-Structures/coordinates.html

Input - PSF

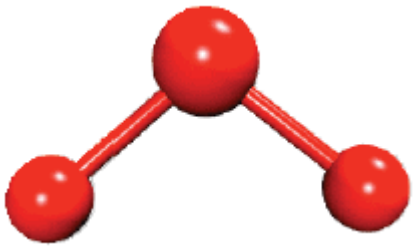
HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

PSF – protein structure file

- from PDB, and topology file (name begin with 'top')
- include atoms, bonds, angles, dihedrals, impropers
- X-PLOR atom type in NAMD, atom type by name



Bond



Angle

Dihedral

Improper

Input - PSF

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

Atoms

atom name

atom type

charge

mass

```
PSF

       1 !NTITLE
 REMARKS original generated structure x-plor psf file

    1231 !NATOM
       1 PROT 1    MET  N    NH3   -0.300000       14.0070          0
       2 PROT 1    MET  HT1  HC     0.330000        1.0080          0
       3 PROT 1    MET  HT2  HC     0.330000        1.0080          0
       4 PROT 1    MET  HT3  HC     0.330000        1.0080          0
       5 PROT 1    MET  CA   CT1    0.210000       12.0110          0
       6 PROT 1    MET  HA   HB     0.100000        1.0080          0
       7 PROT 1    MET  CB   CT2   -0.180000       12.0110          0
```

Input - PSF

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

Bonds (4 x 2 atoms per line)

```
1237 !NBOND: bonds
   1       5       2       1       3       1       4       1
   5       6       7       5       7       8       7       9
  10       7      10      11      10      12      13      10
```

Angles (3 x 3 atoms per line)

```
2257 !NTHETA: angles
   1       5       6       1       5      18       2       1       5
   2       1       4       2       1       3       3       1       5
   3       1       4       4       1       5       5      18      19
```

Input - PSF

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

Dihedrals and impropers (2 x 4 atoms per line)

```
3293 !NPHI: dihedrals
   1       5       7      10       1       5       7       8
   1       5       7       9       1       5      18      20
   1       5      18      19       2       1       5       7

 204 !NIMPHI: impropers
  18       5      20      19      20      18      22      21
  30      32      27      31      30      27      32      31
  32      30      33      34      32      30      34      33
```

Input - topology

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

Topology

- define the type, mass and (partial) charge of every atom of the each residue - H's are now added *i.e.* included
- masses are the same for each atom of the same element
- charges vary from residue to residue
- joining the residues

```
MASS     7 HR1    1.00800 H ! his he1, (+) his HG,HD2
MASS     8 HR2    1.00800 H ! (+) his HE1
MASS     9 HR3    1.00800 H ! neutral his HG, HD2
MASS    10 HS     1.00800 H ! thiol hydrogen
MASS    11 HE1    1.00800 H ! for alkene; RHC=CR
MASS    12 HE2    1.00800 H ! for alkene; H2C=CR
MASS    20 C     12.01100 C ! carbonyl C, peptide backbone
MASS    21 CA    12.01100 C ! aromatic C
MASS    22 CT1   12.01100 C ! aliphatic sp3 C for CH
MASS    23 CT2   12.01100 C ! aliphatic sp3 C for CH2
MASS    24 CT3   12.01100 C ! aliphatic sp3 C for CH3
```

```
DECL -CA
DECL -C
DECL -O
DECL +N
DECL +HN
DECL +CA
```

- not only bonds, but angles and dihedr. must be defined

Input - topology

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

- first and last residue is connected to one residue, the all other to two residues
- patches (patch residue is applied)
- first residue is NTER, last residue is CTER
- terminal of non-protein segment is TER
- angles and dihedrals are automatically generated for every 3 consecutively covalently connected atoms

Input - topology

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

Residues

```
RESI ALA              0.00
GROUP
ATOM N     NH1      -0.47   !       |
ATOM HN    H         0.31   !    HN-N
ATOM CA    CT1       0.07   !       |       HB1
ATOM HA    HB        0.09   !       |     /
GROUP                       !    HA-CA--CB-HB2
ATOM CB    CT3      -0.27   !       |     \
ATOM HB1   HA        0.09   !       |       HB3
ATOM HB2   HA        0.09   !     O=C
ATOM HB3   HA        0.09   !       |
GROUP                       !
ATOM C     C         0.51
ATOM O     O        -0.51
```

Each group of atoms have integer charge

- NAMD *do not* use group definitions

- **impropers**, maintain planarity of amide bond

```
IMPR N -C CA HN   C CA +N O
```

the first atom is connected to other 3

- **bonds**

```
BOND CB CA   N   HN   N   CA
BOND C  CA   C   +N   CA HA   CB HB1   CB HB2   CB HB3
DOUBLE O  C
```

Input - topology

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

- in CHARMM ff there is no explicit H bonds, so this part of record is ignored
- IC holds for internal coordinates (different that z – matrix)
- four atoms connected to each other 1 2 3 4

bond length 1-2 | angle1-2-3 | dihedral 1-2-3-4 | angle 2-3-4 | length 3-4

```
IC -C   CA   *N   HN    1.3551 126.4900  180.0000 115.4200  0.9996
IC -C   N    CA   C     1.3551 126.4900  180.0000 114.4400  1.5390
IC N    CA   C    +N    1.4592 114.4400  180.0000 116.8400  1.3558
IC +N   CA   *C   O     1.3558 116.8400  180.0000 122.5200  1.2297
IC CA   C    +N   +CA   1.5390 116.8400  180.0000 126.7700  1.4613
IC N    C    *CA  CB    1.4592 114.4400  123.2300 111.0900  1.5461
IC N    C    *CA  HA    1.4592 114.4400 -120.4500 106.3900  1.0840
```

Input - topology

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

- impropers differently treated in IC records

  1 2 *3 4 - atoms 1 2 4 connected to atom 3

- values: length 1-3, angle 1-3-2, dihedral 1-2-3-4, angle 2-3-4, length 3-4

- for water, generation of angles and dihedrals is disabled

- extra (non-existing) bond to hold rigid HOH in CHARMM, not needed by NAMD

-both topology and parameters available from

http://www.pharmacy.umaryland.edu/faculty/amackere/research.html

Input - parameters

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

Current

CHARMM22 – proteins
CHARMM27 – lipids and nucleic acids
CHARMM – cgenff - small molecules (do not use it for biomolecules !!!)

-hybrid:
par_all27_na_lipid.inp, par_all27_prot_lipid.inp, par_all27_prot_na.inp

- useful link https://www.paramchem.org/

- parameter file include numerical constants to evaluate forces and energies, using data from PSF and PDB

Input - parameters

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

bonds - pair of atom, force constant and equil. length

```
BONDS
!
!V(bond) = Kb(b - b0)**2
!
!Kb: kcal/mole/A**2
!b0: A
!
!atom type Kb          b0
!
C    C     600.000     1.3350 ! ALLOW ARO HEM
                ! Heme vinyl substituent (KK, from propene (JCS))
CA   CA    305.000     1.3750 ! ALLOW   ARO
                ! benzene, JES 8/25/89
CE1  CE1   440.000     1.3400   !
                ! for butene; from propene, yin/adm jr., 12/95
CE1  CE2   500.000     1.3420   !
                ! for propene, yin/adm jr., 12/95
CE1  CT2   365.000     1.5020   !
```

Input - parameters

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

angles - three atoms, force constant, equil. angle
- Urey-Bradley term for atoms 1 and 3

```
ANGLES
!
!V(angle) = Ktheta(Theta - Theta0)**2
!
!V(Urey-Bradley) = Kub(S - S0)**2
!
!Ktheta: kcal/mole/rad**2
!Theta0: degrees
!Kub: kcal/mole/A**2 (Urey-Bradley)
!S0: A
!
!atom types     Ktheta    Theta0   Kub     S0
!
CA   CA   CA    40.000    120.00   35.00   2.41620 ! ALLOW   ARO
                ! JES 8/25/89
CE1  CE1  CT3    48.00    123.50   !
                ! for 2-butene, yin/adm jr., 12/95
CE1  CT2  CT3    32.00    112.20   !
                ! for 1-butene; from propene, yin/adm jr., 12/95
CE2  CE1  CT2    48.00    126.00   !
                ! for 1-butene; from propene, yin/adm jr., 12/95
```

Input - parameters

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

dihedrals - sine function

P is plane angle, first plane defined by atoms 1 2 3, second by atoms 2 3 4

- X record 'try to find closest match'

```
DIHEDRALS
!
!V(dihedral) = Kchi(1 + cos(n(chi) - delta))
!
!Kchi: kcal/mole
!n: multiplicity
!delta: degrees
!
!atom types             Kchi    n   delta
!
C    CT1  NH1  C        0.2000  1   180.00 ! ALLOW PEP
               ! ala dipeptide update for new C VDW Rmin, adm jr., 3/3/93c
C    CT2  NH1  C        0.2000  1   180.00 ! ALLOW PEP
               ! ala dipeptide update for new C VDW Rmin, adm jr., 3/3/93c
C    N    CP1  C        0.8000  3     0.00 ! ALLOW PRO PEP
               ! 6-31g* AcProNH2, ProNH2, 6-31g*//3-21g AcProNHCH3 RLD 4/23/93
```

```
X    C    C    X        4.0000  2   180.00 ! ALLOW HEM
                ! Heme (6-liganded): substituents (KK 05/13/91)
X    C    NC2  X        2.2500  2   180.00 ! ALLOW   PEP POL ARO
                ! 9.0->2.25 GUANIDINIUM (KK)
```

Input - parameters

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

impropers – to maintain planarity

R is angle between plane of atoms 1 2 3 and plane of atoms 2 3 4

- X as above

```
IMPROPER
!
!V(improper) = Kpsi(psi - psi0)**2
!
!Kpsi: kcal/mole/rad**2
!psi0: degrees
!note that the second column of numbers (0) is ignored
!
!atom types           Kpsi                   psi0
!
CPB  CPA  NPH  CPA    20.8000          0      0.0000 ! ALLOW HEM
                ! Heme (6-liganded): porphyrin macrocycle (KK 05/13/91)
CPB  X    X    C      90.0000          0      0.0000 ! ALLOW HEM
                ! Heme (6-liganded): substituents (KK 05/13/91)
```

- Non bonded, partially used in NAMD

- explicit H-bond *not* used

```
END
```

- end of record

Protein –some issues

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

We need all atoms of protein for simulation, problems

- missing parts of side chains

- usually at the outer surface of the protein, high flexibility
- so you can find only CB
- lysine, arginine, glutamine…

- need to retain crystal water – can be important during simulation

- by default both psfgen and Deep View remove this water

- you can 'save' it in separate file, but than you shouldn't use roto-translations in GUI

- by Deep View you can retain cofactors, phosphorilated residues..

- by Vega you can 'skip' such problems

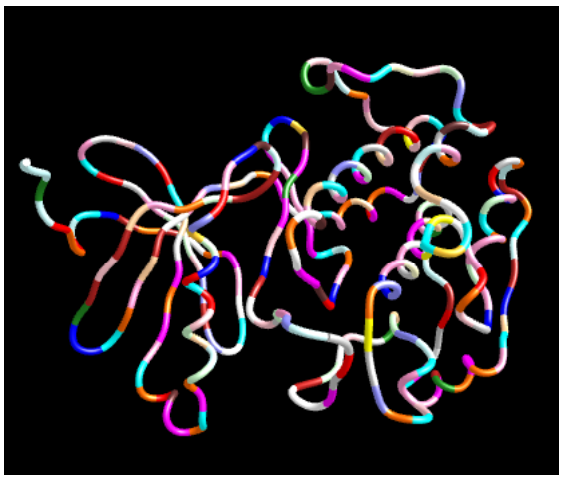
Protein

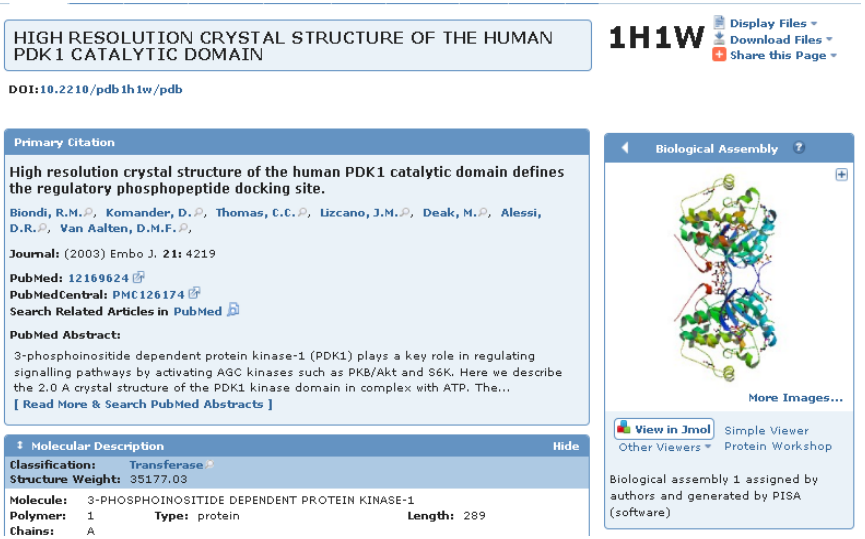
HP-SEE
High-Performance Computing Infrastructure for South East Europe's Research Communities

- REMARK 470 in PDB file – missing atoms

- example 1H1W

```
REMARK 470
REMARK 470 MISSING ATOM
REMARK 470 THE FOLLOWING RESIDUES HAVE MISSING ATOMS (M=MODEL NUMBER;
REMARK 470 RES=RESIDUE NAME; C=CHAIN IDENTIFIER; SSEQ=SEQUENCE NUMBER;
REMARK 470 I=INSERTION CODE):
REMARK 470   M RES CSSEQI  ATOMS
REMARK 470     GLN A  73    CG   CD   OE1  NE2
REMARK 470     ARG A  75    CG   CD   NE   CZ   NH1  NH2
REMARK 470     ARG A 238    CG   CD   NE   CZ   NH1  NH2
REMARK 470     GLU A 303    CG   CD   OE1  OE2
REMARK 470     LYS A 304    CG   CD   CE   NZ
REMARK 470     GLU A 348    CG   CD   OE1  OE2
REMARK 470     LYS A 357    CG   CD   CE   NZ
```

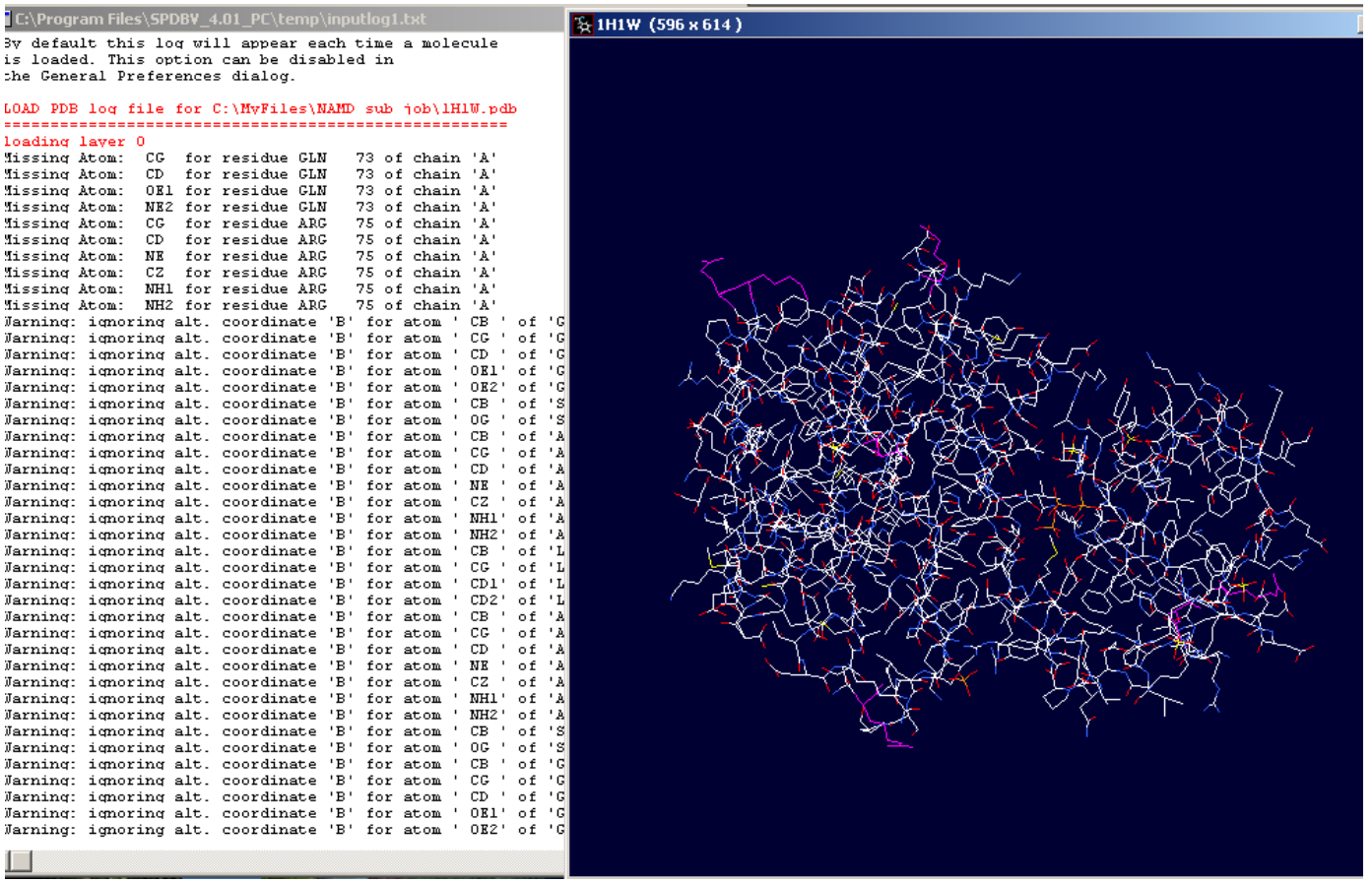




Protein

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

- reconstruction by Deep View



- save the Deep View log file - copy from temp folder of the exc.

-reconstructed side chains given in pink

Protein

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

- reconstruction by VegaZZ

  - remove the CB of missing chains manually
    Edit/Remove/Atom - atom by atom

  - add missing side chains
    Edit/Add/Side chains - automatically

  - less manual work, and you will retain everything existing in PDB file

- there are other tools for such purposes, some of which should be paid

Input preparation psfgen

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

- 1UBQ and top_all27_prot_lipid.inp in your home directory

structure topology

- load structure to VMD, than open console, 'go' to your home (all training files are stored in your home directory)
- select protein – as segment - to new PDB file

```
set ubq [atomselect top protein]
$ubq writepdb ubqp.pdb
```

-ubqp.pdb appear in directory, manually change HIS to HSE (text editor)

- there is no HIS in top, but HSD, HSE and HSP

Input preparation psfgen

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

- copy and paste to empty file in text editor

```
package require psfgen
topology top_all27_prot_lipid.inp
alias atom ILE CD1 CD
segment U {pdb ubqp.pdb} coordpdb ubqp.pdb U
guesscoord
writepdb ubq.pdb
writepsf ubq.psf
```

- save this file as ubq.pgn to your home directory

- type in console source ubq.pgn

- in this way you were add H and make PSF file, close VMD

-psfgen require separate *segment* for each chain, water, heteroatoms etc.

Input preparation psfgen

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

- two new files appear in your home directory

ubq.pdb ubq.psf

- to solvate protein with 5 Å water box, open VMD, open console, go to your home directory (cd path) and type

```
package require solvate
solvate ubq.psf ubq.pdb -t 5 -o ubq_wb
```

- in this way you obtained protein in water box, 5 Å from protein 'edges' (files ubq_wb.pdb and ubq_wb.psf files appear in your home direct.)

- lets try to do the similar using VegaZZ

Input preparation VegaZZ

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

- take original file 1UBQ from PDB, rename it to 1UBGO

- open VegaZZ, import 1UBQO.PDB (drag and drop from the file browser)

- go to Edit/Add/Hydrogens dialog block appears

- chose 'Protein' and 'Residue end', 'Use IUPAC atom nomenclature' are chosen by default

- go to Calculate/Charge & Potential

-in dialog block chose 'CHARM22 PROT' for FF, and 'CHARM CHAR' for charge, click on 'Fix'

- in console you will see that system is neutral 'Total charge 0.0'

Input preparation VegaZZ

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

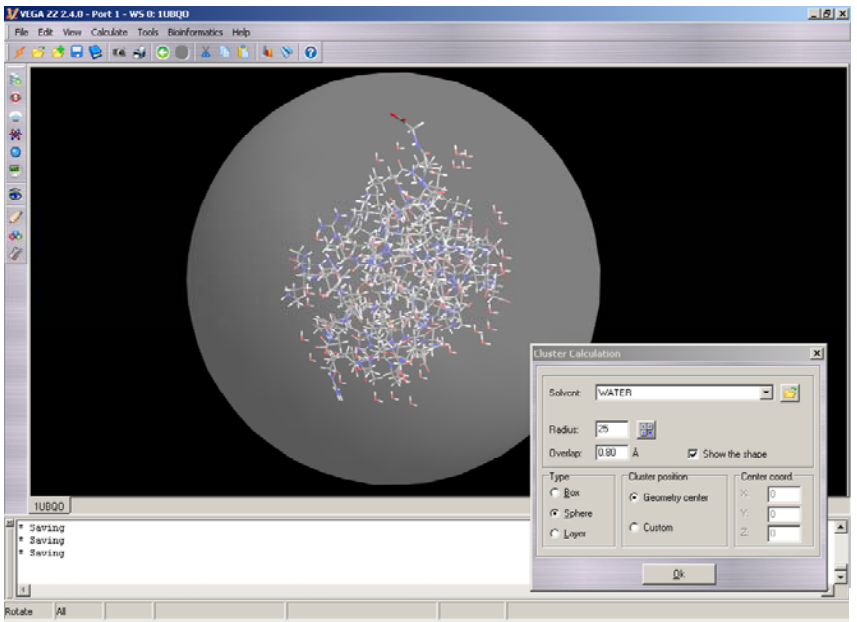
- save the file with ‘iff’ extension
- save it again as psf
- save it again as PDB2.2 (exclude connectivity)

- firstly saved file (iff) retain all data needed - atomic positions and topology, so you could load and use this file later

- to solvate protein

- Edit/Add/Cluster

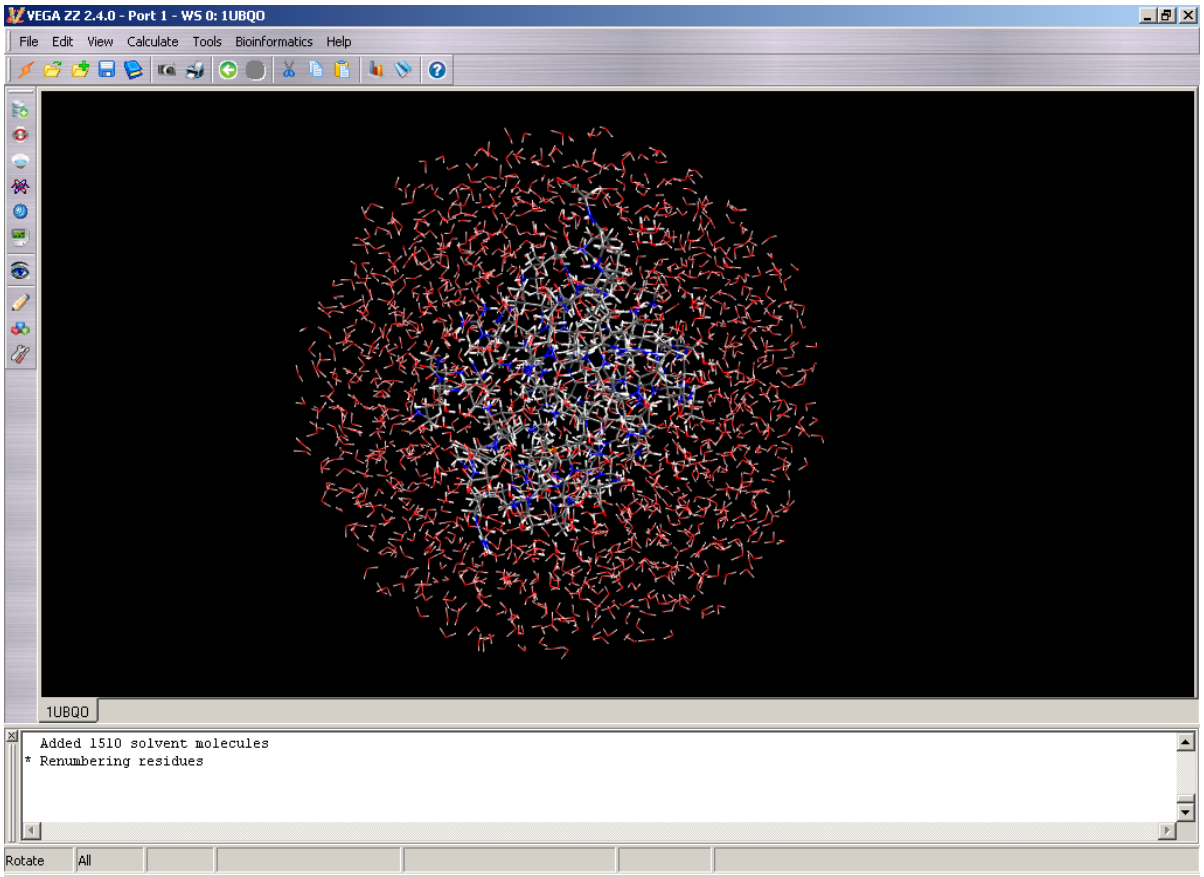
-in dialog block chose
water
sphere
geometry center
27 Å
and mark ‘Show the shape’



Input preparation VegaZZ

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

- you will see protein in the nice water sphere



- ascribe charge and potential

- save file as 1UBQOws in iff, pdb, and psf extension

Molecular dynamics - configuration

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

- minimization
- heating
- equilibration (stable system should be promptly equilibrated)
- molecular dynamics (MD)
- input files needed – for all 4 above

coordinates	file.pdb
structure	file.psf
parameters	par_all27_prot_lipid.prm

- we will use hybrid protein-lipid parameters
- native CHARMM parameter files start with 'par'
- extension can be '.inp' and '.prm'

Molecular dynamics - configuration

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

- NAMD 'accept' more that one parameter file
- all parameters *must* correspond to topology used for psf generation
- equilibration and MD also need
  - binary coordinates file.coor
  - velocities file.vel
- this files include information on 'history of the system'
- position of atoms after (for example) heating, and velocities of atoms (during heating we 'add' energy to the whole system)
- during simulation two 'backups' appear, for both coor and vel files (restart.coor, restart.vel and restart.coor.old, restart.vel.old)

Molecular dynamics - configuration

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

- name of output files should be chosen

outputname file2 (without extension – all files generated will have this name)

DCDfile file2.dcd (trajectory, same name as outname)

- for simplicity no explanation of all lines needed is included but you can find it in NAMD 2.8 ug

http://www.ks.uiuc.edu/Research/namd/2.8/ug/

- or tutorial

http://www.ks.uiuc.edu/Training/SumSchool/materials/tutorials/02-namd-tutorial/namd-tutorial.pdf

Molecular dynamics - configuration

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

- minimization and the heating of system are short tasks
- we will perform this parts in user terminal
- MD will run on Paradox
  (configuration files provided to attendees)

- semi-automatic configuration file preparation by VegaZZ
  - load 1UBQOws.pdb and 1UBQOws.psf to Vega
  - go to Calculate/NAMD, dialog block appears
  - in the very last (Other) 'card' chose 'Min all free (generic)'
  - in the 3rd 'card' (Output) change name of Output and trajectory files to 1UBQOws-min

Molecular dynamics - configuration

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

- in the second ‘card’ uncheck automatic input, in ‘Parameter file’ field, browse to your home directory and chose ‘par_all27_prot_lipid.prm’

- I the 4th card change ‘Distance for inclusion in the pair list’ from 12 (this is default) to 13.5

- in the main part of dialog block (Run mode) check ‘prepare the input file only’, than click on Run

- configuration file 1UBQOws-min.namd appear in directory

- open the configuration file in text editor

- remove the ‘paths’ in coordinates, structure, parameters and output lines (delete part of line), save file

Molecular dynamics - configuration

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

- open command prompt from your home directory, than type 'namd2 +p2 1UBQOws-min.namd > 1UBQOws-min.out'

- 10000 steps take some 15 min; +p2 is number of processors

- open '.out' file and compare CPUtime and WallClock at the very end of the file – times are the same – good scaling

(repeat the same for the ubq_wb.pdb and ubq_wb.psf files obtained by psfgen as exercise)

-open 1UBQOws.psf in Vega, go to Calculate/Analysis, drag and drop 1UBQOws-min.dcd

- click on 'Energy', graph show decrees of pot. energy of system

- go to last frame, save this frame to file as 1UBQOws-heat iff

Molecular dynamics - configuration

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

- save the same file again as PDB and PSF
- close the Vega, open program again and load PDB and PSF files
- go to Calculate/NAMD, dialog block appears
- in the very last (Other) 'card' chose 'Heating 300K (generic)'
- in the 3rd 'card' (Output) change name of Output and trajectory files to 1UBQOws-heat, in the first card (Basic) change 'Number of timesteps' from 9000 to 10000
- repeat the same procedure, as described, from the beginning of the slide 36 to the red line at the slide 37, this time your input is 1UBQOws-heat.namd
- heating procedure should be finished within 15 min

Molecular dynamics - configuration

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

File preparation for productive run – this files will be transferred to Paradox

- script.sh and task.pbs are in the your home directory and will be transferred to Paradox user directory along with other files

- open 1UBQOws-heat.pdb in Vega.

- go to Calculate/NAMD

- in “Basic’ card set ‘Number of steps’ to 300000

- in the same card set ‘Starting timestep value’ to 10000

- in ‘Input’ card uncheck ‘automatic’

Molecular dynamics - configuration

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

- now we will use binary coordinates and velocities obtained from heating phase in our input
- uncheck ‘automatic’ in ‘Input’ card
- fill fields as follows

Coordinates PDB file	1UBQOws-heat.pdb
Binary PDB file	1UBQOws-heat.coor
PSF file	1UBQOws-heat.psf
Parameter file	par_all27_prot_lipid.prm
Velocities file	1UBQOws-heat.vel

- in ‘Output’ card fill:

Output file	1UBQOws-md
Trajectory file	1UBQOws-md.dcd

Molecular dynamics - configuration

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

- in lower left part of ‘PME’ card click ‘calculate’

- same numbers- (50) points should appear in x y and z direction

- in BC, chose ‘periodic’, than enable this part. Click blue button left from the ‘Cell origin’, numbers close to zero should appear

- click blue button above, numbers near 50 should appear in diagonal fields

- at the bottom of this dialog box turn to on ‘Wrap water’ and ‘Wrap nearest’

- in the upper part of the dialog box retain ‘prepare input file only’

- click run

Molecular dynamics - configuration

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

-in your home directory new file should appear
1UBQOws-md.namd

- execute this file by NAMD installed locally to be sure that all settings are OK. Kill the job when 1UBQOws-md.dcd file appear

- check in text editor are some path is retained to input or output files. Remove this paths if exist

- copy following files to separate (sub)directory of your home, with name 'training'

1UBQOws-heat.pdb; 1UBQOws-heat.coor
1UBQOws-heat.psf; 1UBQOws-heat.vel
par_all27_prot_lipid.prm; 1UBQOws-md.namd

as well as script.sh, job.sh and task.pbs

Molecular dynamics - configuration

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

- pack 'training' directory

```
tar zcvf training.tgz training/
```

- transfer file to your user home directory on Paradox, using pscp

```
pscp training.tgz username@ui.ipb.ac.rs:
```

- login to your account on Paradox, than transfer file to directory from which it should be executed

```
cp training.tgz /nfs/username/
```

- it is good idea to make 'NAMD-devoted' sub-directory, now you are in /nfs/username/

```
mkdir namd
```

- move, than unpack your tgz to new namd sub-directory

```
mv training.tgz namd/
cd namd/
tar zxvf training.tgz
```

Molecular dynamics - configuration

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

- in the training directory, open pbs file, go to insert mode

  cd training

  vi task.pbs

  type I after file opening

- change number of nodes that you will use. In the heating .out file you can see that system comprise 3 by 3 by 3 patches (look below 'structure summary' in 'info' lines).

- you will use processors efficiently if employ less processors than patches in the system (27). So chose 3 nodes and 8 processors per node (3 x 8 = 24).

- chose 5 hours as a time of simulation (more than needed)

(basic 'console commands' was provided to attendees few days before the training)

Molecular dynamics - configuration

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

- exit from insert mode

  Esc

- save your change closing the file

  :wq

- now we should modify script file in order to be functional for our simulation, open file

  vi script.sh

- at the very beginning of the file, change the number of processors. This number must be the same as you requested in the .pbs file. So, you should chose 24 process.

  PROC_NUM=24

Molecular dynamics - configuration

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

- to overcome slow communication between the nodes, in some extent, we will 'leave' the one processor for communication. So, in the second line from the bottom of our script.sh add +p23 after …../namd2.8/namd2

…./namd2.8/namd2 +p23

- we should change script in order to allow system to find our NAMD configuration file (1UBQOws-md.namd)
- so we should give the 'absolute' path to our configuration file
- in the very last line of the script change existing path to /nfs/username/namd/training/1UBQOws-md.namd
- save changes and exit

Molecular dynamics - configuration

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

- to speed up our computation a bit more, open 1UBQOws-md.namd and change DcdFreq to 10000

```
vi 1UBQOws-md.namd
 I
DCDfreq    10000
 Esc
:wq
```

- if you need more frequent sampling of trajectory, such change is not favorable

- also you could increase outputiming, but just to some extent. During simulation output files become larger and larger, and all 'backups' are written to the memory till new restart file was written. As on your lab computer, there is some memory limit on cluster also

Job submission

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

Now we will submit our job

- type

  `qsub task.pbs`

- in console something like

  `1234567.ce64.ipb.ac.rs` should appear

- you could monitor your job

  `qstat 1234567`

- or all jobs that you submitted

  `qstat –u username`

- if your job running successfully after less than minute file nodelist should appear in the directory

(detail user guide to PBS submission is provided to users)

Job submission

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

- open this file and remember the nodes number
- you could monitor running jobs at ganglia

http://ganglia.scl.rs
(chose Grid64bit WNs)

using web browser, without logging to ipb.ui

- suitable for time-consuming tasks
- so far output files are written to your directory at the end of simulation, still you could monitor size of dcd file to gain an impression of the job progress
- script.sh will be improved in future to allow out file to be written to home directory from the start of calculation

Job monitoring

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

- each error are written to both out and err files

1234567.ce64.ipb.ac.rs.out
1234567.ce64.ipb.ac.rs.err

that appear in your home directory (*i.e.* directory from which you run simulation) instantly when job stops

- visual inspection of this two files allow you to fix the problem, and resubmit the job. There is unified job ID for the each submission.
- when your job is successfully finished, check the output. Open out file and go to the end of file
  Shift+G

Job monitoring

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

- you should see something like

```
ENERGY: 30010000      1563.8725      2257.0826       467.9256       963.7698         -27326
.3774     -2118.6158         0.0000         0.0000      5000.3895         -19191.9532
 304.5643    -24192.3427    -19147.3316       309.2949            -14.4491      -400.1310
  117833.5258      -402.4146      -402.4867

WRITING EXTENDED SYSTEM TO RESTART FILE AT STEP 30010000
WRITING COORDINATES TO DCD FILE AT STEP 30010000
WRITING COORDINATES TO RESTART FILE AT STEP 30010000
FINISHED WRITING RESTART COORDINATES
The last position output (seq=30010000) takes 0.005 seconds, 17.301 MB of memory in use
WRITING VELOCITIES TO RESTART FILE AT STEP 30010000
FINISHED WRITING RESTART VELOCITIES
The last velocity output (seq=30010000) takes 0.007 seconds, 17.288 MB of memory in use
WRITING EXTENDED SYSTEM TO OUTPUT FILE AT STEP 30010000
CLOSING EXTENDED SYSTEM TRAJECTORY FILE
WRITING COORDINATES TO OUTPUT FILE AT STEP 30010000
CLOSING COORDINATE DCD FILE
The last position output (seq=-2) takes 0.008 seconds, 16.845 MB of memory in use
WRITING VELOCITIES TO OUTPUT FILE AT STEP 30010000
The last velocity output (seq=-2) takes 0.003 seconds, 16.820 MB of memory in use
=============================================================

WallClock: 449145.593750  CPUTime: 407575.093750  Memory: 16.820145 MB
```

- compare WallClock and CPUtime

Job monitoring

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

- if those two numbers are close to each other you has good scaling and efficiently used processors
- close editor, and exit your ‘simulation' directory (‘training’ in our case)

```
cd ..
```

- now pack whole directory to tgz

```
tar  zcvf  training_output.tgz  training/
```

- this can take some time because files size in directory
- copy or move this file to your home directory (same one in which you was transfer input files from your computer at the start of this tutorial)

```
cp training_output.tgz  /home/username/
```

Downloading output

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

- download output to your lab computer, using pscp

- open command prompt from directory in your lab comp., where pscp is installed and type

pscp username@ui.ipb.ac.rs:training_output.tgz training_output.tgz

Now we finished our exercise
Many details are included because training 'target' chemist,
Much more discussion was done during training… 
Main goal is to allow one to submit, monitor jobs,
and 'collect' output without help of more experienced colleagues

Hope that all of this can be useful

CompChem (RS)

THEORETICAL *and* COMPUTATIONAL BIOPHYSICS GROUP

NIH RESOURCE FOR MACROMOLECULAR MODELING AND BIOINFORMATICS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Home | Software | Research | Publications | Instruction | Media | Facilities | About Us

- Home
- Overview
- Publications
- Research
- Software
  - VMD Molecular Graphics Viewer
  - NAMD Molecular Dynamics Simulator
  - BioCoRE Collaboratory Environment
  - MD Service Suite
  - Structural Biology Software Database
  - Computational Facility
- Outreach

Download NAMD

Download VMD

Parallel Programming Laboratory

NAMD
Scalable Molecular Dynamics

NAMD, recipient of a **2002 Gordon Bell Award**, is a parallel molecular dynamics code designed for hi pe a of large biomolecular systems. Based on **Charm++ parallel objects**, NAMD **scales** to hundreds processors hi ral s of processors on commodity **clusters** using gigabit ethernet. NAMD uses the popular mole h pro sim ajectory analysis, but is also file-compatible with AMBER, CHARMM, and X-PL urc ild NAMD yourself or download **binaries** for a wi ariety of platfo MD bio g.

The ferenc at h 2 .

Se

S

Thanks for your attention and time

Other Spotlights

M ng cells to another, for example transport cell components along th ein, such as myosin VI, has to "walk" or "run" along the cellular highway of n the transport. In the case of myosin VI, snapshots from crystallography revealed that the p e too short to explain the step size taken. Computational and experimental biophysicists have now solved the mystery of how myosin VI dimers realize their large step size despite their short legs. The investigation, based on the program **NAMD** and **reported recently**, demonstrates that the answer lies in the flexibility of the legs. Myosin VI is able to triple the length of each leg, made of a short bundle of up-down-up connected alpha-helices, by extending the bundle to a stretched-out down-down-down geometry of segments, like turning a letter z into a a single long line. In the telescoping process, myosin VI also gets help from its well-known binding partners, namely calmodulins. The calmodulins direct the telescoping of the protein legs as well as strengthen the extended legs. Together with an earlier study of the "neck" region of the molecule (see December 2010 highlight on **Opposites Attract in a Motor Protein**), the scientists have established how walking myosin VI achieves its wide stride. More information can be found on our **motor protein website**.

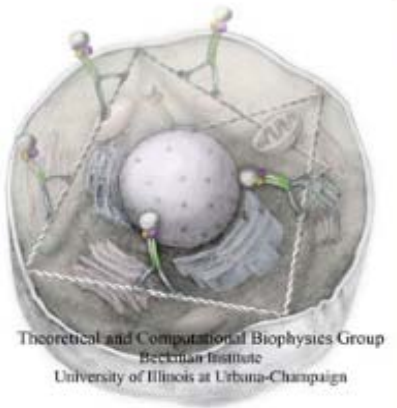


image size: 266.3KB