Applied Statistical Thermodynamics

How to analyse using GROMOS

Chris Oostenbrink
Institute of molecular modeling and simulation
University of Natural Resources and Life Sciences

Institute of molecular modeling and simulation

Applied Statistical Thermodynamics

Schedule

Time	Mon. 20/11/17	Tue. 21/11/17	Wed. 22/11/17	Thur. 23/11/17	Fri. 24/11/17	Mon. 27/11/17	Tue. 28/11/17	Wed. 29/11/17	Thur. 30/11/17	Fri. 01/12/17
9:00	Free Time	Welcome JAG	Free Time	Free Time	Free Time	Free Time	Free Time	Free Time	Free Time	Free Time
9:30		Lecture 1: Overview and Introduction JAG	Lecture 4: Thermodynamics JAG	Lecture 7: Classical Mechanics I JAG	Lecture 10: Ensembles I CO	Lecture 13: Free energies: reaction coordinates JAG	Lecture 16: Calculating properties from simulations JAG	Lecture 19: Electrostatics WvG	Lecture 22: On the ethics of the academic endeavour: where do we go?	Lecture 24: Left-overs/ questions and future perspectives WvG
10:15		Break	Break	Break	Break	Break	Break	Break		Break
10:30		Lecture 2: Molecular Simulations MD/SD/MC CO	Lecture 5: Force-Field Development CO	Lecture 8: Classical Mechanics II JAG	Lecture 11: Ensembles II JAG	Lecture 14: Boundary Conditions I WvG	Lecture 17: Comparison with Experiments WvG	Lecture 20: Polarization WvG	WvG	Lecture 25: Students plans I
11:15		Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break
11:45		Lecture 3: How to simulate using GROMOS CO	Lecture 6: Structure Refinement CO	Lecture 9: Analizing with GROMOS CO	Lecture 12: Free energies: alchemistry CO	Lecture 15: Boundary Conditions II WvG	Lecture 18: Searching & Enhanced Sampling WvG	Lecture 21: Multi -resolution simulations WvG	Lecture 23: QM/MM WvG	Lecture 26: Students plans II
12:30		Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Break for lunch, self-study, discussion.*	Tutorial 10: **
14:00	Registration Tutorial 1 Linux, NLHPC and OS-dongle installation	Tutorial 2: Running MD GROMOS Tutorial	Tutorial 3: Running MD Students Plans	Tutorial 4: Running MD Students Plans	Tutorial 5: Running MD Students Plans	Tutorial 6: Statistical Mechanics exercises	Tutorial 7 Analizing MD: GROMOS Tutorial	Tutorial 8: Analizing MD: Students Plans	Tutorial 9: Analizing MD: Students Plans	How to prepare a barbecue
16:00	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	Coffee Break	
17:30	End of session	End of session	End of session	End of session	End of session	End of session	End of session	End of session	End of session	
20:30	Free time	Free time	Free time	Free time	Beer, Science & Friendship Good Stock Bar	Free time	Free time	Free time	Free time	Farewell

Applied Statistical Thermodynamics

7. Analysis

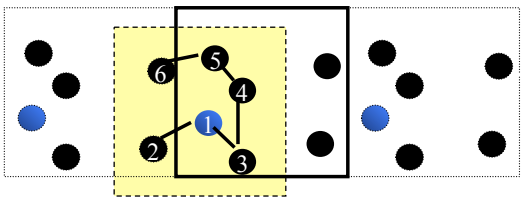
- What properties are we typically interested in?

- **Structural properties**
  - occurence of hydrogen bonds
  - root-mean-square deviations from the initial structure
  - agreement with NMR experiments (NOE analysis),
  - clustering of observed conformations
- **Energetic properties**
  - interaction energy between components of the system
  - free energy of binding, folding, ...
- **Properties of solutions**
  - density
  - viscosity
  - diffusion
- **Correlations, Distributions**
- **...**

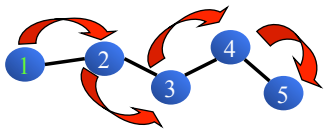
Applied Statistical Thermodynamics

Boundaries and gathering

- Periodic boundary conditions are commonly used
- The molecule may seem to be split by the box boundary



- Analysis programs use the nearest image convention to calculate distances:
  - Take the closest periodic copy
- For visualization and some analyses, it is important to reconstruct the molecule: gathering
  - Take subsequent nearest images
  - Follow atom number, or the bonds



□□□□□□□□□□□□□□□□□□□□

box shape *gather method*

Selecting the proper gather method can be quite a challenge!

Atomspecifiers

- Often you need to use specific atoms for an analysis
- Atomspecifiers
  - consists of one string (no spaces!)
  - examples:

□□□	third atom of the second molecule
□□□□□□□	atoms 3,4,5,6 and 8 of the first molecule
□□□□□□ □□□□	atoms 4 to 12 of the solvent
□□□□□	atoms 1 and 5 of all molecules
□□□□	all atoms with name “CA” in molecule 1
□□□□□□□□□□□	CA atoms of residue 4 in molecule 1
□□□□	all atoms with a name starting with “O” in molecule 1
□□□□□□□□	atoms 1 to 7 and all “CA” of molecule 2
□□□□	all “OW” atoms in the solvent

- Use the program □□□□□□□□ to check what the atomspecifier does

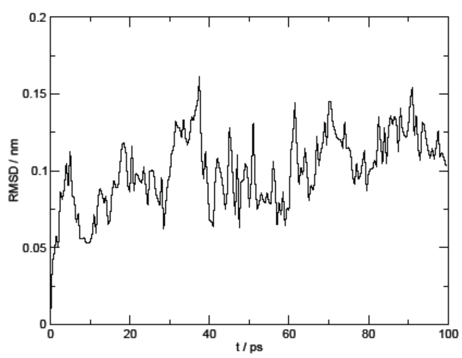
  □□□□□□□□□ □□□□□□□□□□□□□□□□□□□□□□□□□□□
  □□□□□□□□□□□□□□□□□□

Applied Statistical Thermodynamics

7. Analysis: RMSD

- Atom-positional root-mean-square deviations (RMSD)
- Measure of structural deviation between two configurations
  - reference configuration: e.g. X-ray or NMR structure

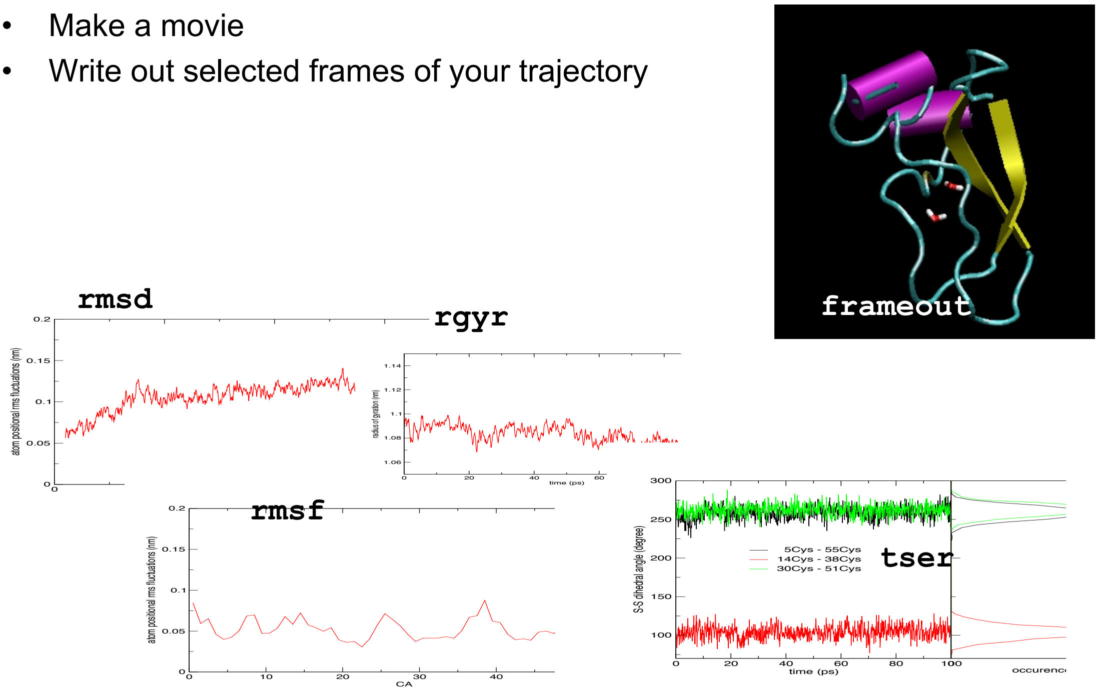
□□□□ □□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□*atoms to consider for rmsd*
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□□□□□□□



- Similarly, program □□□□ calculate the atom-positional root-mean-square fluctuations around the average position for selected atoms

Basic analysis

- Atom positional root-mean-square fluctuations
- Radius of gyration
- Monitor time series of structural properties
- Make a movie
- Write out selected frames of your trajectory

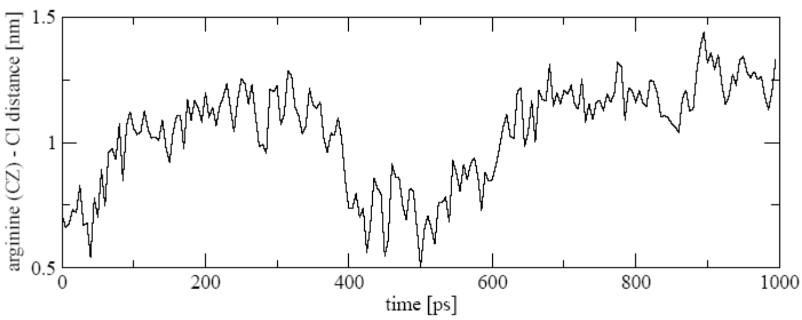


Applied Statistical Thermodynamics

7. Analysis: time series

- Time series of different properties, given by the property specifier

[illegible] [illegible]
[illegible]
[illegible]
[illegible] *distance*
[illegible] *angle*
[illegible] *torsional angle*
[illegible]
[illegible]
[illegible]
[illegible]



7. Analysis: Hydrogen bonds

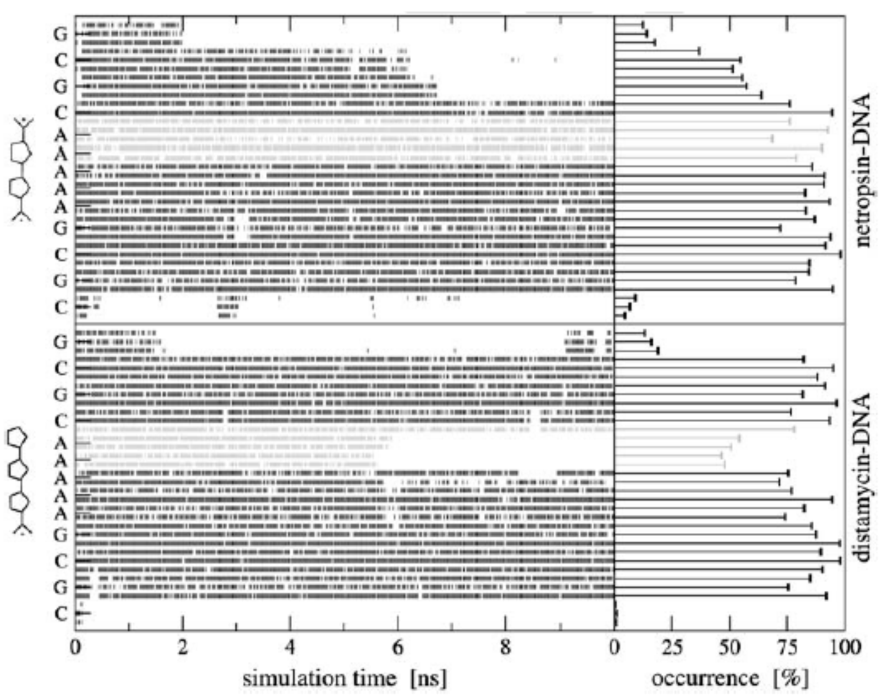
- We can monitor the occurrence of hydrogen bonds
- Defined using
  - the atomspecifiers for the donors and acceptors
  - geometric criteria, e.g.:
    - H...acceptor distance maximally 0.25 nm
    - donor...H...acceptor angle minimally 135°

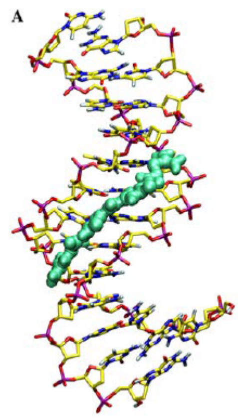
this gives us all hydrogen bonds between molecules 1 and 2

filters for relevant atoms

7. Analysis: Hydrogen bonds

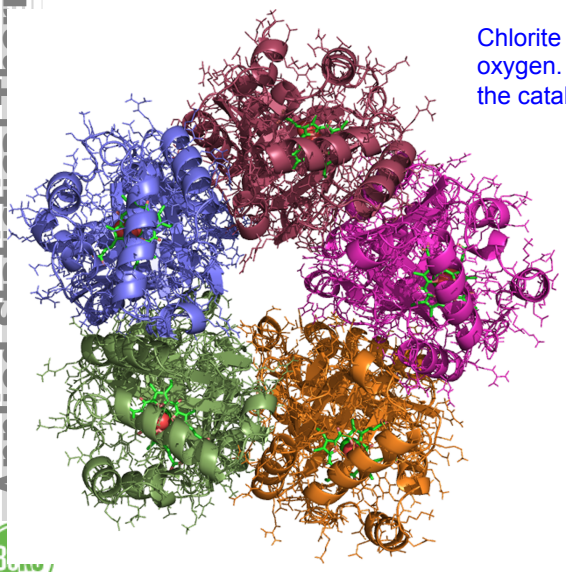
- Example: monitoring the Watson-Crick hydrogen bonds in a DNA-ligand complex



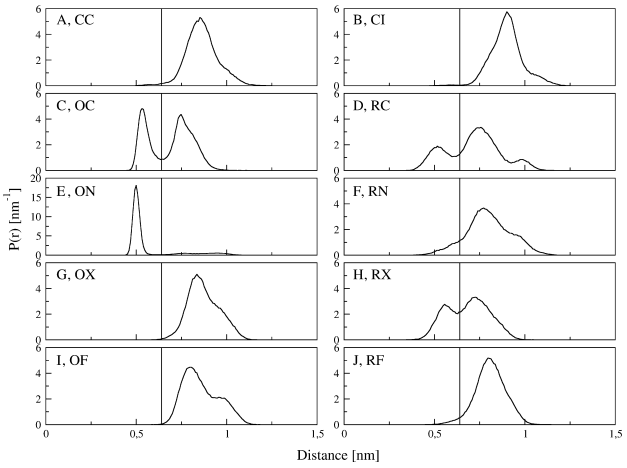


Measuring distances, distributions

- We can monitor the distances between molecules as a function of time
- Or represent them as a distribution

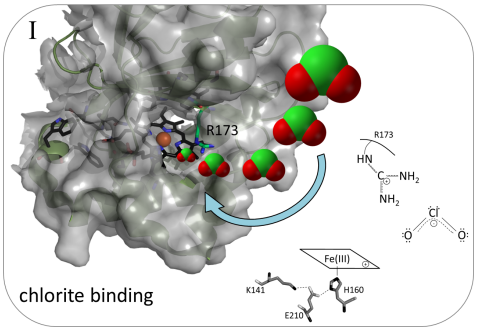


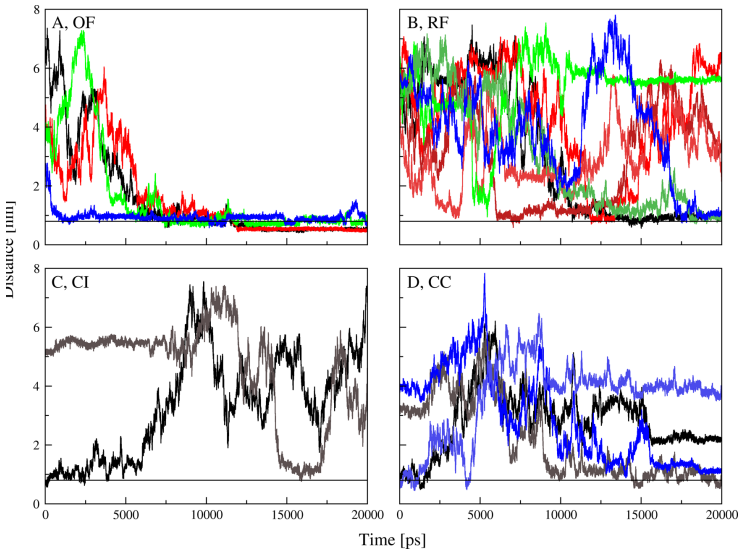
Chlorite dismutase converts chlorite into chloride and molecular oxygen. Here we monitor the distance between various ions and the catalytic heme group.



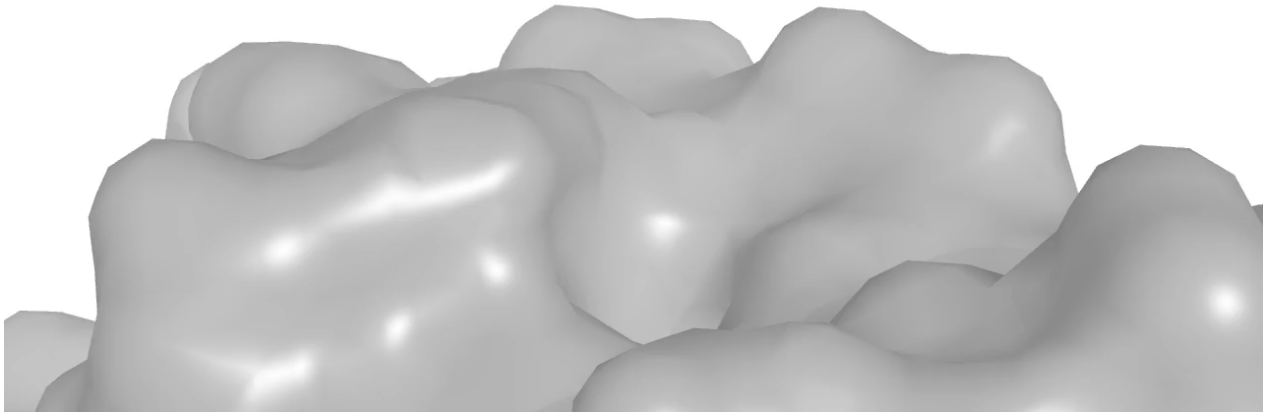
Spontaneous binding events

- In some simulations of the chlorite dismutases we observed spontaneous binding



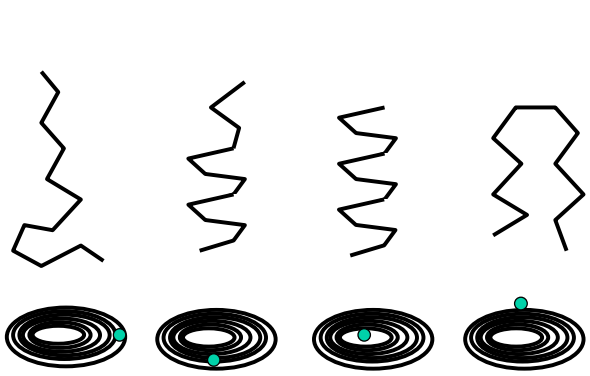


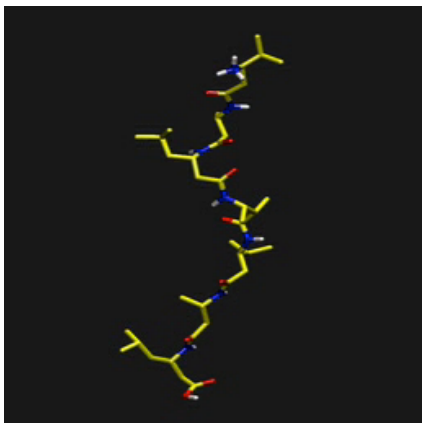
Sünderman et al. Biochemistry (2014) 53:4869-4879

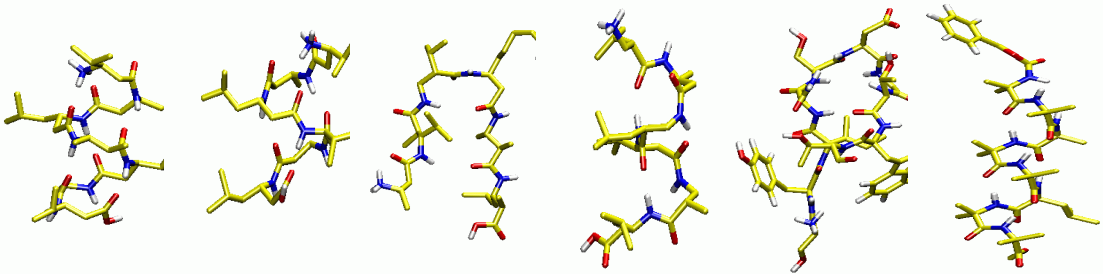


Clustering of conformations

- Extract mostly populated conformations from simulation



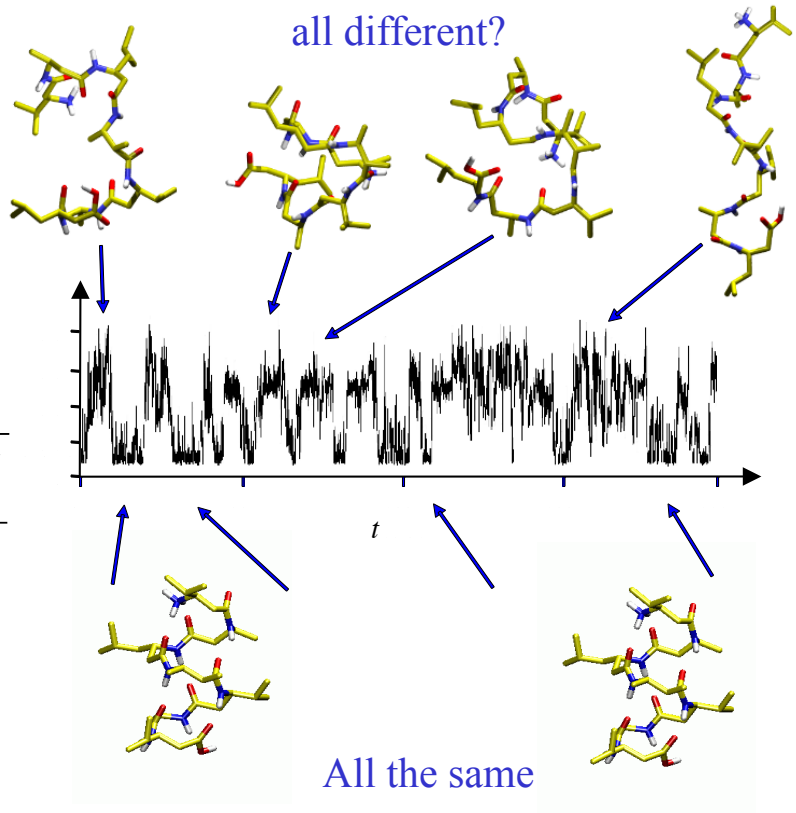




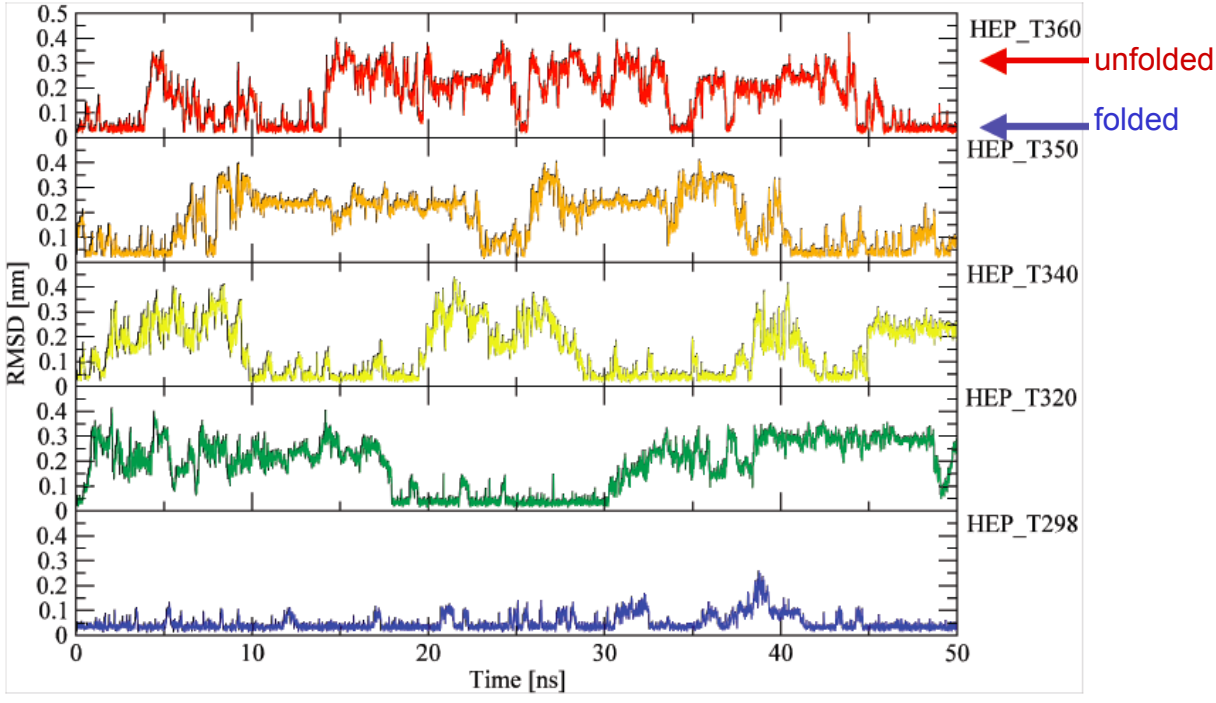
RMSD: Root mean square deviation

- A measure to compare two structures
- Here we compare the structures seen in the simulation to the experimentally determined 'folded' structure

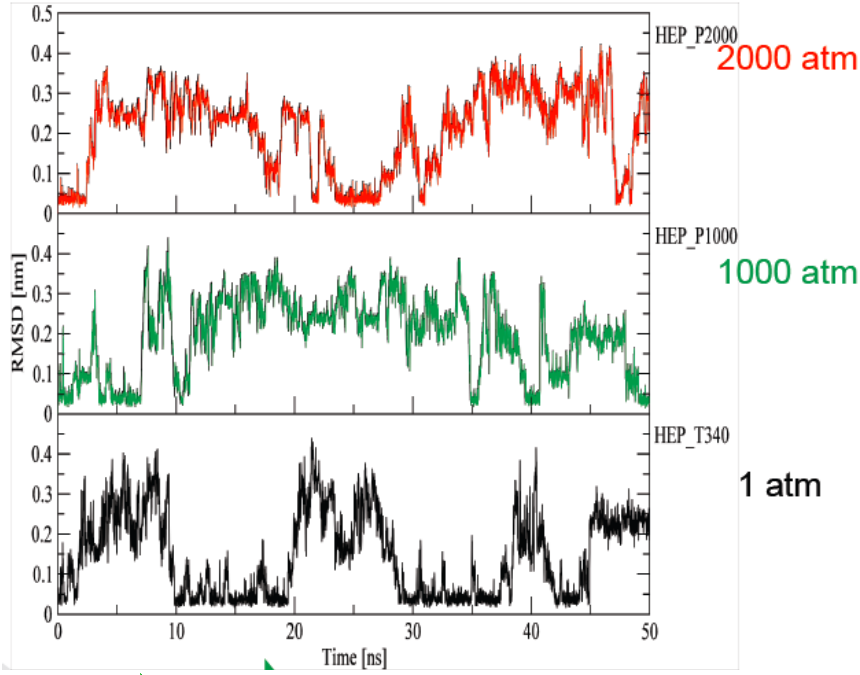
$$D = \sqrt{\frac{\sum_i \left|\mathbf{q}_i - \mathbf{q}_i^{ref}\right|^2}{N}}$$



Temperature dependency



Pressure dependency



folding equilibrium depends on pressure

Applied Statistical Thermodynamics

7. Analysis: structural clustering

- map the structures sampled during the simulations onto a set of generic configurations
- three steps: calculation of an rmsd matrix, clustering, analysis

1. gromos++ program [illegible] calculates the atom-positional root-mean-square deviation between all pairs of structures in a trajectory file:

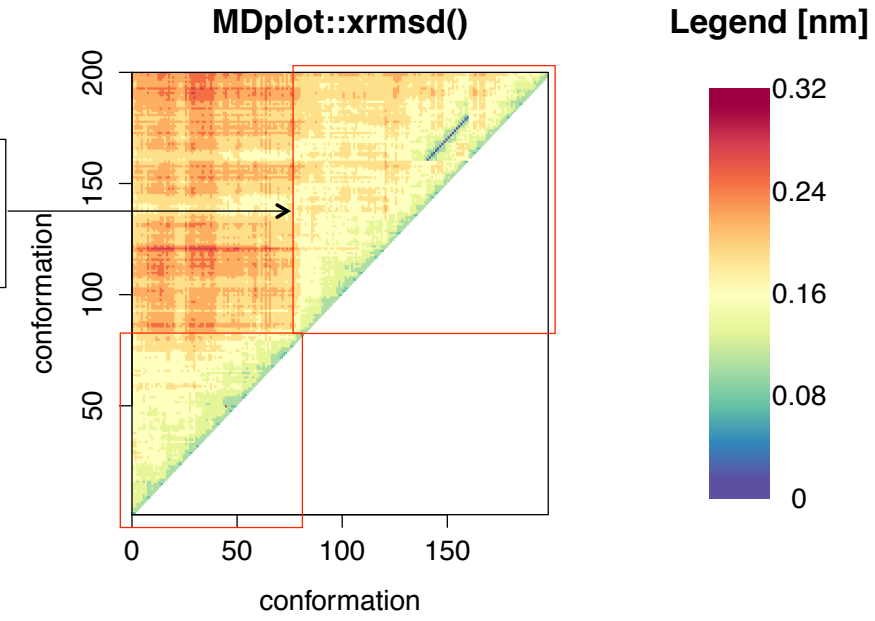
[illegible] [illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]
[illegible]

- writes the output file [illegible] (large file)

Conformational clustering

- Compare all snapshots in a simulation against each other
- Very many comparisons: big files

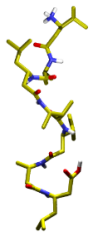
Special algorithms to identify groups of structures that look similar



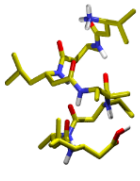
7. Analysis: structural clustering

2. gromos++ program □□□□□□□ performs conformational clustering based on the similarity matrix, rmsdmat.bin.
   - The structures with RMSD smaller than the cutoff are part of the same cluster

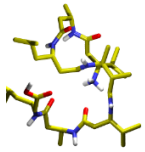
□□□□□□□ □□□□□□□□□□□□□□□□□□□□
□□□□□□□□□□□□
□□□□□□□□□□□□



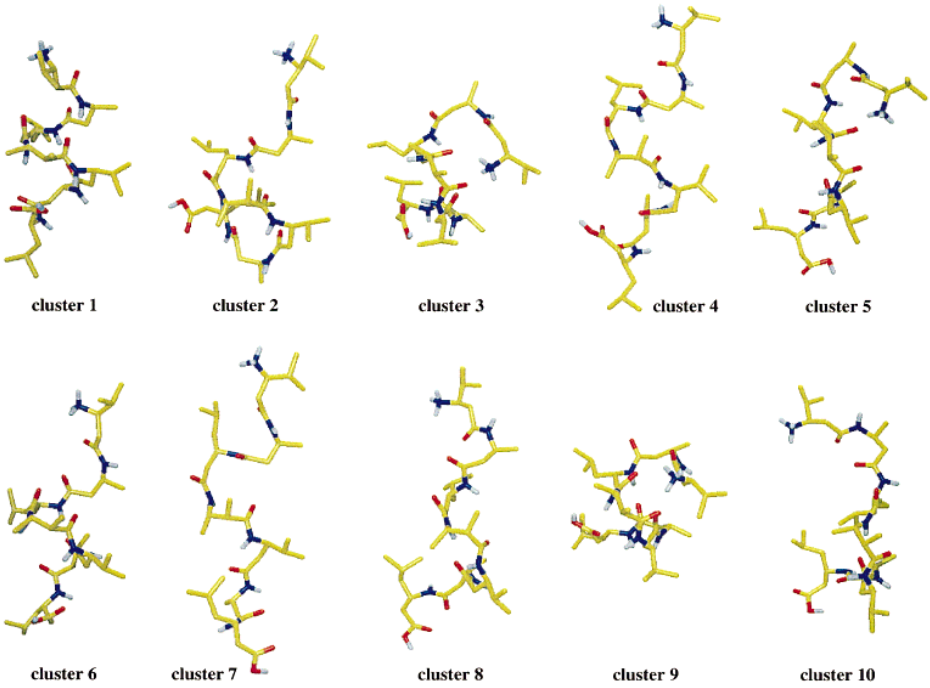
   - Gives output with cluster sizes, number of clusters, representative structures, etc.



3. gromos++ program □□□□□□□□□□□ can do additional analysis on the output of the □□□□□□□ program
   - Life-time analysis
   - Filtering of trajectories into the clusters
   - Mark cross clustering over multiple trajectories



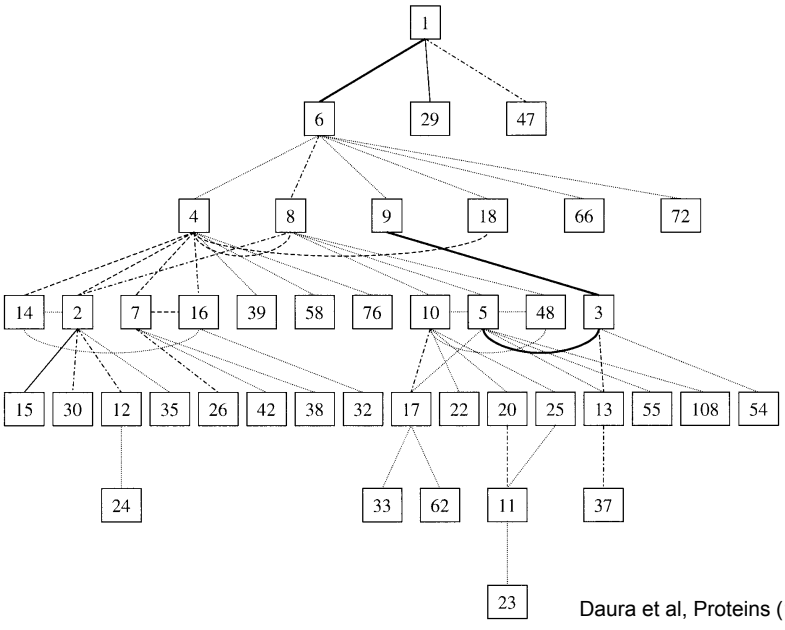
Conformational clustering



Applied Statistical Thermodynamics

Conformational clustering

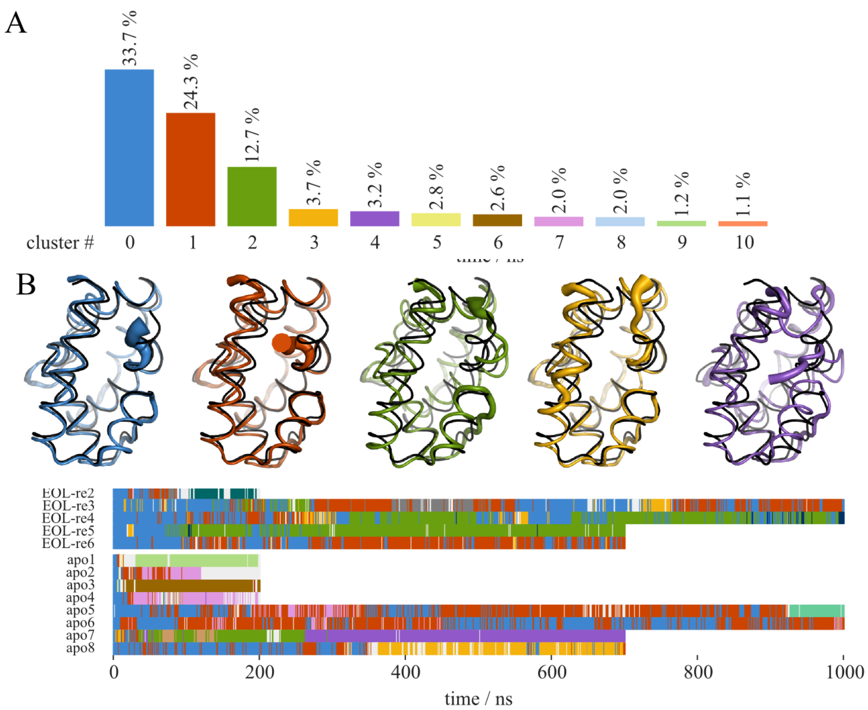
- Visualize the folding pathways of the peptide



Daura et al, Proteins (1999) 34:269-280

Conformational clustering

another example



Pechlaner and Oostenbrink, J. Chem. Inf. Model (2015) 55:2633-2643

7. Analysis: time-series of energies

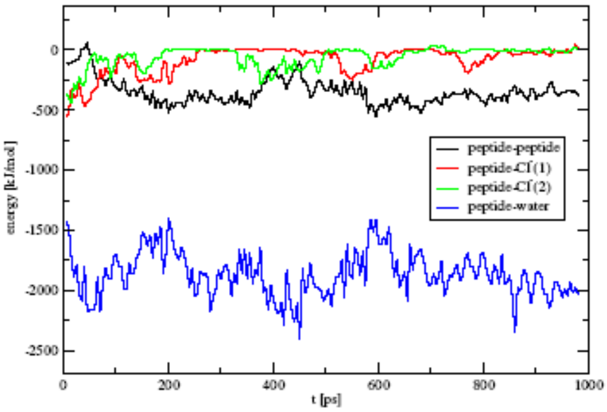
- GROMOS can write (free) energies to separate trajectory files for later analysis
- program extracts time series of different properties from the energy trajectories
- the properties are defined in the library

this identifies e.g. the second value in the block to the variable

7. Analysis: time series of energies

- Use the program together with the library file

- Writes summary to output, and time series to files



Applied Statistical Thermodynamics

8. Error messages

- Read the message
  - does it mean anything to you?
- Check your input
  - input flags, paths to the files, atomspecifiers, file names)
- Check your topology
  - are you using the correct one?
- Check your topology in combination with the coordinates
  - can other programs read this combination?
- Check your input, again(!)

Did you make manual changes to the topology or the trajectory?
dangerous, try to avoid

- This all did not help?
- *Maybe something is wrong with the program. Catch the bug!*