Molecular mechanics and dynamics

Molecular mechanics

- Force fields
- Energy minimization
- Conformational analysis
- Conformational searching
- Thermodynamic *versus* kinetics
- Systematic search
- Monte Carlo
- Genetic algorithm

Molecular dynamics

- Newton's 2nd Law of Motion
- Running a simulation
- Analysis
- Applications

Force field

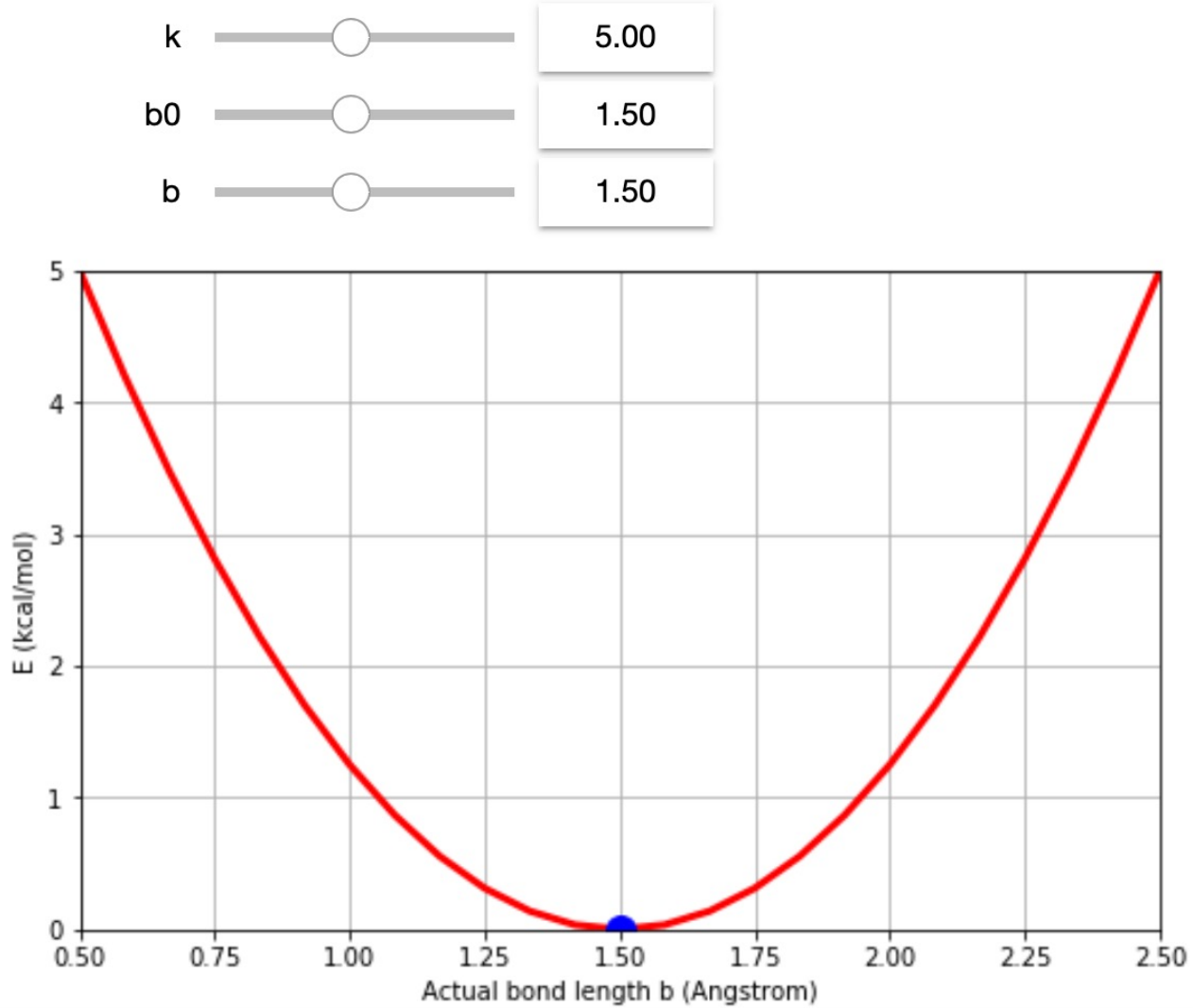
$$E_{total} = E_{bonded} + E_{nonbonded}$$

$$E_{bonded} = E_{bonds} + E_{angles} + E_{dihedrals}$$

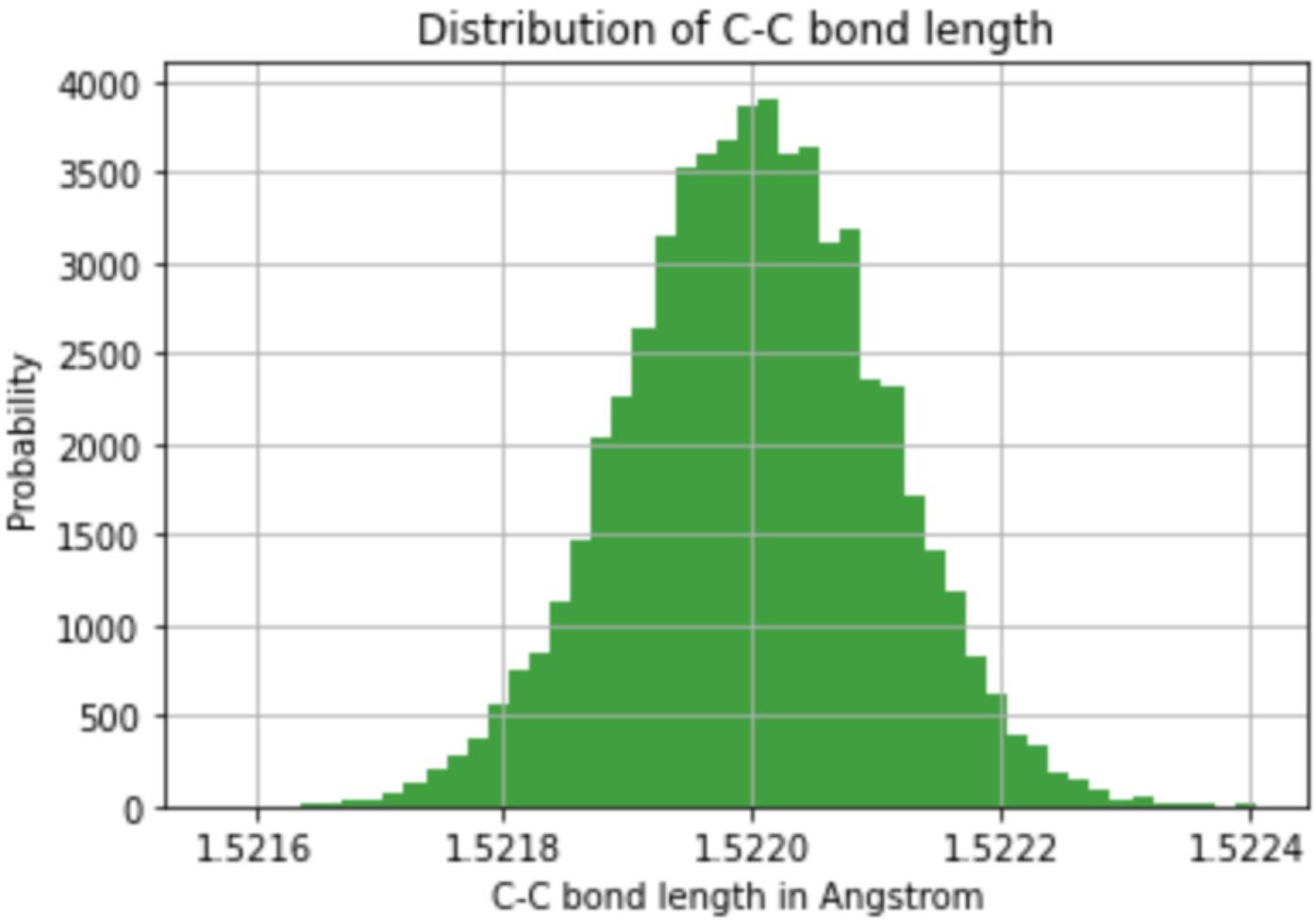
$$E_{nonbonded} = E_{electrostatic} + E_{vdw}$$

$$E_{bonded} = E_{bonds} + E_{angles} + E_{dihedrals}$$

$$E_{bonds} = \sum_{allbonds} k(b - b_0)^2$$

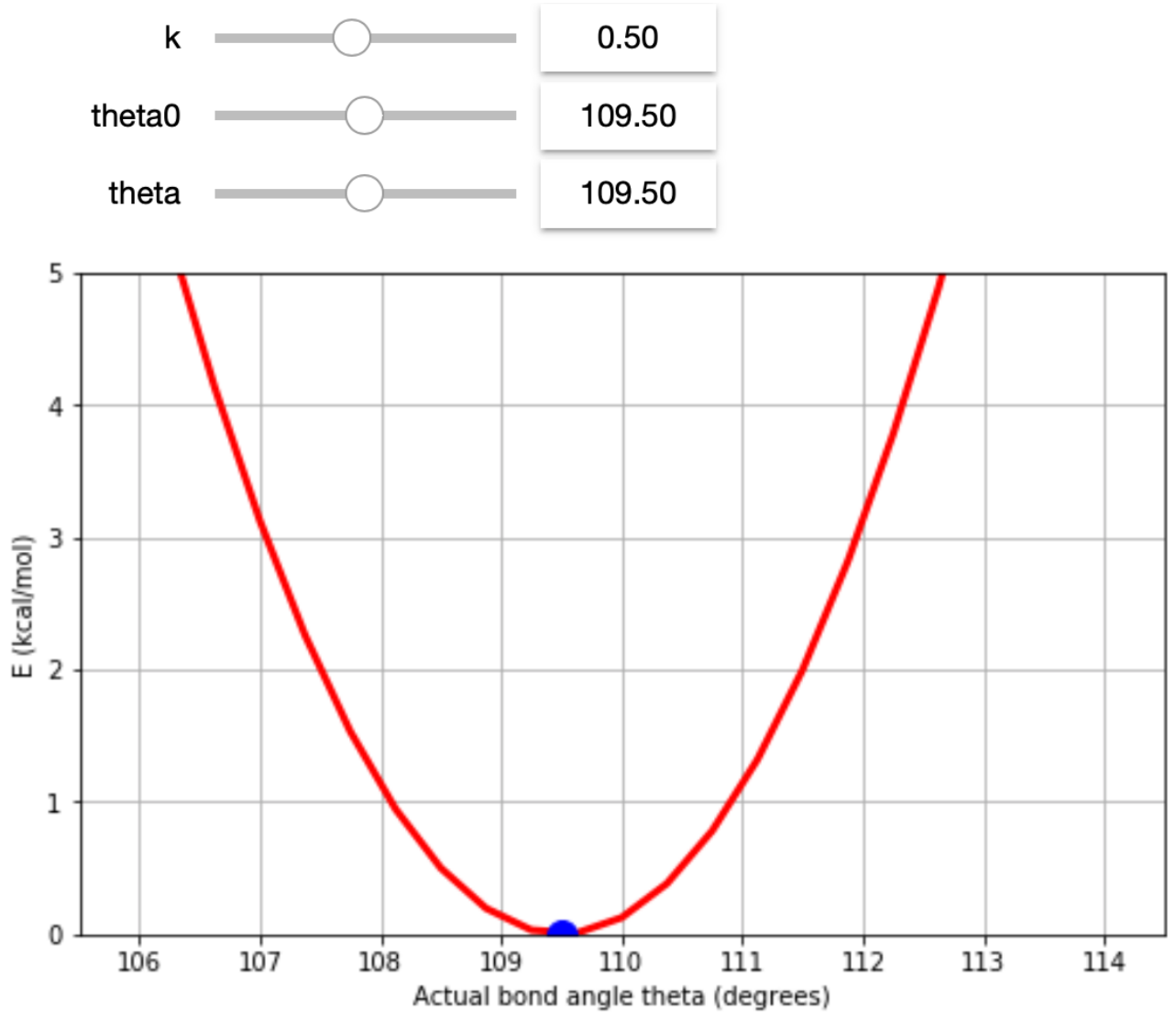


Bond stretching



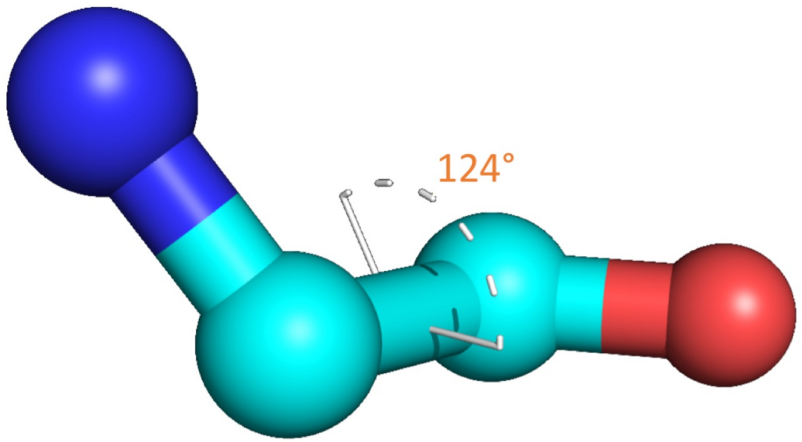
$$E_{bonded} = E_{bonds} + E_{angles} + E_{dihedrals}$$

$$E_{angles} = \sum_{allangles} k(\theta - \theta_0)^2$$



$$E_{bonded} = E_{bonds} + E_{angles} + E_{dihedrals}$$

$$E_{dihedrals} = \sum_{alldihedrals} k(1 + \cos(n\phi - \delta))$$

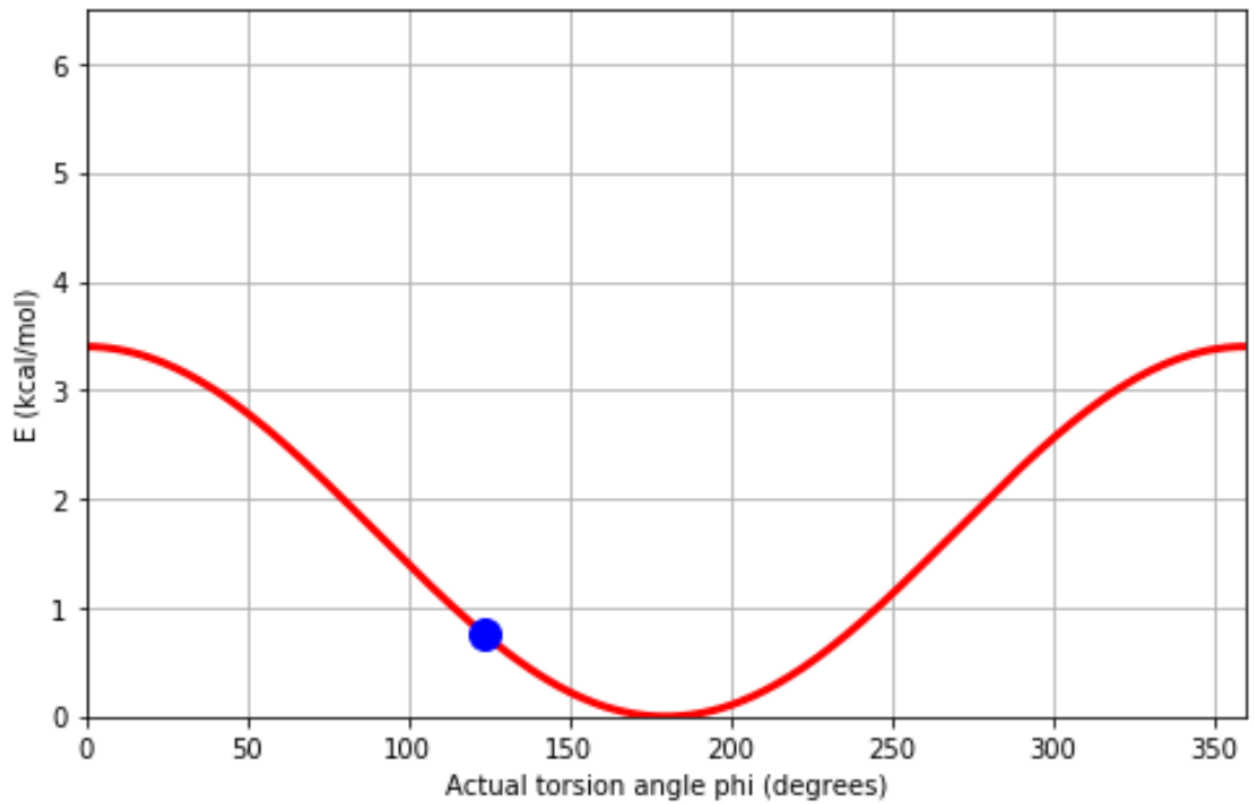


k 1.70

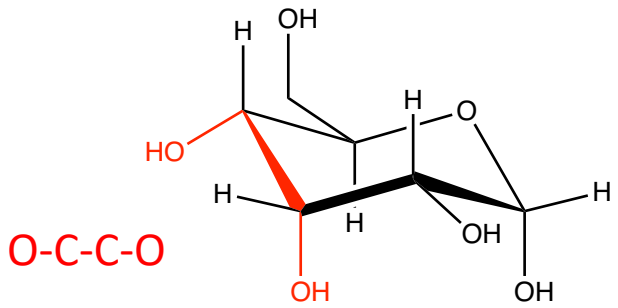
phi 124

delta 0

n 1



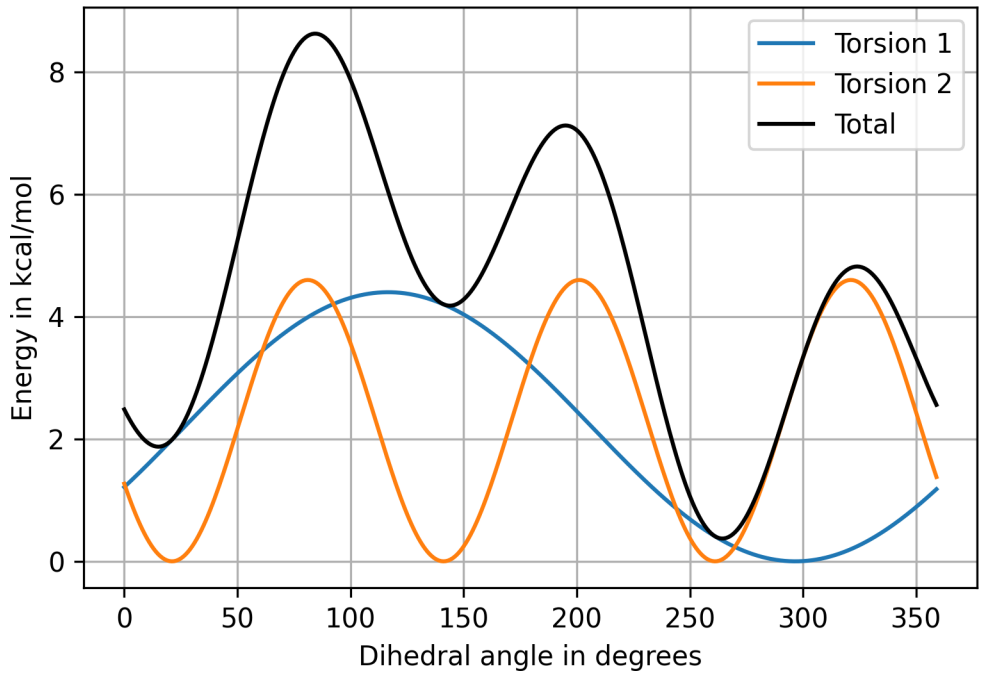
Torsion example



$$k\left(1 + cos\left(n\phi - \delta\right)\right)$$

	k (kcal/mol)	n	δ (degrees)
Torsion 1	2.2	1	-90
Torsion 2	2.3	3	+90

Gromos 54A6 parameters

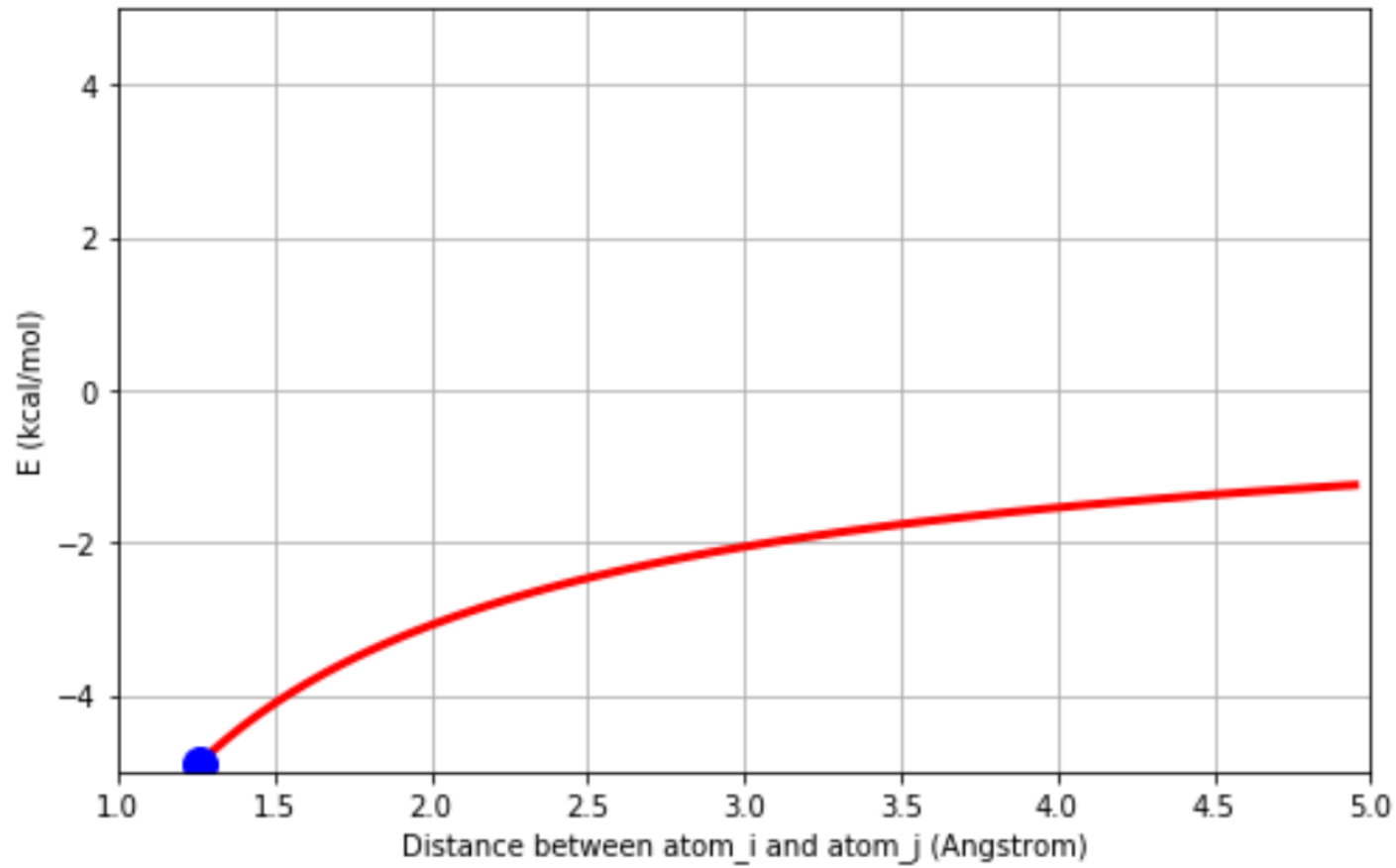


$$E_{nonbonded} = E_{electrostatic} + E_{vdw}$$

$$E_{electrostatic} = \sum_{i}^{natoms} \sum_{j}^{natoms} \frac{q_i q_j}{kDr_{ij}}$$

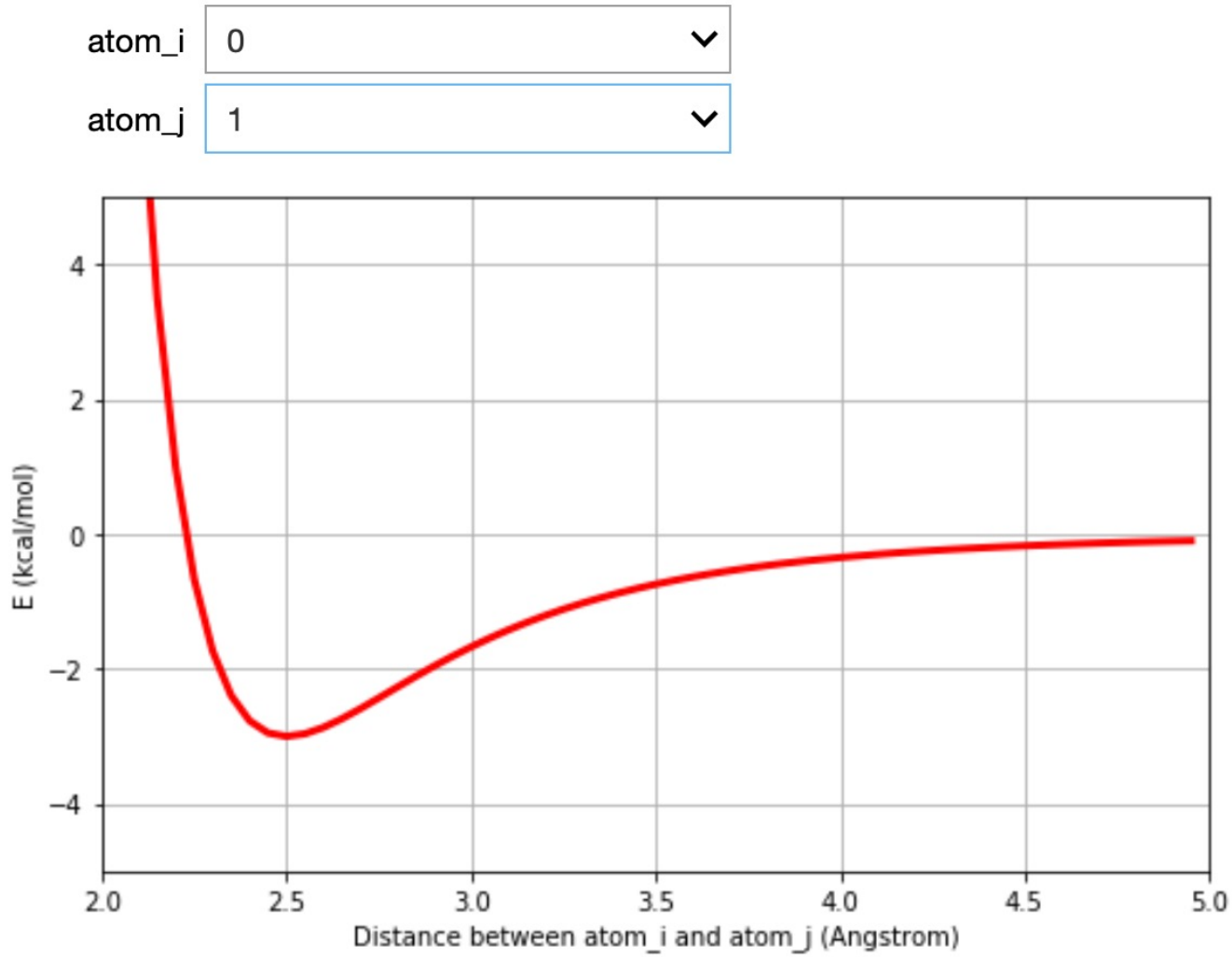
atom_i 2 (charge: -0.28)

atom_j 1 (charge: +0.22)



$$E_{nonbonded} = E_{electrostatic} + E_{vdw}$$

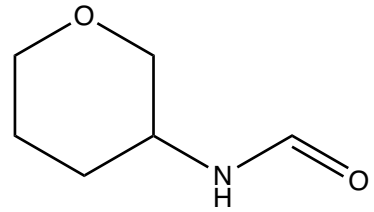
$$E_{vdw} = \sum_i^{natoms} \sum_j^{natoms} \epsilon_{ij}[(\frac{R_{min,ij}}{r_{ij}})^{12} - 2(\frac{R_{min,ij}}{r_{ij}})^{6}]$$



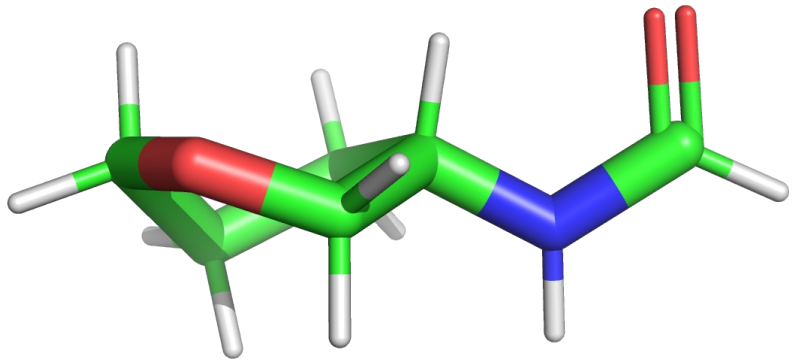
Converting 2D to 3D

Distance geometry approach in RDKit:

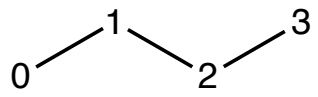
1. The molecule's distance bounds matrix is calculated based on the connection table and a set of rules.
2. The bounds matrix is smoothed using a triangle-bounds smoothing algorithm.
3. A random distance matrix that satisfies the bounds matrix is generated.
4. This distance matrix is embedded in 3D dimensions (producing coordinates for each atom).
5. The resulting coordinates are cleaned up somewhat using a crude force field and the bounds matrix.
6. Finally, a **minimization step** removes energetically unfavorable conformations




```
[4]  from rdkit import Chem
     from rdkit.Chem import AllChem
     mol = Chem.MolFromSmiles("C1CCOCC1NC=O")
     mol =Chem.AddHs(mol)
     AllChem.EmbedMolecule(mol)
```



Distance matrix with lower and upper distances



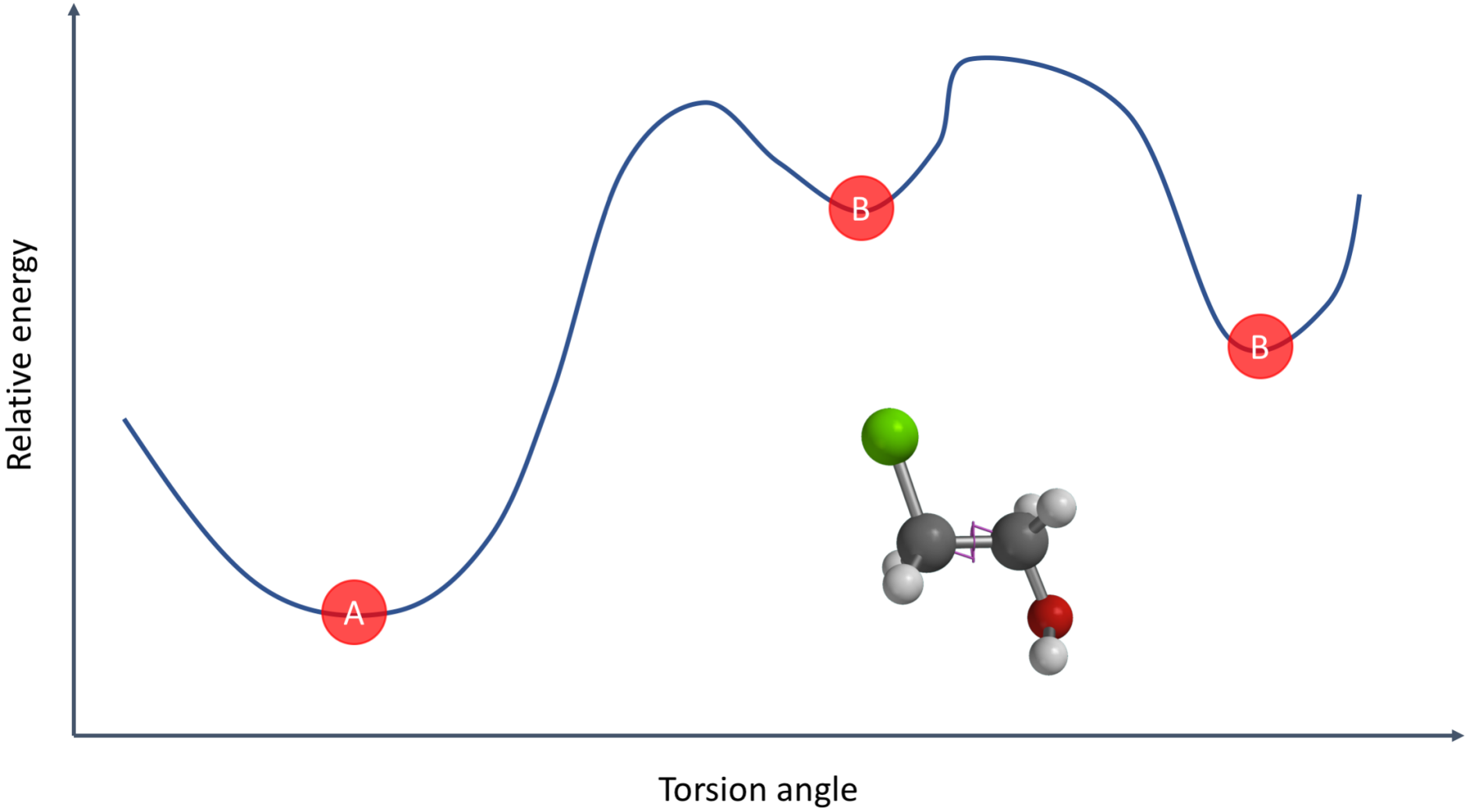
Each bond length is 1.5 Å
VDW radius of each atom is 1.8 Å

MIN	0	1	2	3
0	0	1.5	3.6	3.6
1	1.5	0	1.5	3.6
2	3.6	1.5	0	1.5
3	3.6	3.6	1.5	0

MAX	0	1	2	3
0	0	1.5	3.0	4.5
1	1.5	0	1.5	3.0
2	3.0	1.5	0	1.5
3	4.5	3.0	1.5	0

LIMITS	0	1	2	3
0	0	1.5	3.0/3.6	3.6/4.5
1	1.5	0	1.5	3.0/3.6
2	3.0/3.6	1.5	0	1.5
3	3.6/4.5	3.0/3.6	1.5	0

Energy minimzation



Steepest descent versus conjugate gradient method for energy minimization

Initial minimization (gradient < 1 kcal/A2)

Method	CPU time(s)	Number of iterations
Steepest descent	67	98
Conjugate gradients	149	213

Stringent minimization (gradient < 0.1 kcal/A2)

Method	CPU time(s)	Number of iterations
Steepest descent	1,405	1,893
Conjugate gradients	257	367

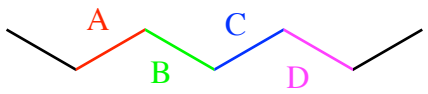
Conformational analysis

- Thermodynamic *versus* kinetics
- Systematic search
- Monte Carlo
- Genetic algorithm

Conformational analysis

- Thermodynamic *versus* kinetics
- Systematic search
- Monte Carlo
- Genetic algorithm

Systematic search: looping over all torsions



- Rotate torsion **D** from 0° to 360° in steps of 10°
  - Rotate torsion **C** from 0° to 360° in steps of 10°
    - Rotate torsion **B** from 0° to 360° in steps of 10°
      - Rotate torsion **A** from 0° to 360° in steps of 10°

CALCULATE ENERGY

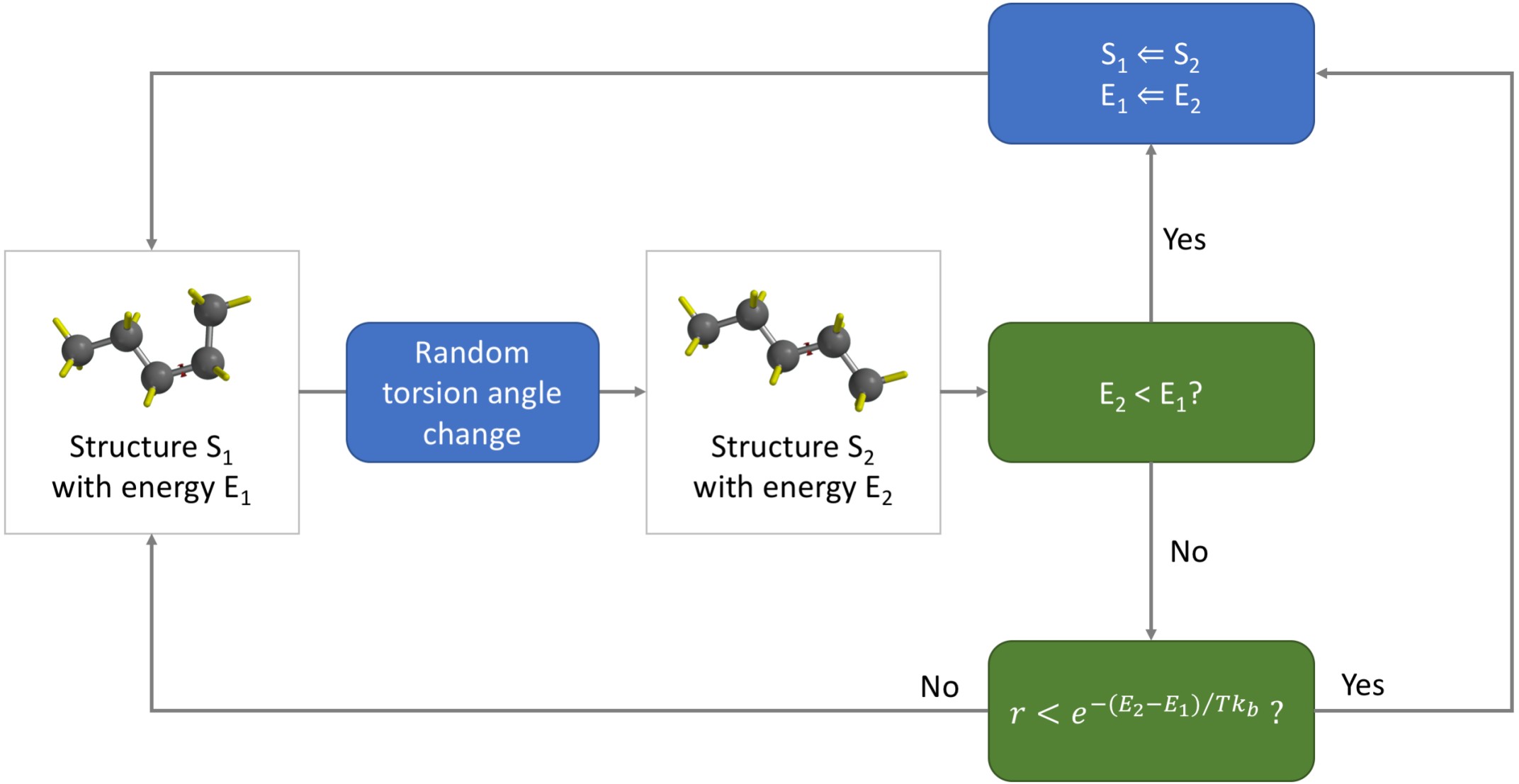
Systematic search

	Number of torsions			
Angle increment	**5**	**10**	**20**	**40**
30	1	9.1×10^{5}	2.1×10^{17}	3.2×10^{39}
15	3.2×10^{1}	9.1×10^{8}	2.2×10^{23}	3.5×10^{51}
8	7.4×10^{2}	5.0×10^{11}	6.5×10^{29}	2.9×10^{62}
4	2.4×10^{4}	5.0×10^{14}	6.8×10^{35}	3.2×10^{72}
2	7.6×10^{5}	5.3×10^{17}	7.1×10^{40}	3.5×10^{86}

Conformational analysis

- Thermodynamic *versus* kinetics
- Systematic search
- Monte Carlo
- Genetic algorithm

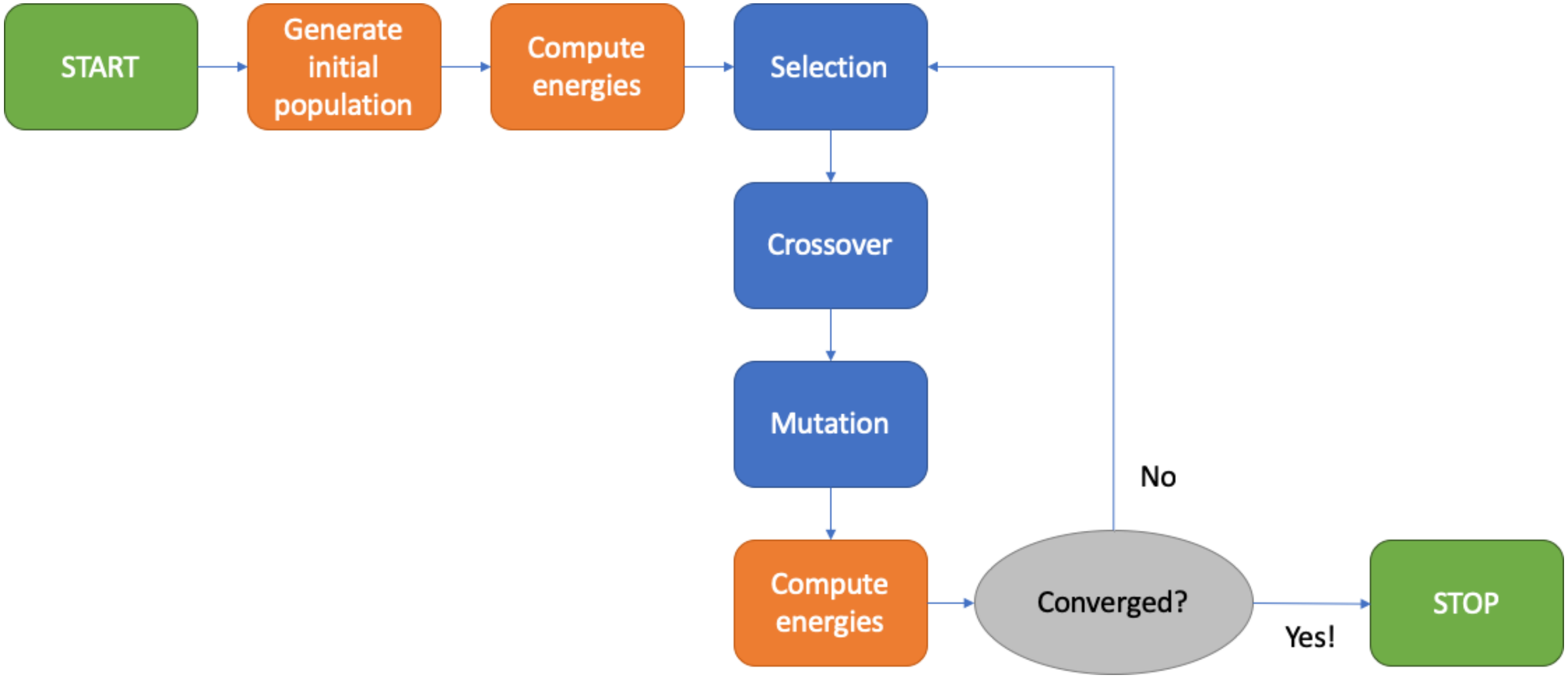
Metropolis Monte Carlo



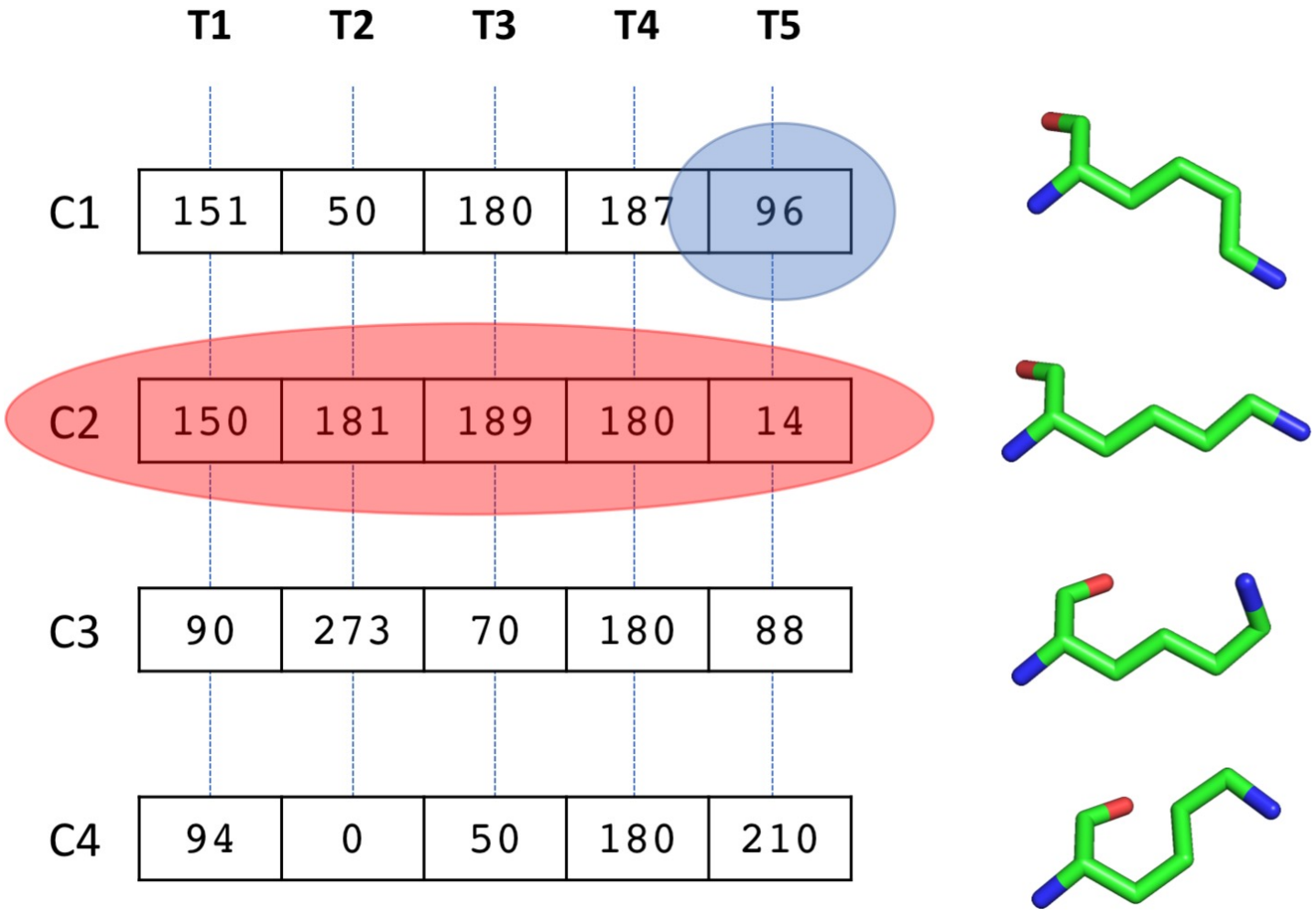
Conformational analysis

- Thermodynamic *versus* kinetics
- Systematic search
- Monte Carlo
- Genetic algorithm

Genetic Algorithm



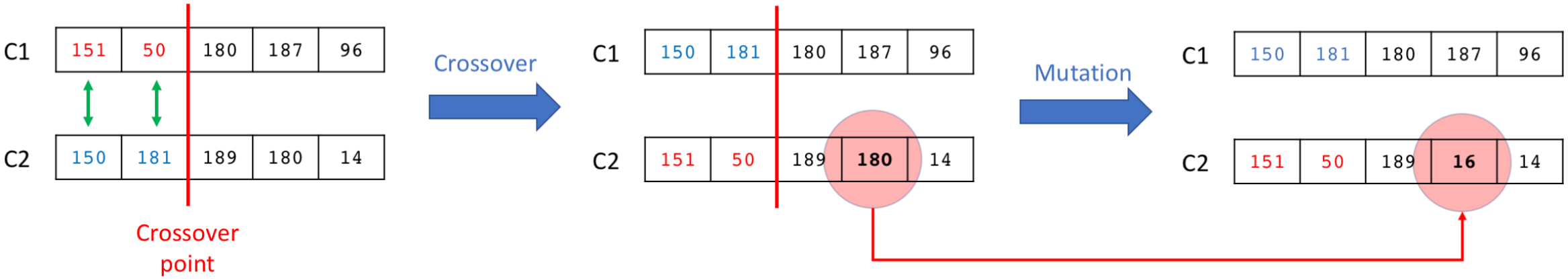
Initial population represented by chromosomes and genes



Five phases

- Initial population
- Fitness function
- Selection
- Crossover
- Mutation

Crossover and mutation



Molecular mechanics and dynamics

Molecular mechanics

- Force fields
- Energy minimization
- Conformational analysis
- Conformational searching
- Thermodynamic *versus* kinetics
- Systematic search
- Monte Carlo
- Genetic algorithm

Molecular dynamics

- Newton's 2nd Law of Motion
- Running a simulation
- Analysis
- Applications

Popular MD packages

- Gromacs
- Amber
- NAMD

Different steps in a MD simulation

1. Calculation of the force on each atom, using the force field parameters
2. Calculation of the acceleration on each atom, using the forces from step 1
3. Repositioning of each atom according the acceleration from step 2
4. Repeat, or stop

Step 1: calculation of the force on each atom

$$F = -\nabla V(\vec{r})$$

Bonds: $F_b = -2k(b - b_0)$

Angles: $F_a = -2k(\theta - \theta_0)$

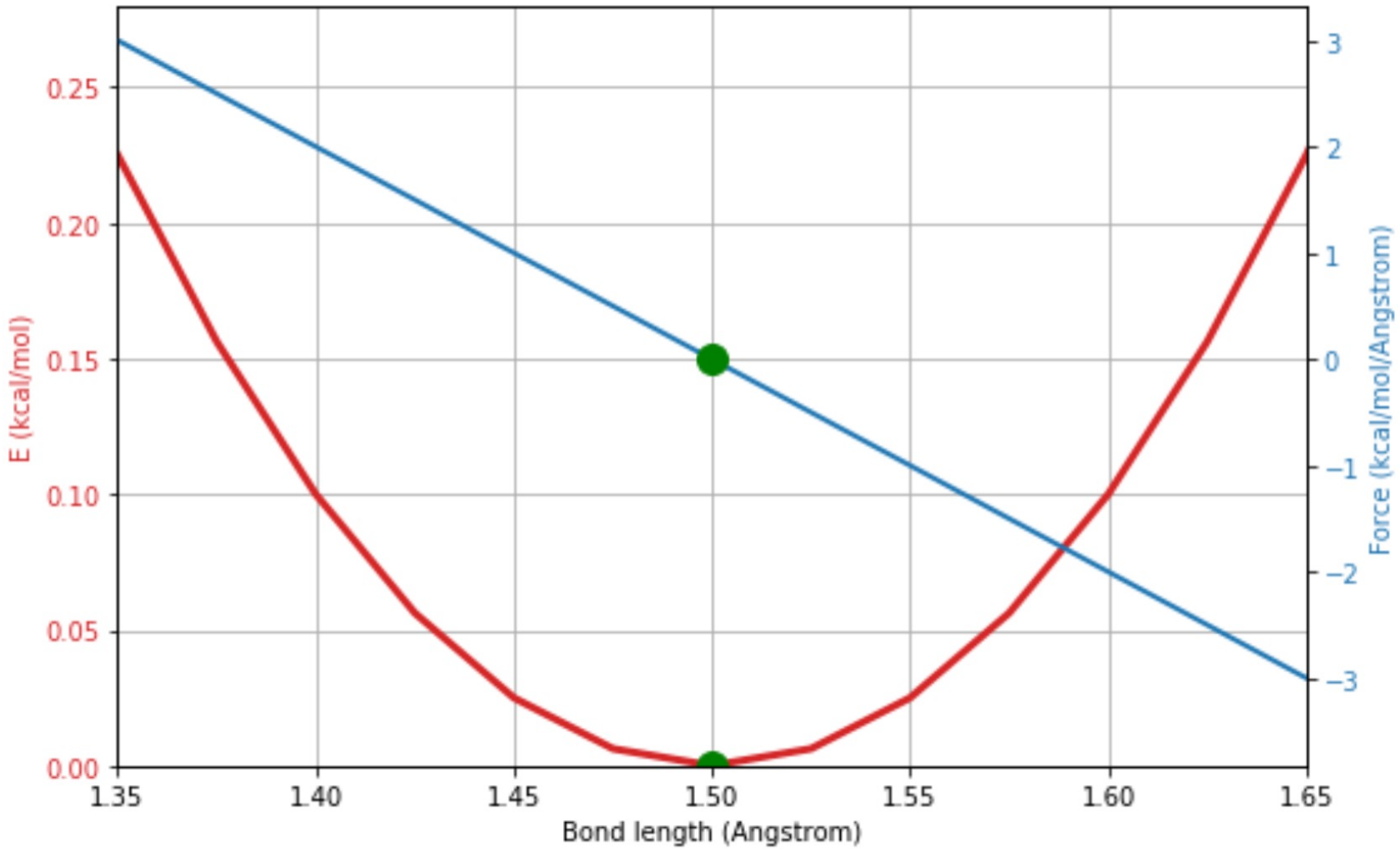
Dihedrals: $F_d = kn\sin(n\phi - \delta)$

Electrostatics: $F_e = \frac{q_i q_j}{Dkr_{ij}^2}$

VDW: $F_{vdw} = \frac{-12\epsilon_{ij}R^6_{min,ij}(r_{ij}^6 - R_{min,ij}^6)}{r_{ij}^{13}}$

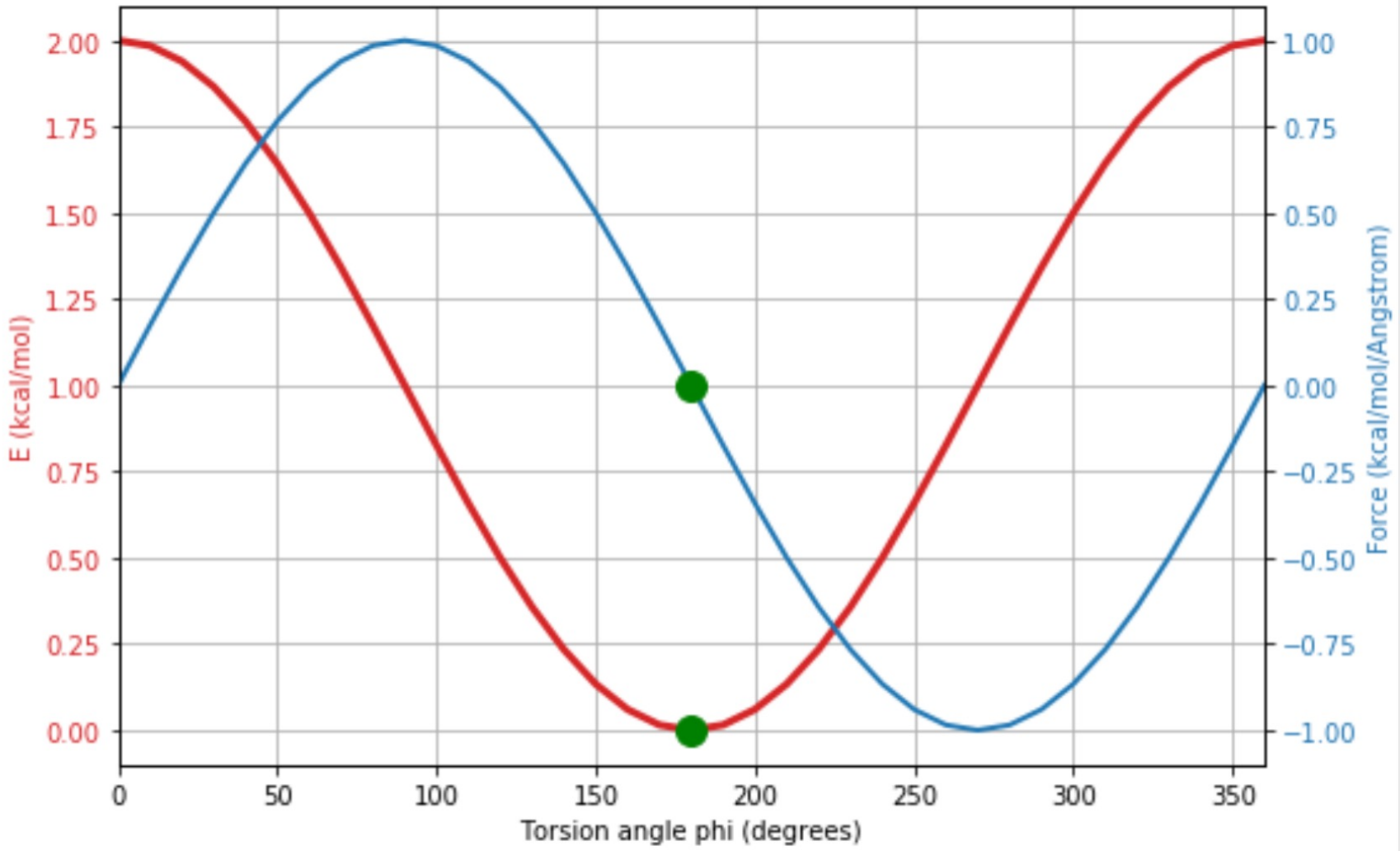
Forces on the bonds: $F_b = -2k(b - b_0)$

b 1.50



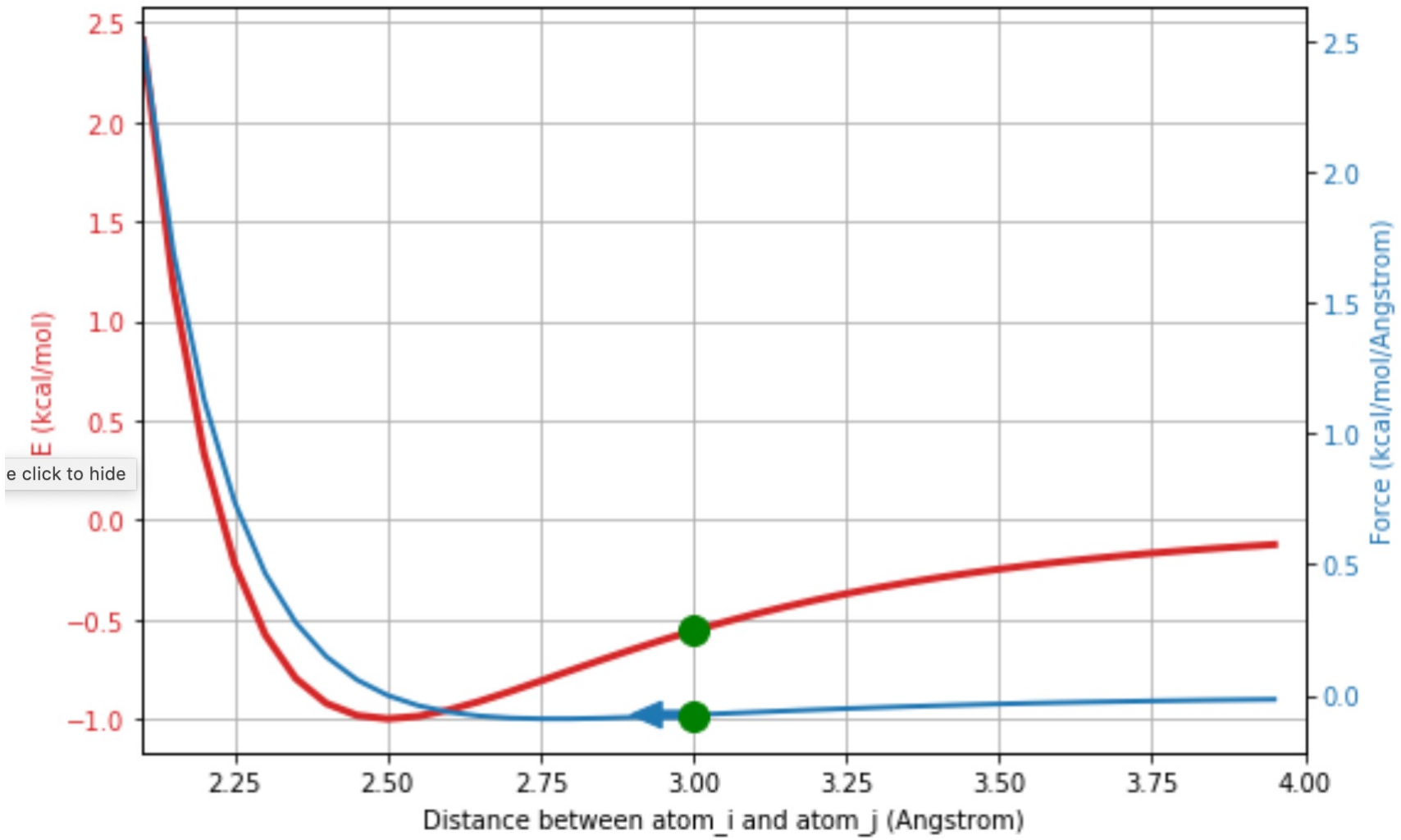
Forces on the torsion angles: $F_d = kn\sin(n\phi - \delta)$

phi 180



VDW forces: $F_{vdw} = \frac{-12\epsilon_{ij}R^{6}_{min,ij}(r_{ij}{}^{6} - R_{min,ij}{}^{6})}{r_{ij}^{13}}$

d 3.00



Step 2: calculate the acceleration on each atom

$$\vec{a}_i = \frac{\vec{F}_i}{m_i}$$

$\vec{F}_i$ was calculated in step 1

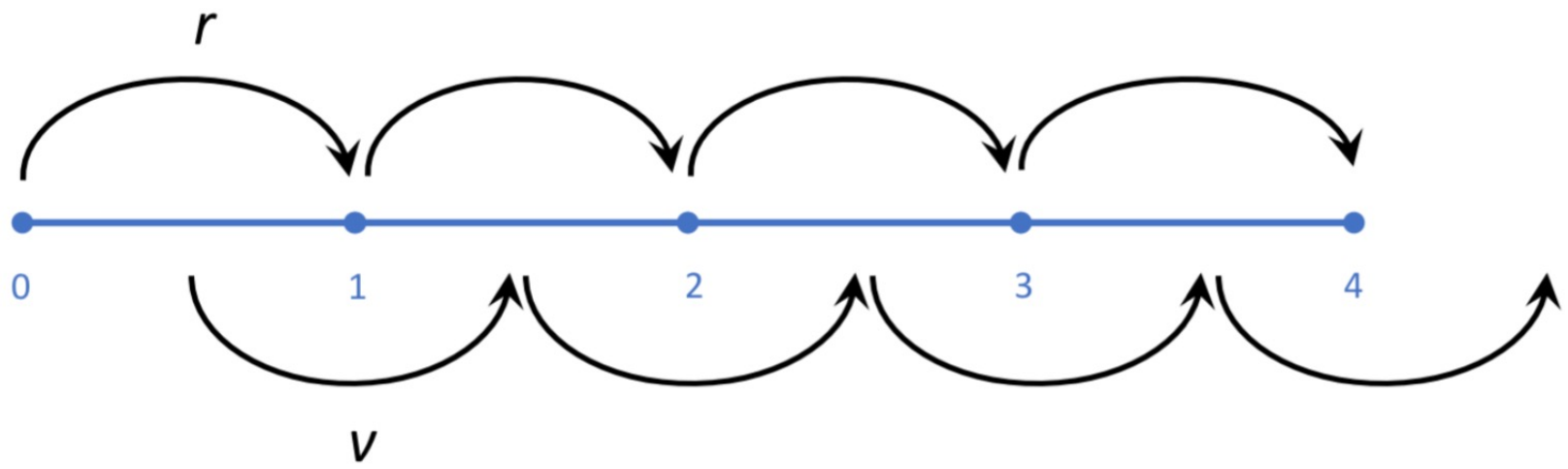
Step 3: repositioning of each atom according the acceleration from step 2

Method 1: Verlet algorithm:

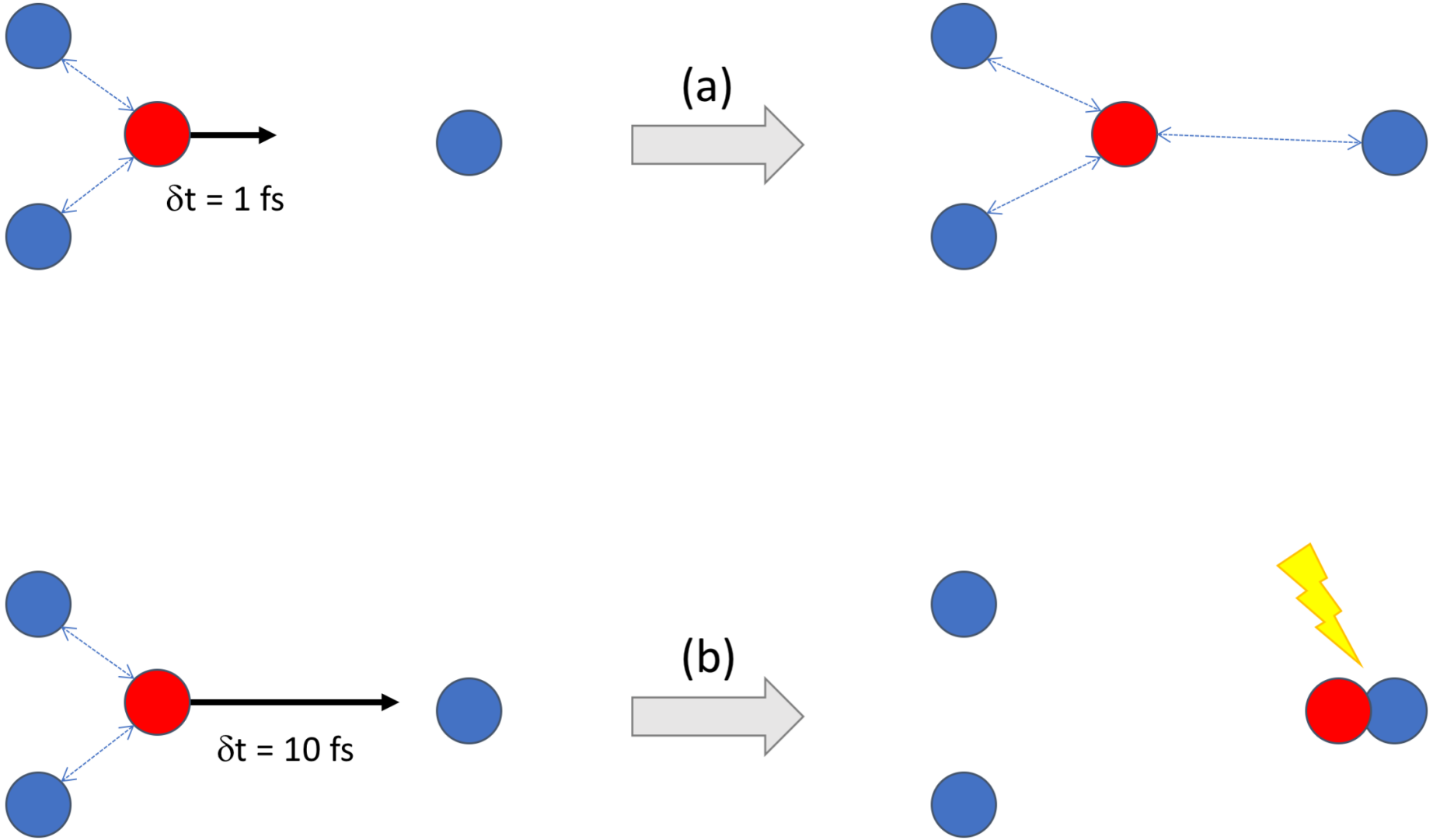
$$r(t+\delta t) = 2r(t) + \delta t^2 a(t) - r(t-\delta t)$$

Method 2: Leap-frog algorithm:

$$r(t+\delta t) = r(t) + \delta t v(t+\tfrac{1}{2}\delta t)$$
$$v(t+\tfrac{1}{2}\delta t) = v(t-\tfrac{1}{2}\delta t) + \delta t a(t)$$



Choosing an appropriate timestep

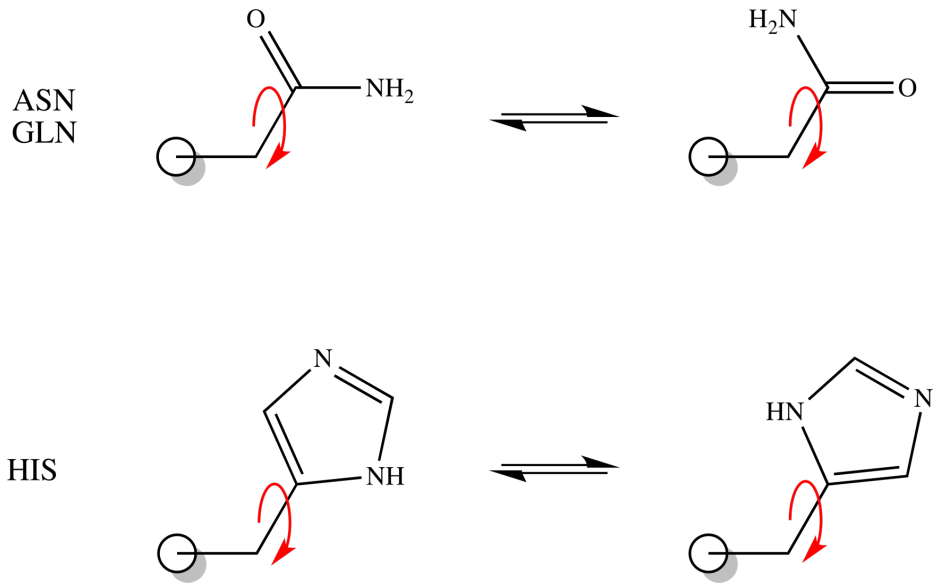


Running a simulation: setting up the protein and environment (part 1)

- Obtaining protein coordinates
  - Protein databank (PDB)
- Add missing loops
  - ModLoop
  - RAPPER
  - GalaxyWEB
  - ArchPRED
  - RCD+
  - ROSIE

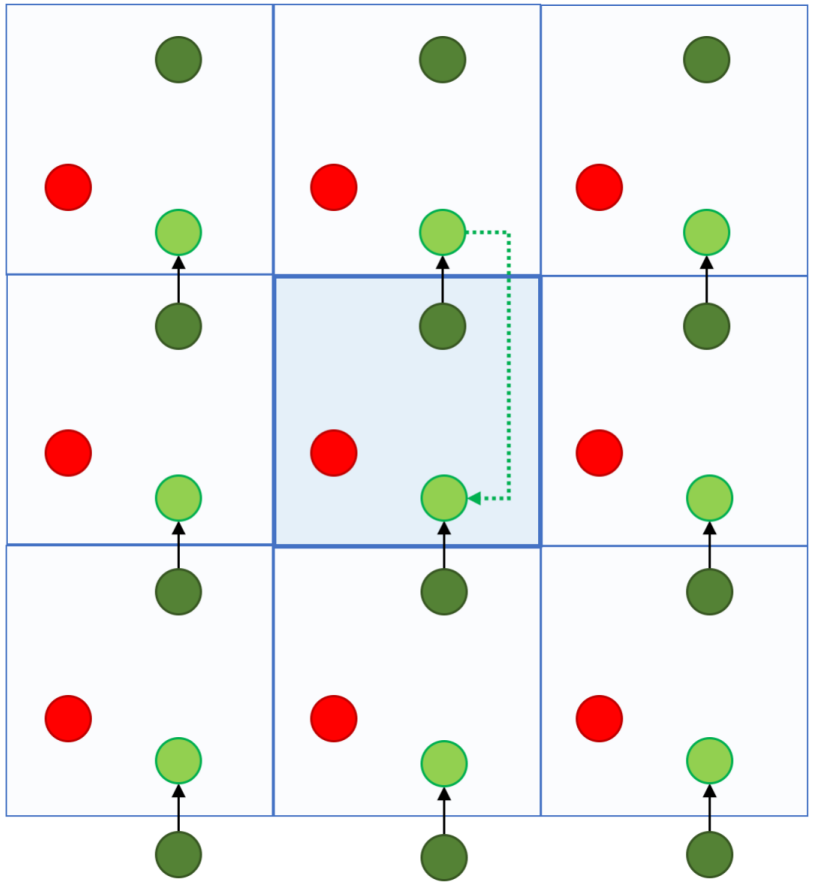
Running a simulation: setting up the protein and environment (part 2)

- Make sure the protein is in its correct protonation state
  - Karlsberg+
  - H++
  - PDB2PQR
- Asn, Gln, His flips
  - MolProbity

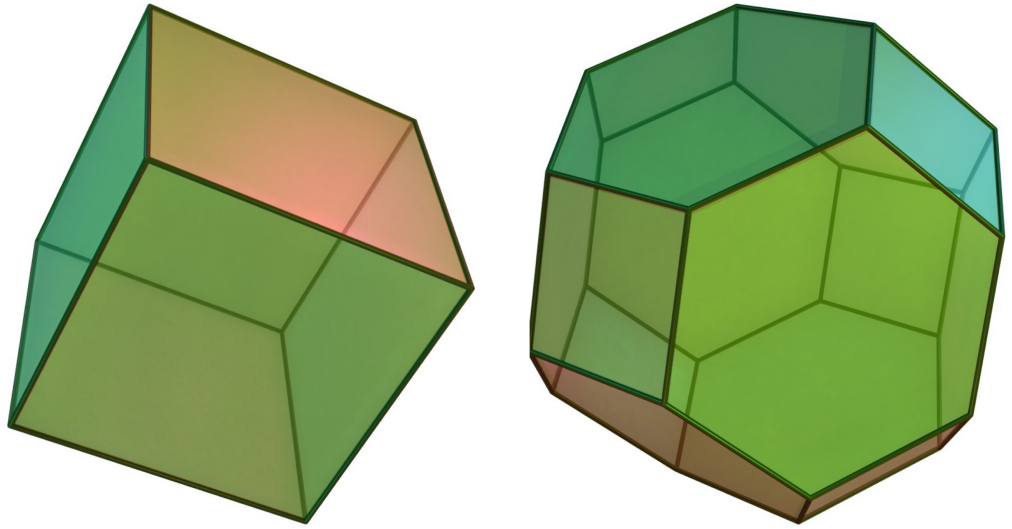


Running a simulation: setting up the protein and environment (part 3)

- Periodic boundary conditions

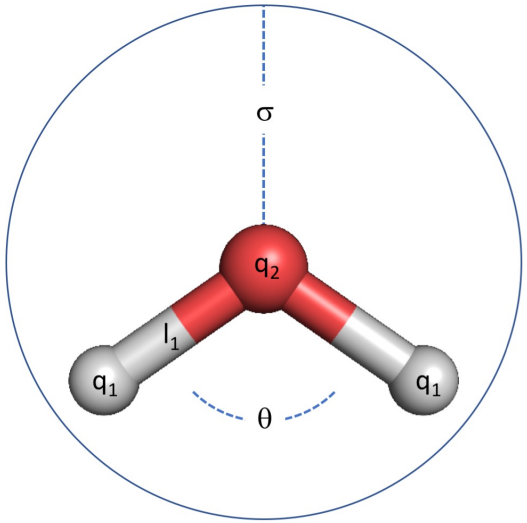


- Different box shapes



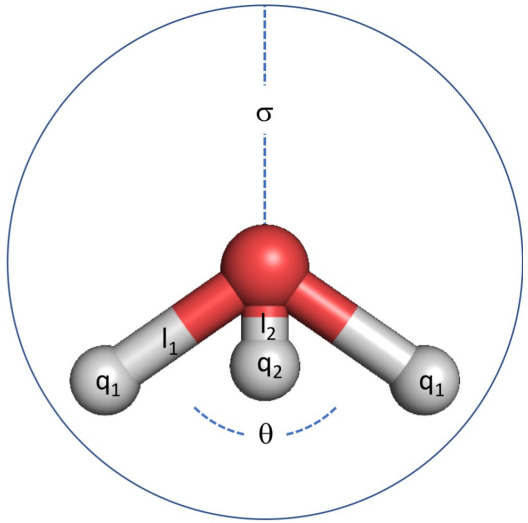
Running a simulation: setting up the protein and environment (part 4)

- Adding water and counterions



Model a

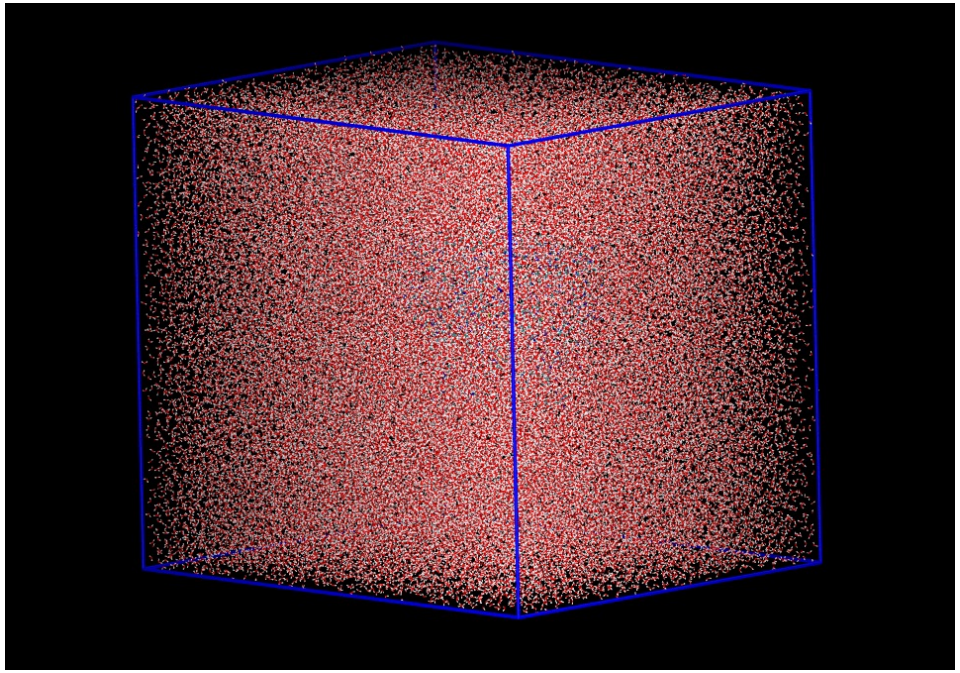
TIP3P
SPC



Model b

TIP4P

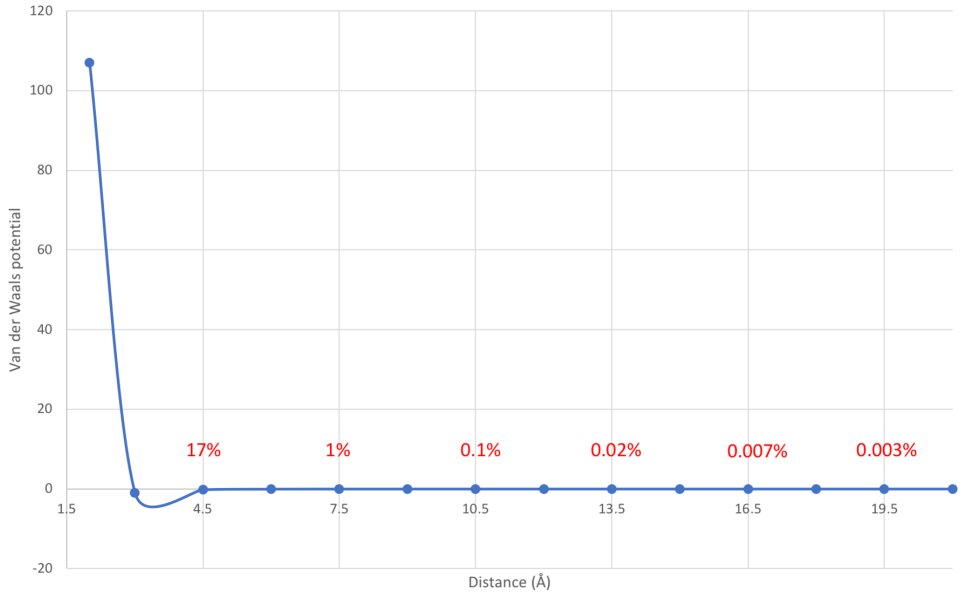
- Example box



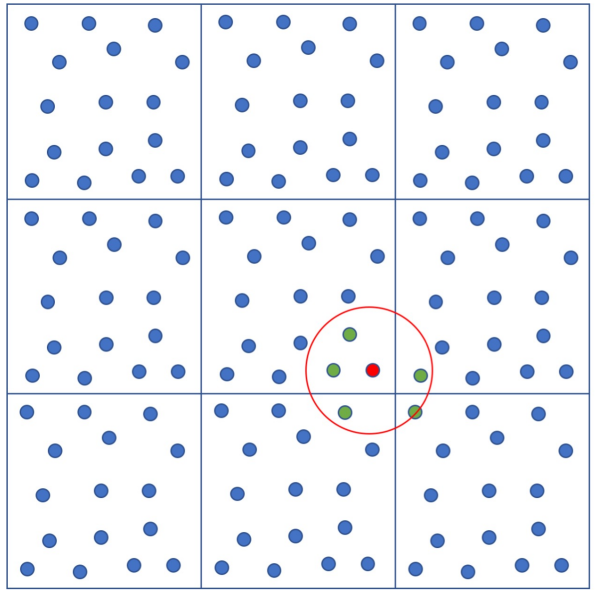
Running a simulation: maintaining T and P

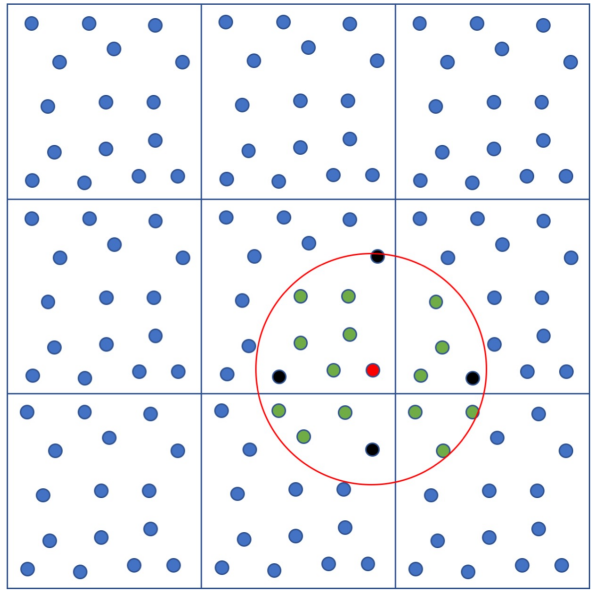
- Canonical ensemble
  - NVT constant
- Isothermal-isobaric ensemble
  - NPT constant

Short- and longe-range interactions



- Minimum image convention





Flow of a typical MD run

Minimization:

- Time:
  1,000-10,000 steps
- Temperature:
  0 K
- Restraints:
  10 kcal/mol/A^2

Heating:

- Time:
  5 ps
- Temperature:
  0 → 300 K
- Restraints:
  10 kcal/mol/A^2

Relaxation:

- Time:
  450 ps
- Temperature:
  300 K
- Restraints:
  10 → 0 kcal/mol/A^2

Equilibration:

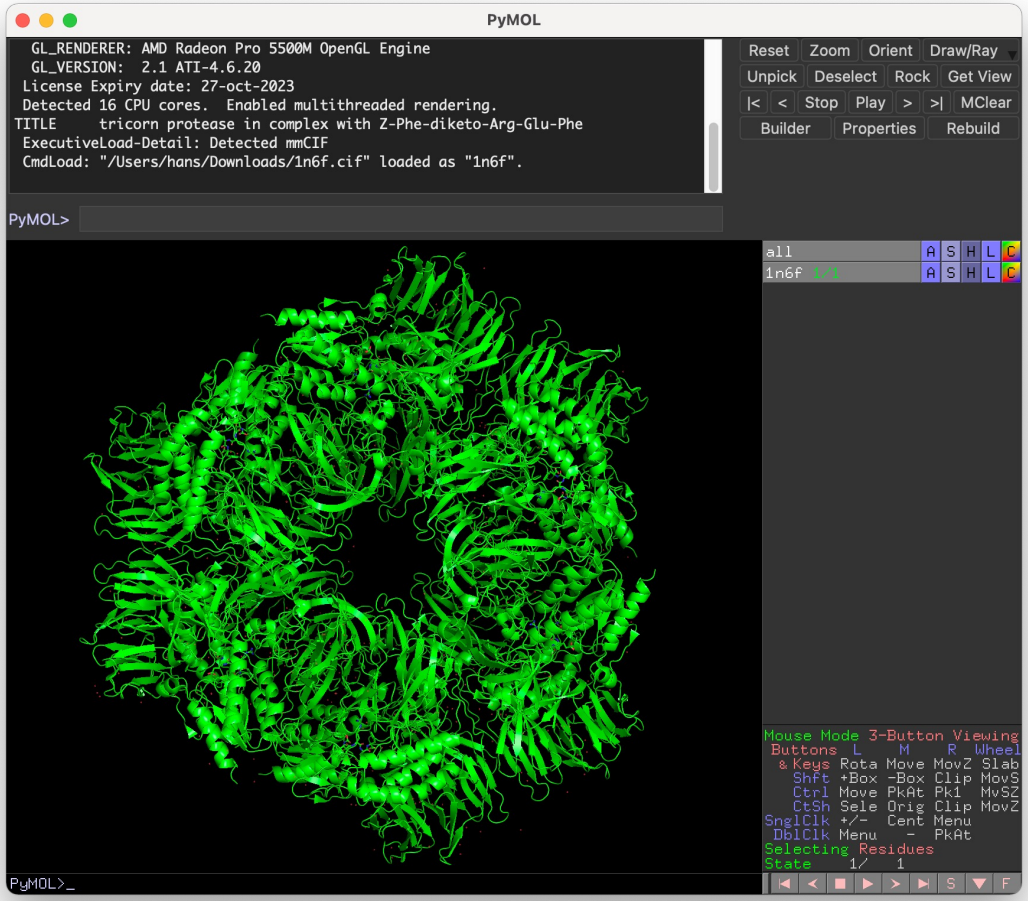
- Time:
  500 ps
- Temperature:
  300 K
- Restraints:
  -

Production:

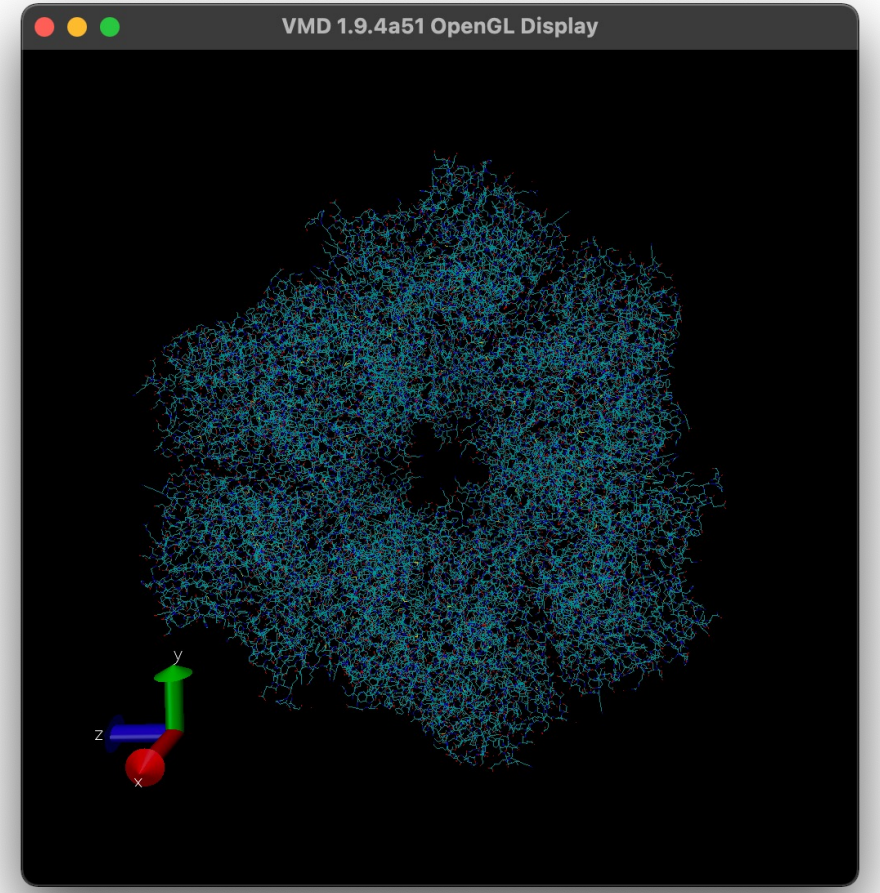
- Time:
  Multiple ns or µs
- Temperature:
  300 K
- Restraints:
  -

MD analysis: visualisation

PyMol



VMD

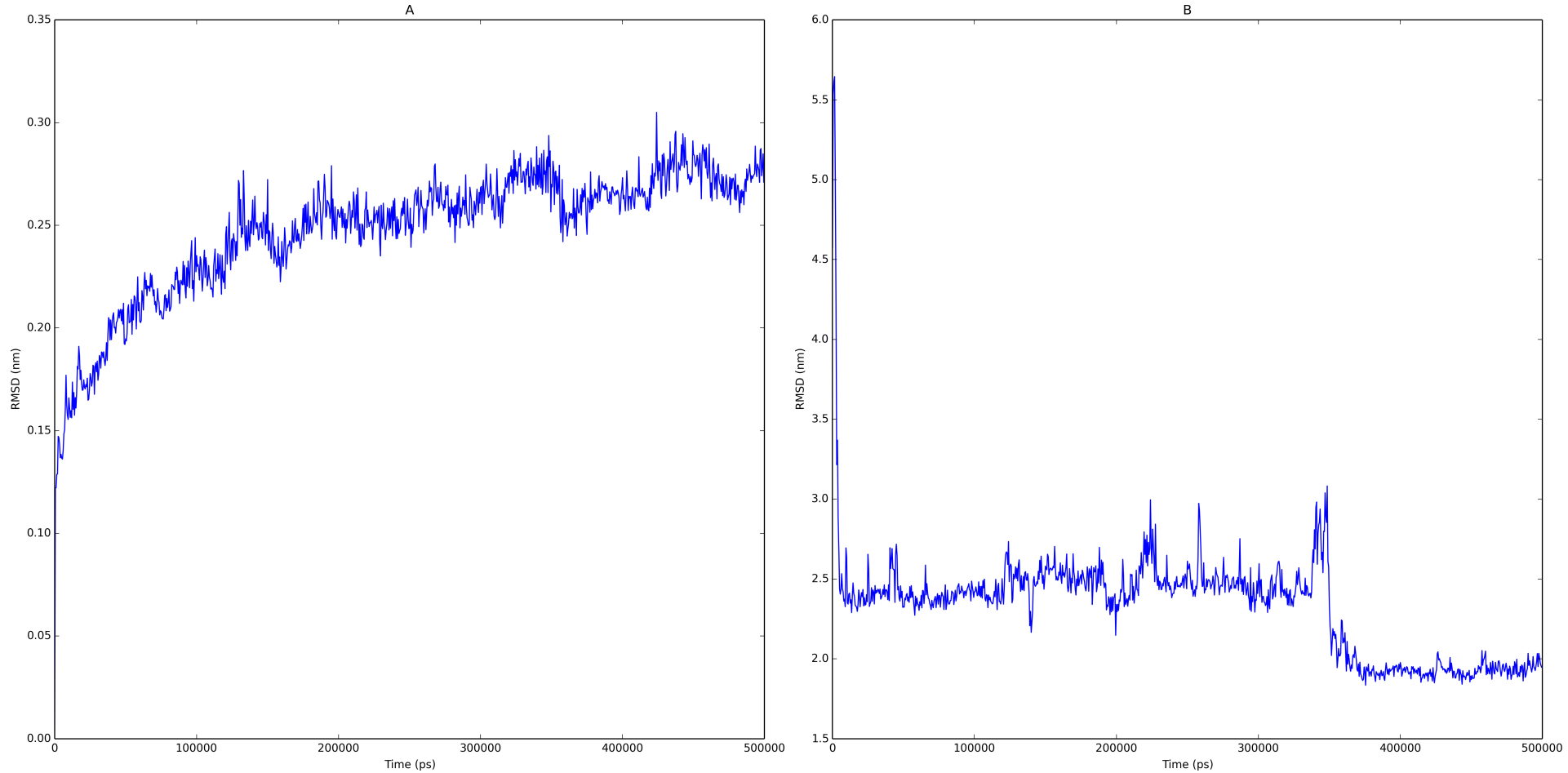


Applications of MD

- Protein dynamics and protein folding
  - RMSD, RMSF, ROG, SASA, ...
  - Case study: design of modified oligonucleotides as antisense therapeutics
- Study of protein-ligand complexes, binding events, affinities
  - Case study: design of novel inhibitors of PRCP with **umbrella sampling**
  - Case study: investigating the location of binding of NPD to LCMT with GaMD **absolute binding free energy calculations** (ABFE)

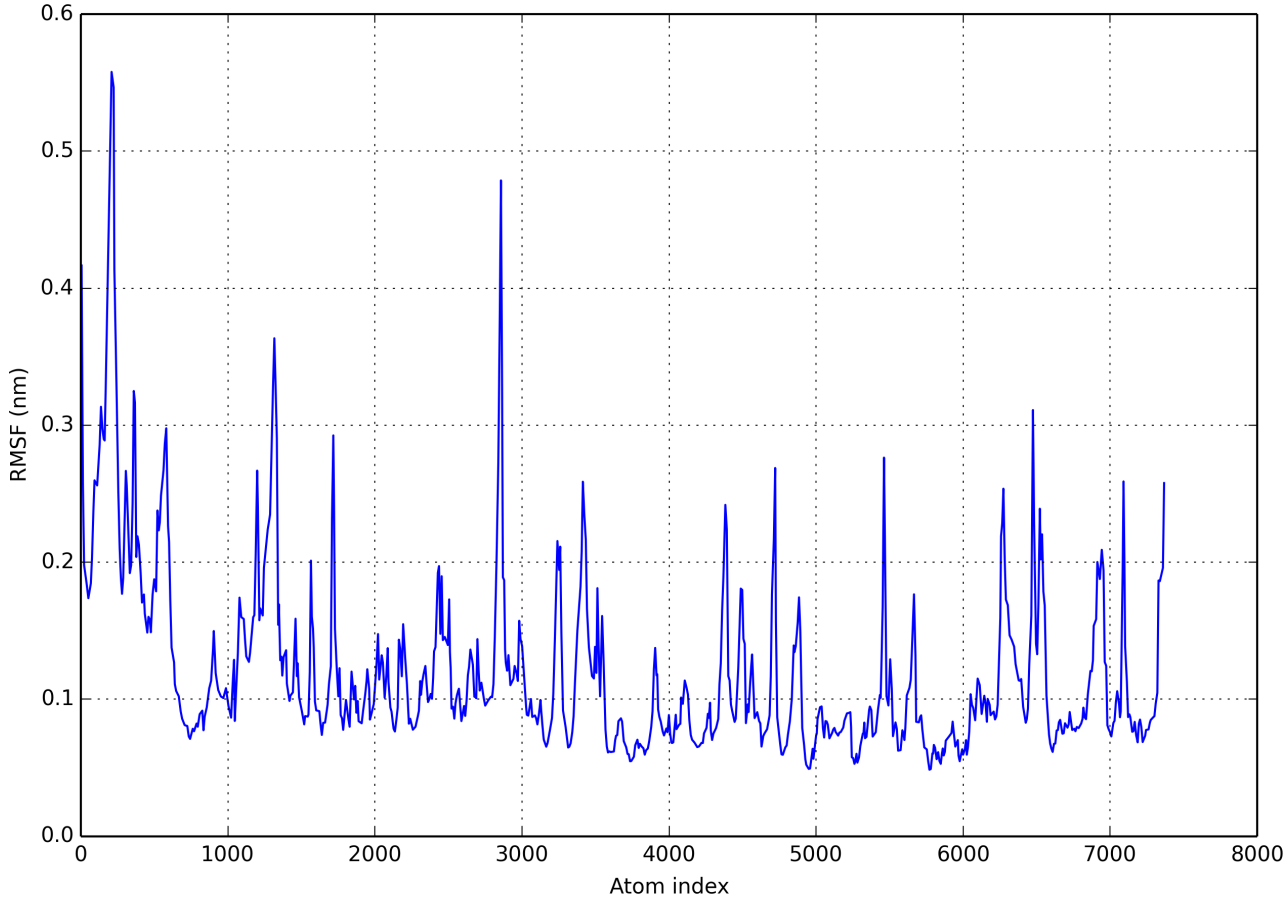
MD analysis: root-mean-square deviation (RMSD)

$$RMSD(t) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\vec{r}_i(t) - \vec{r_{i,ref}})^2}$$

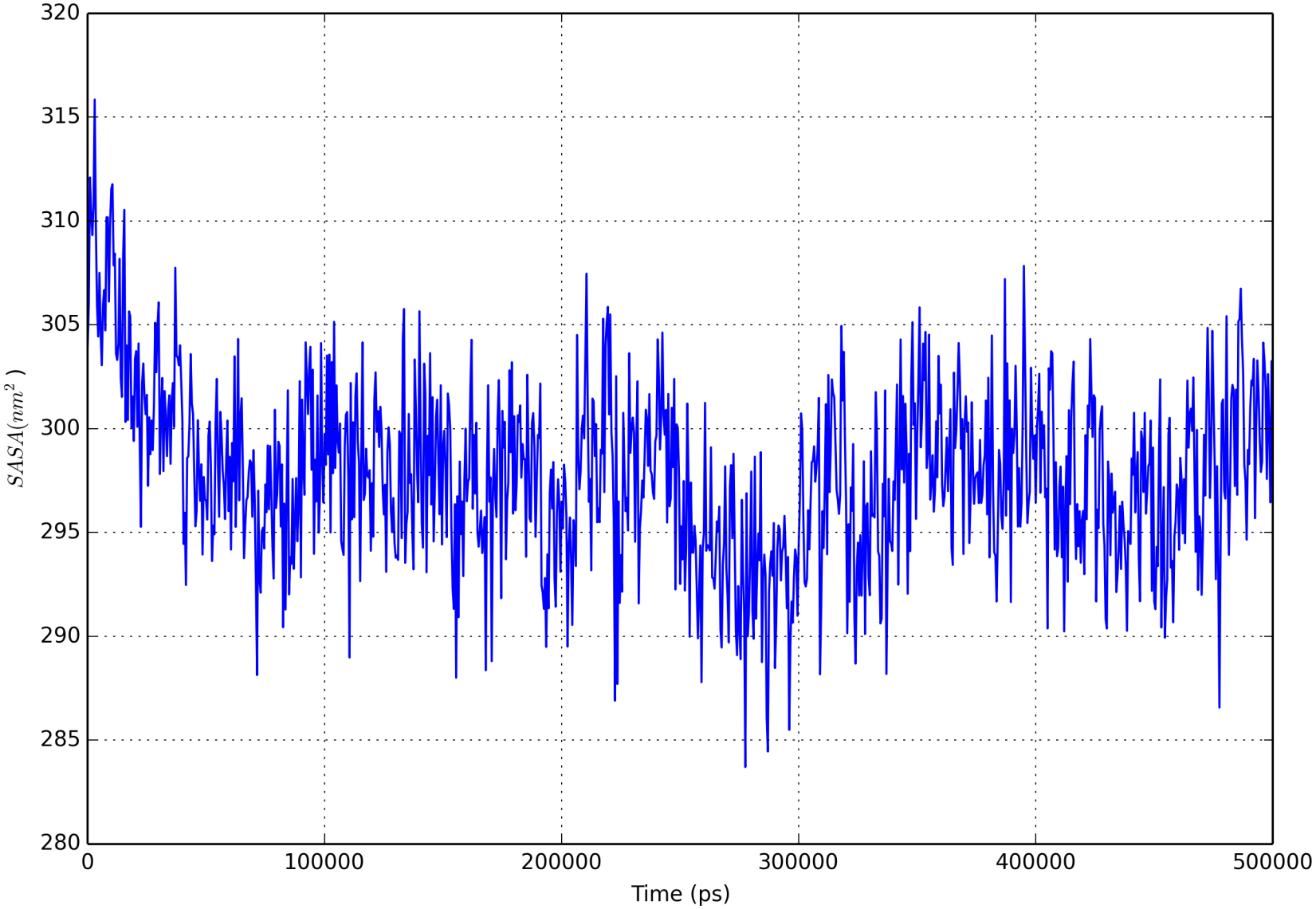


MD analysis: root-mean-square fluctuation (RMSF)

$$RMSF(i) = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(\vec{r}_i(t) - \vec{r_{i,ref}})^2}$$

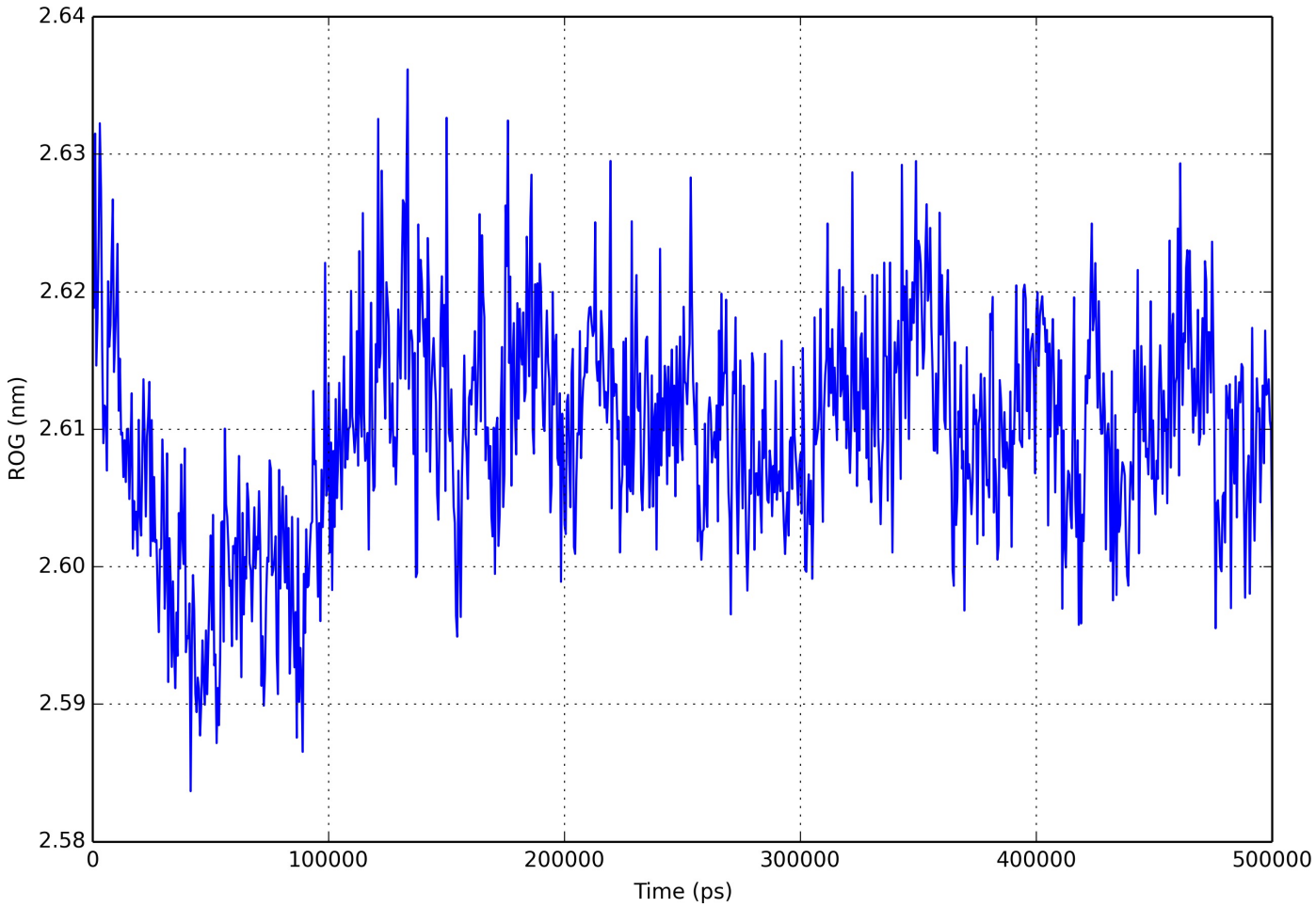


MD analysis: solvent-accessible surface area (SASA)



MD analysis: radius of gyration (ROG)

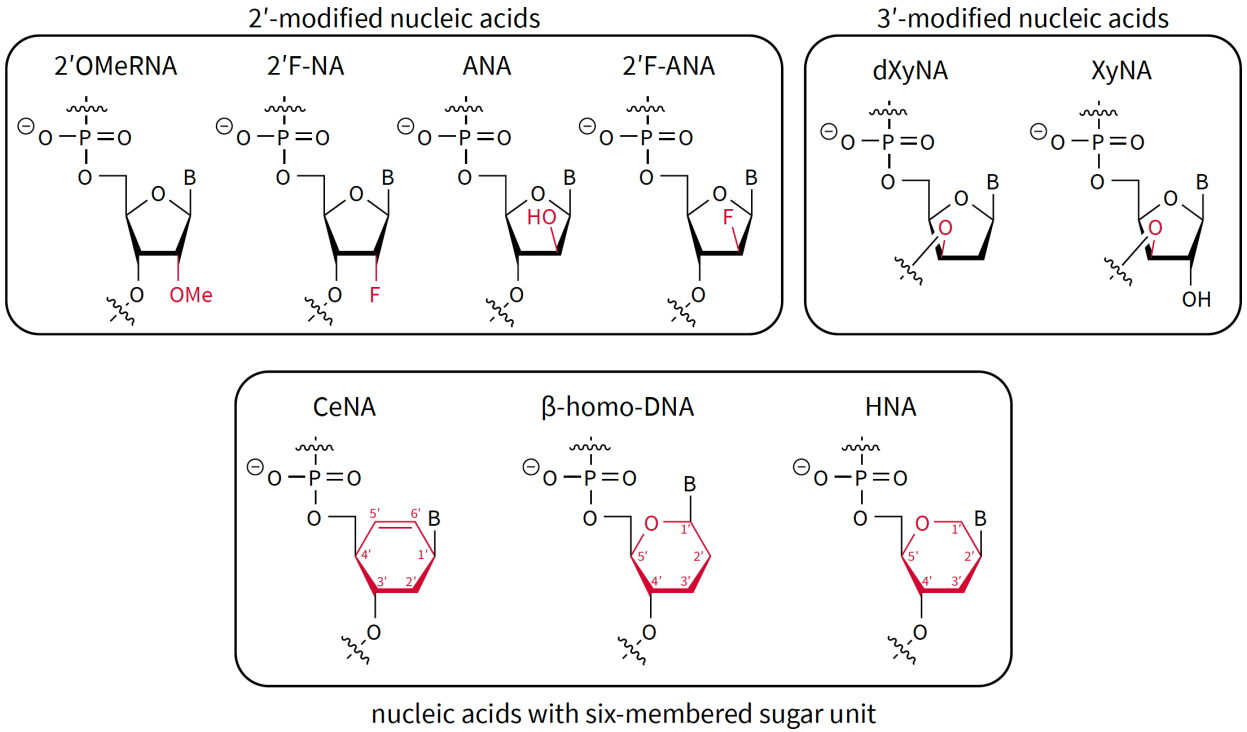
$$ROG(t) = \frac{1}{N}\sum_{i=1}^{N}(\vec{r}_i(t) - \vec{r_{mean}}(t))^2$$

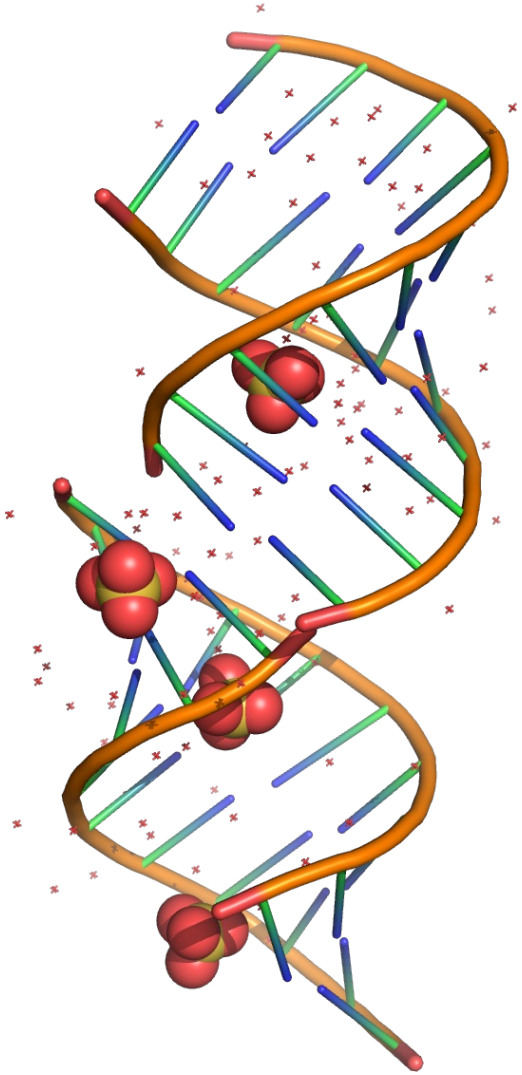


Applications of MD

- Protein dynamics and protein folding
  - RMSD, RMSF, ROG, SASA, ...
  - Case study: design of modified oligonucleotides as antisense therapeutics
- Study of protein-ligand complexes, binding events, affinities
  - Case study: design of novel inhibitors of PRCP with **umbrella sampling**
  - Case study: investigating the location of binding of NPD to LCMT with GaMD **absolute binding free energy calculations** (ABFE)

Antisense oligonucleotides





Applications of MD

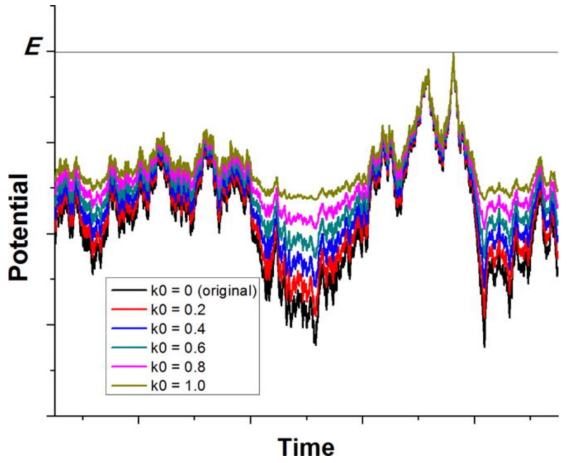
- Protein dynamics and protein folding
  - RMSD, RMSF, ROG, SASA, ...
  - Case study: design of modified oligonucleotides as antisense therapeutics
- Study of protein-ligand complexes, binding events, affinities
  - Case study: design of novel inhibitors of PRCP with **umbrella sampling**
  - Case study: investigating the location of binding of NPD to LCMT with GaMD **absolute binding free energy calculations** (ABFE)

Applications of MD

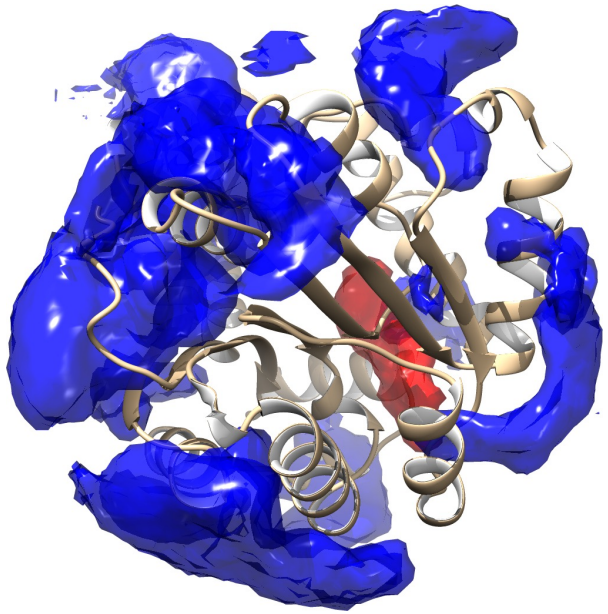
- Protein dynamics and protein folding
  - RMSD, RMSF, ROG, SASA, ...
  - Case study: design of modified oligonucleotides as antisense therapeutics
- Study of protein-ligand complexes, binding events, affinities
  - Case study: design of novel inhibitors of PRCP with **umbrella sampling**
  - Case study: investigating the location of binding of NPD to LCMT with **GaMD absolute binding free energy calculations** (ABFE)

Where does NPD bind?

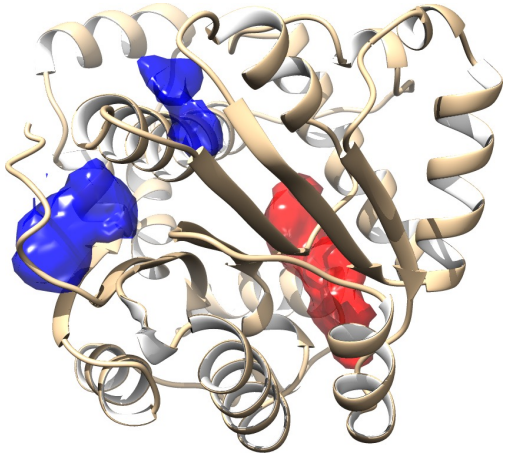
- LiGaMD simulations:
  - 25 independent runs, 4 μs accumulated
  - Each run started with NPD positioned randomly in the solvent, SAH positioned in its pocket
- Results:



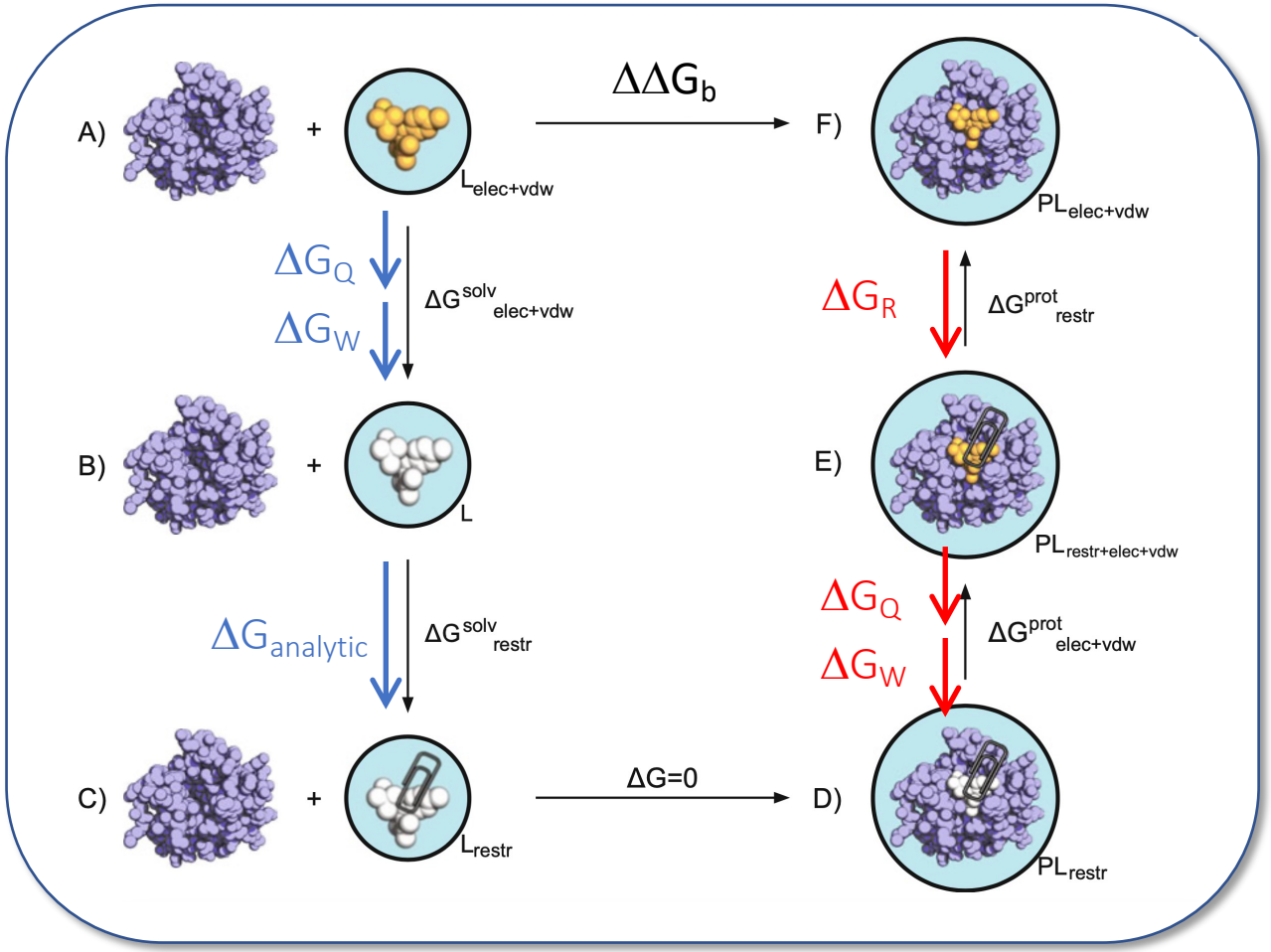
Miao & McCammon (2017) 'Gaussian accelerated molecular dynamics: Theory, implementation, and applications', *Annu. Rep. Comput. Chem.* **13**, 231-278.



SAH
NPD



Absolute binding free energy calculations



$\Delta G_{sol} = \Delta G_Q + \Delta G_W + \Delta G_{analytic}$

$\Delta G_{protein} = \Delta G_R + \Delta G_{Q+} \Delta G_R$

$\Delta\Delta G_b = \Delta G_{sol} - \Delta G_{protein}$

Results

System	$\Delta\Delta G_b$ (kcal/mol)	ΔG_{sol} (kcal/mol)			$\Delta G_{protein}$ (kcal/mol)		
		ΔG_Q	ΔG_W	$\Delta G_{analytic}$	ΔG_R	ΔG_Q	ΔG_W
NPD in NPD1 pocket (no SAH)	-4.5 ± 0.5	+95.74 ± 0.02	-0.03 ± 0.05	+8.88	+1.75 ± 0.11	+93.61 ± 0.12	+13.76 ± 0.51
NPD in NPD1 pocket (with SAH in SAH pocket)	-5.3 ± 1.4	+95.74 ± 0.02	-0.03 ± 0.05	+10.65	+2.30 ± 0.99	+95.15 ± 0.57	+14.17 ± 0.73
NPD in NPD2 pocket (no SAH)	-5.0 ± 1.0	+95.74 ± 0.02	-0.03 ± 0.05	+8.66	+2.10 ± 0.04	+93.14 ± 0.35	+14.12 ± 0.90
NPD in NPD2 pocket (with SAH in SAH pocket)	-1.8 ± 0.7	+95.74 ± 0.02	-0.03 ± 0.05	+9.46	+2.08 ± 0.07	+93.96 ± 0.22	+10.89 ± 0.60
NPD in SAH pocket (no SAH)	-2.8 ± 1.8	+95.74 ± 0.02	-0.03 ± 0.05	+9.90	+5.44 ± 1.76	+91.67 ± 0.23	+11.32 ± 0.35
SAH in SAH pocket (no NPD)	-102.2 ± 1.7	+216.52 ± 0.31	+3.31 ± 0.07	+10.42	+1.74 ± 0.04	+309.01 ± 0.84	+21.72 ± 1.45
SAH in SAH pocket (with NPD in NPD1 pocket)	-19.9 ± 1.7	+216.52 ± 0.31	+3.31 ± 0.07	+11.03	+1.49 ± 0.10	+228.73 ± 0.47	+20.57 ± 1.60
SAH in SAH pocket (with NPD in NPD2 pocket)	-22.2 ± 1.3	+216.52 ± 0.31	+3.31 ± 0.07	+10.79	+2.01 ± 0.11	+230.41 ± 1.06	+20.39 ± 0.74

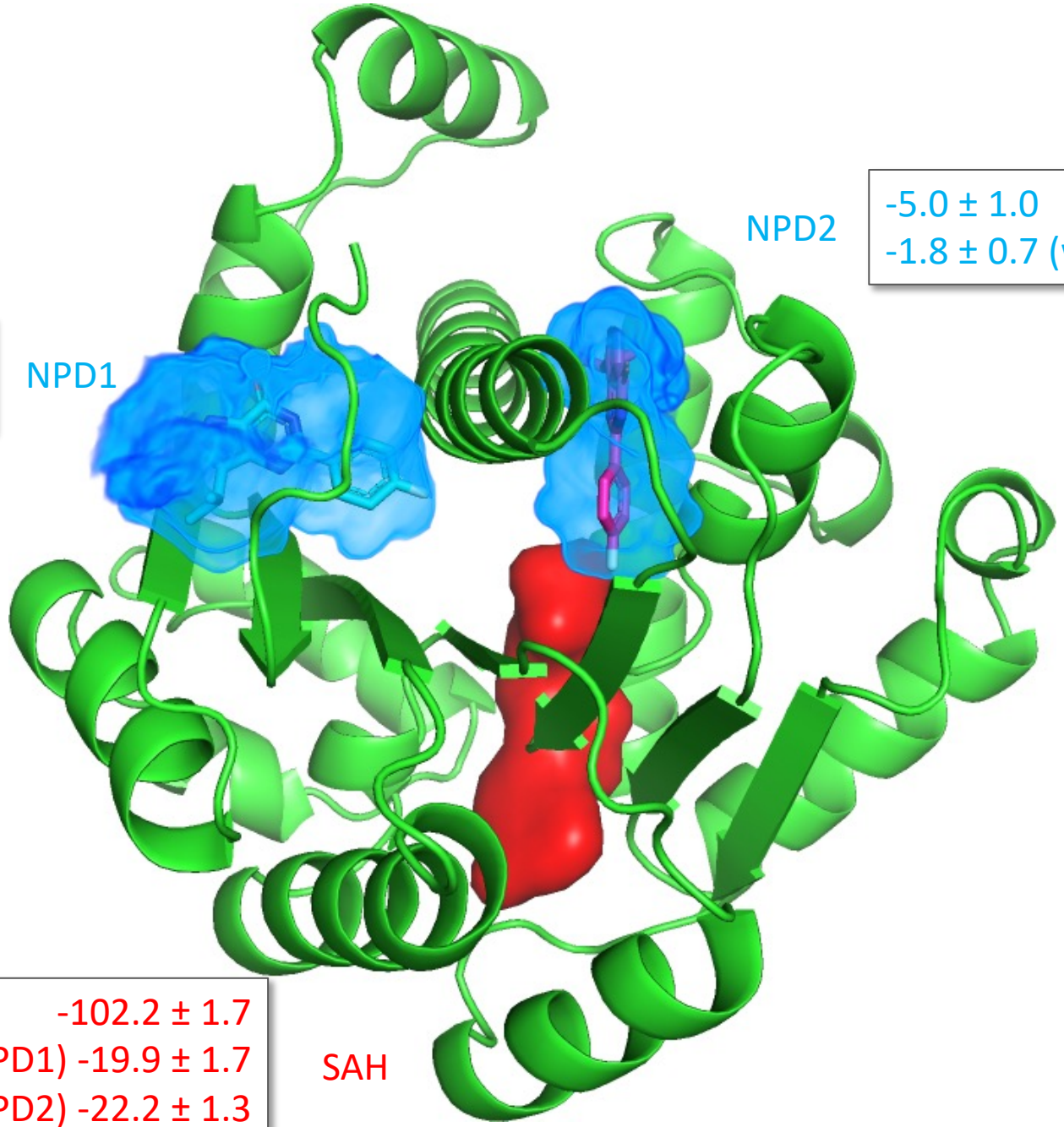
(all values in kcal/mol)

-4.5 ± 0.5
(with SAH) -5.3 ± 0.5

NPD1

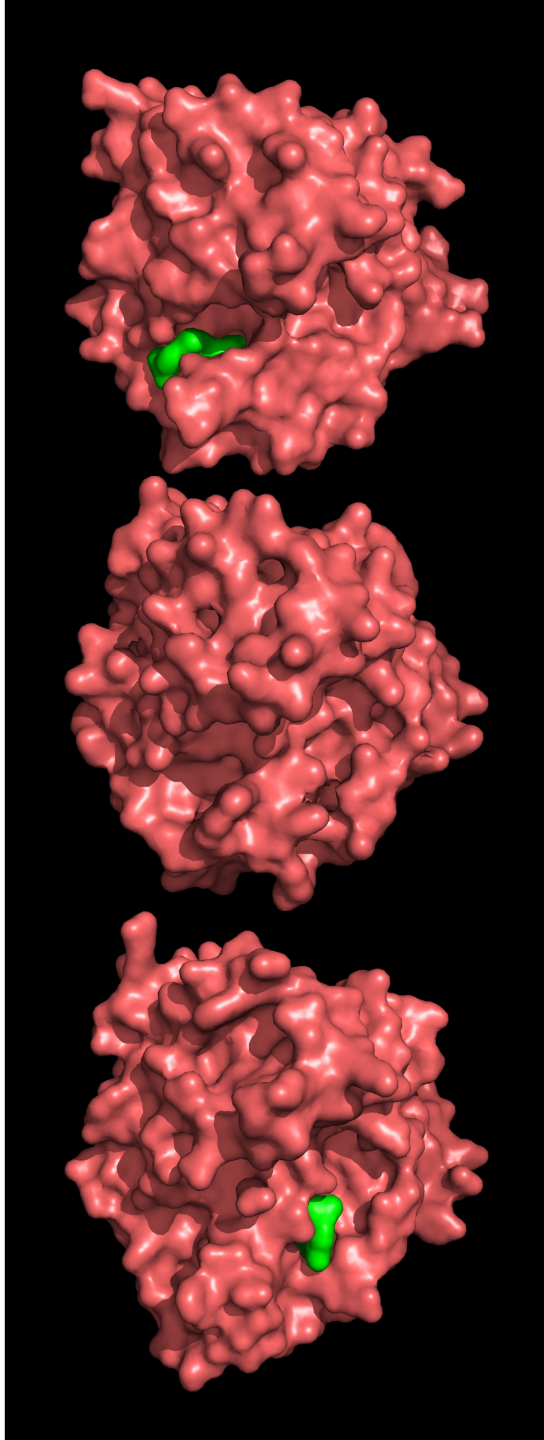
NPD2

-5.0 ± 1.0
-1.8 ± 0.7 (with SAH)



-102.2 ± 1.7
(with NPD1) -19.9 ± 1.7
(with NPD2) -22.2 ± 1.3

SAH



Overview video

- https://www.youtube.com/watch?v=OjV90WVJj2M
  (or `molecular-dynamics.mp4` on BB)