Using GPUs to accelerate MD simulations of RNA/protein complexes

Andrea Pérez-Villa

International School of Advanced Studies, SISSA

Trieste, Italy

Workshop on Accelerated High-Performance Computing

Trieste

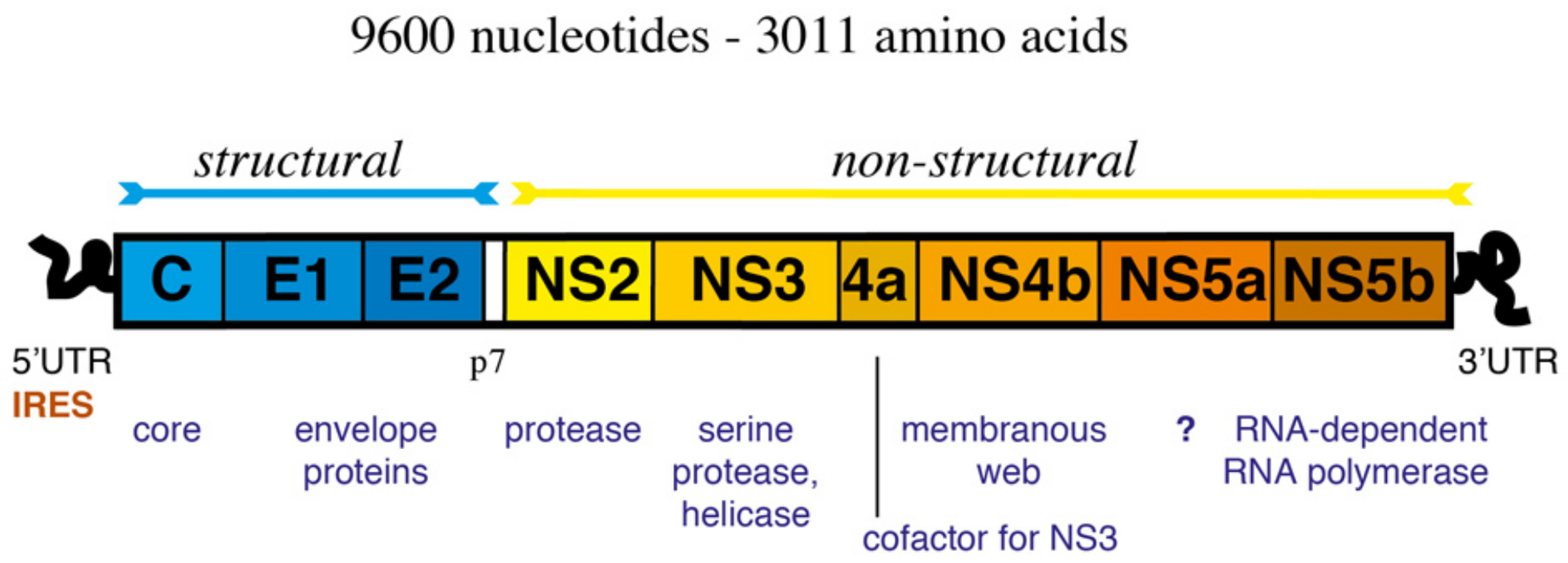
May 29, 2015

Introduction

Hepatitis C virus (HCV)

Hepatitis C Virus:

- Flaviviridae virus, cause of hepatitis C in humans.
- Its (+)-ssRNA encodes for structural and nonstructural (NS) proteins.
- One of the NS proteins is NS3 Helicase(h)/serine protease.



Swiss Med Wkly. (2012)

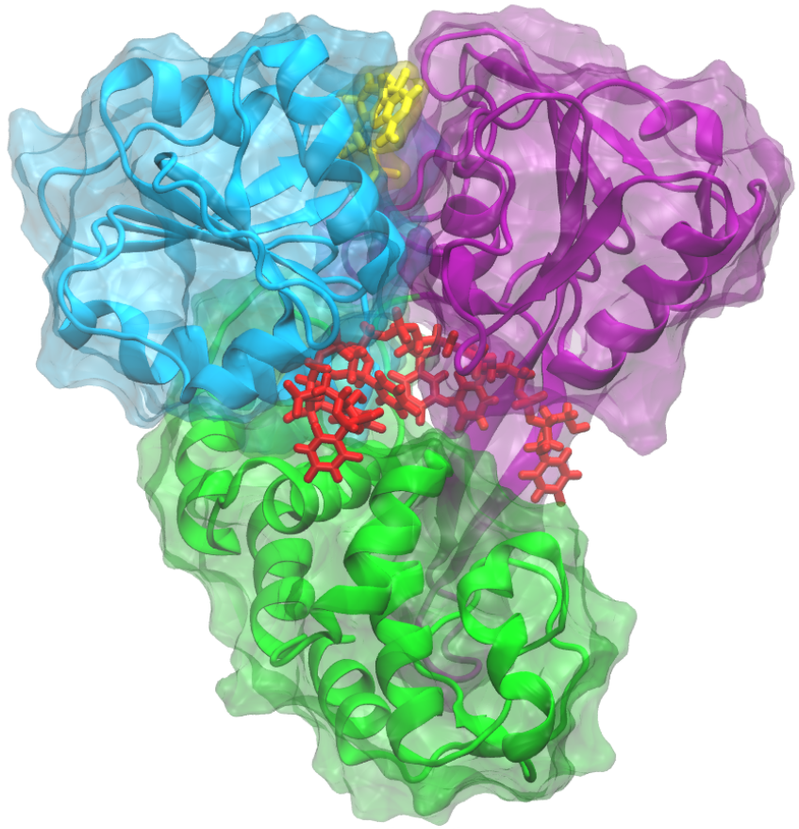
S.L. Tan editor. Hepatitis C Viruses: Genomes and Molecular Biology. Norfolk (UK): Horizon Bioscience (2006)

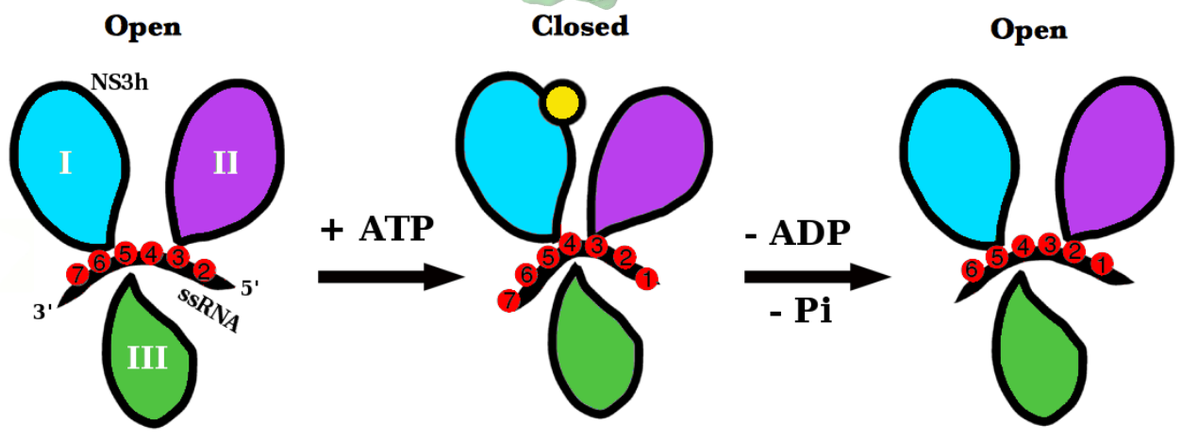
NS3 HCV

NS3h HCV: Motor protein member of Helicase superfamily II. Important for the replication of Hepatitis C Virus.

- x-ray crystal structures[a]
- single molecule experiments[b]
- coarse-grained simulations[c]

"inchworm-like" mechanism. One base pair (of DNA or RNA) unwinding/translocation per one ATP molecule.





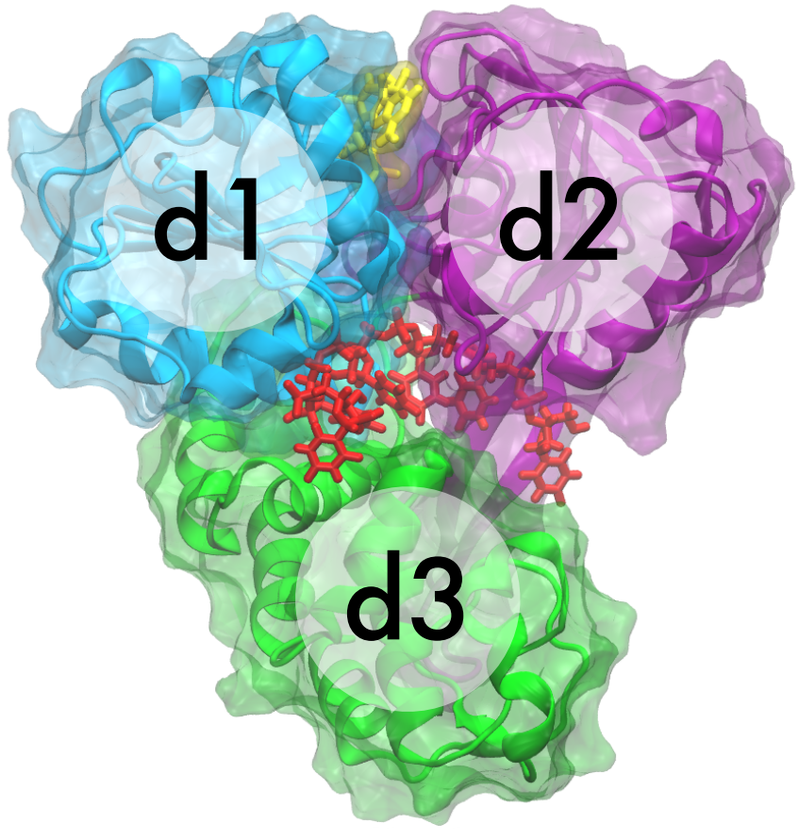
[a]Appleby et al. JMB (2011)
[b]W. Cheng et al. Science. (2011)
[c]H. Flechsing, A. S. Mikhailov. PNAS. (2010)
http://www.alecjacobson.com/weblog/media/worm-walk.gif

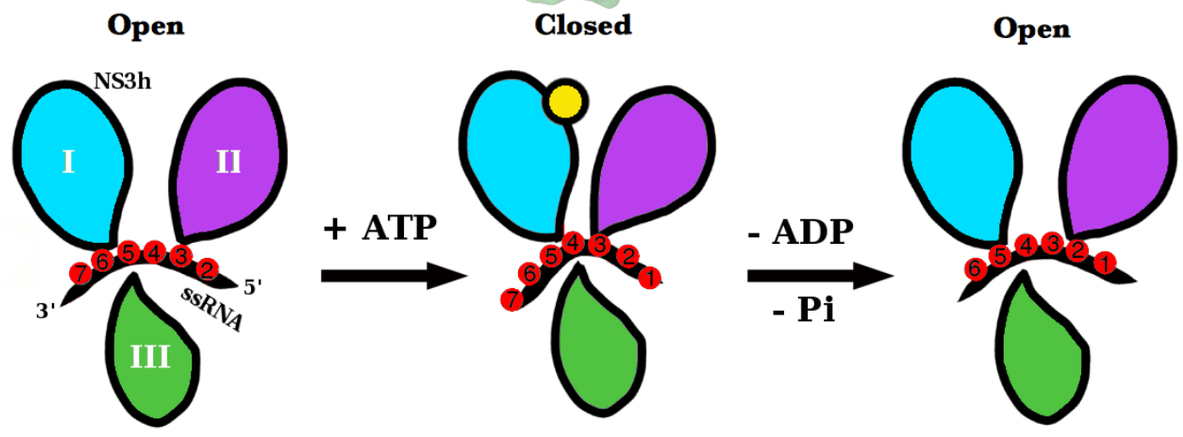
NS3 HCV

NS3h HCV: Motor protein member of Helicase superfamily II. Important for the replication of Hepatitis C Virus.

- x-ray crystal structures[a]
- single molecule experiments[b]
- coarse-grained simulations[c]

"inchworm-like" mechanism. One base pair (of DNA or RNA) unwinding/translocation per one ATP molecule.





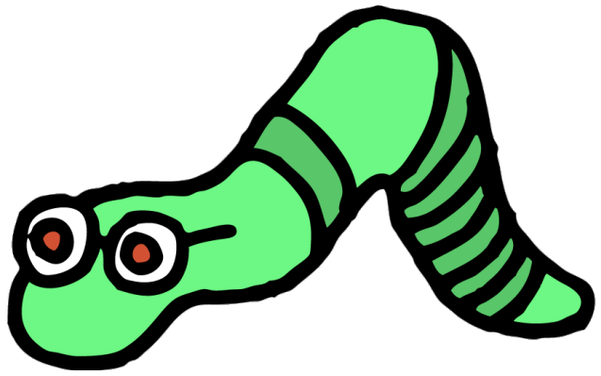
[a]Appleby et al. JMB (2011)
[b]W. Cheng et al. Science. (2011)
[c]H. Flechsing, A. S. Mikhailov. PNAS. (2010)
http://www.alecjacobson.com/weblog/media/worm-walk.gif

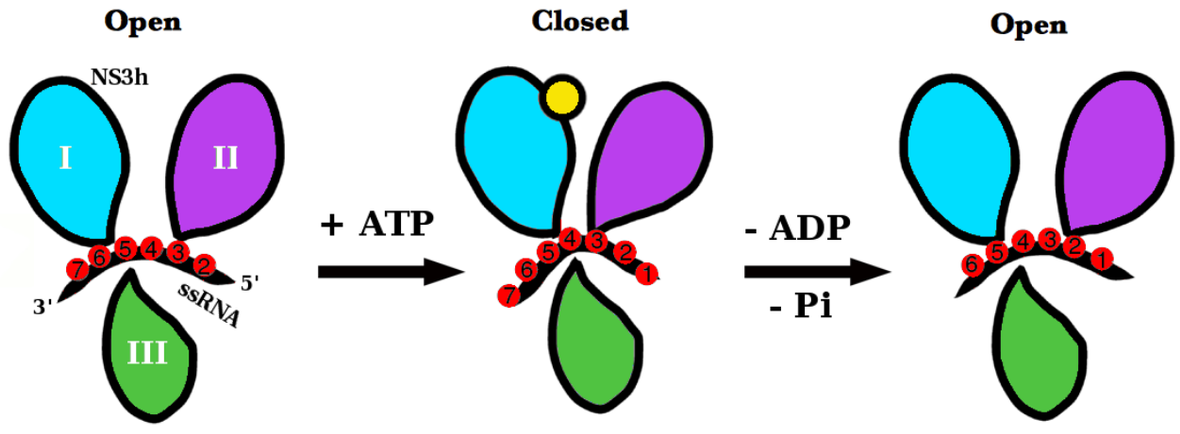
NS3 HCV

NS3h HCV: Motor protein member of Helicase superfamily II. Important for the replication of Hepatitis C Virus.

- x-ray crystal structures[a]
- single molecule experiments[b]
- coarse-grained simulations[c]

"inchworm-like" mechanism. One base pair (of DNA or RNA) unwinding/translocation per one ATP molecule.



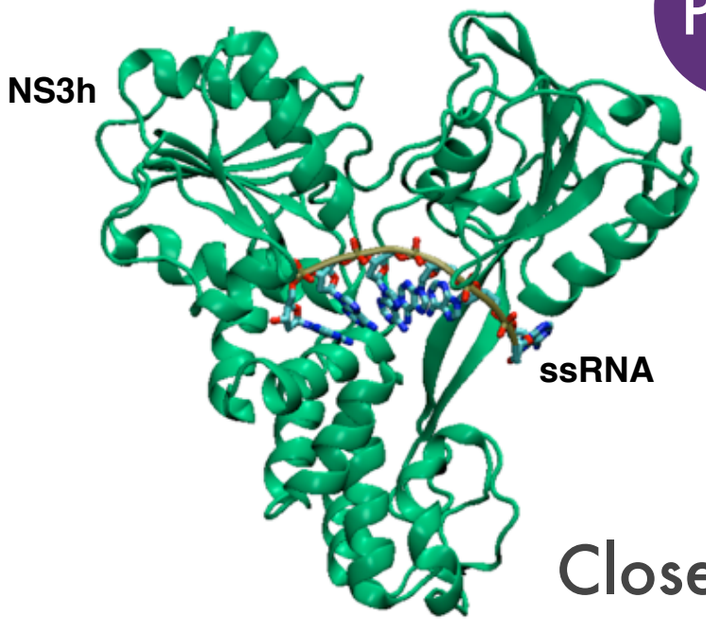


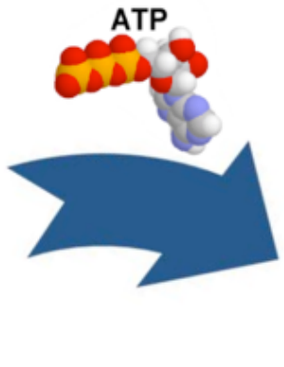
[a]Appleby et al. JMB (2011)
[b]W. Cheng et al. Science. (2011)
[c]H. Flechsing, A. S. Mikhailov. PNAS. (2010)
http://www.alecjacobson.com/weblog/media/worm-walk.gif

Overview of the mechanism

Open snapshot

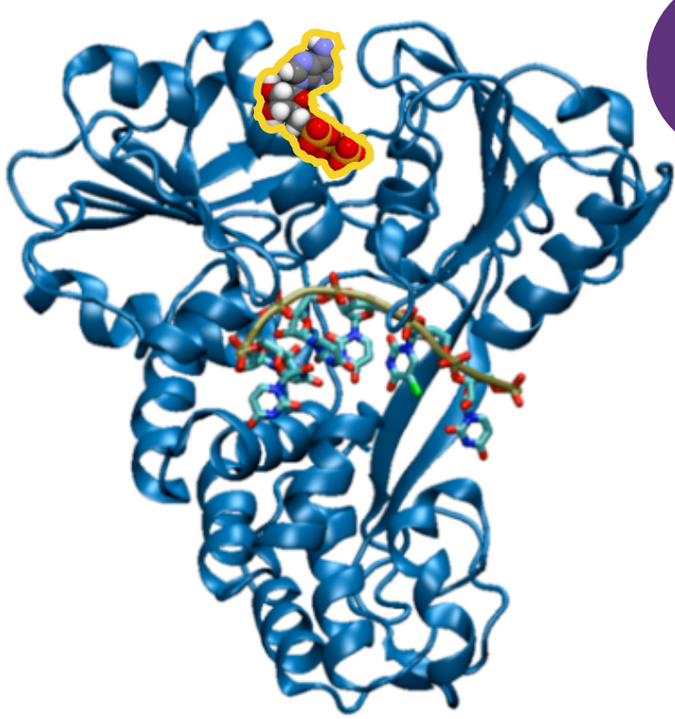
PDB



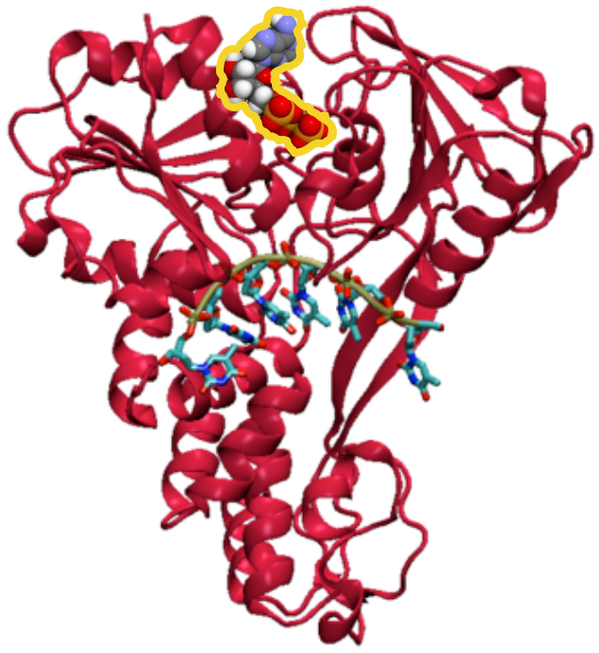


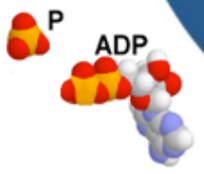
Closed snapshot
ATP ground state

PDB



Closed snapshot
ATP transition state

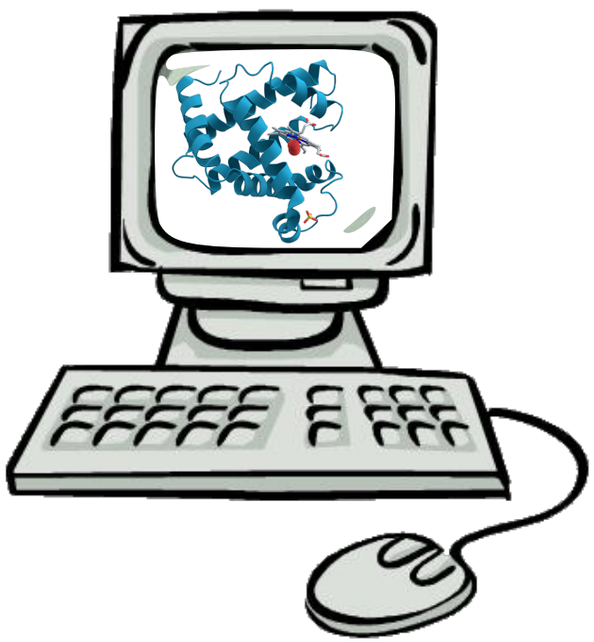




- X-ray snapshots.
- Single molecule experiments (kinetics).
- Hydrolysis reaction: ATP, ADP.

- X-ray snapshots.
- Single molecule experiments (kinetics).
- Hydrolysis reaction: ATP, ADP.
- Atomistic details of the mechanism?

- X-ray snapshots.
- Single molecule experiments (kinetics).
- Hydrolysis reaction: ATP, ADP.
- Atomistic details of the mechanism?



Molecular dynamics simulations

Apo

ADP • Mg^{2+}

ATP • Mg^{2+}

Closed

Open

PDB

PDB

Outline

- Computational details.
- Equilibrium properties (Long Plain MD).
- Performance issue on bias sampling.
- Summary.

Computational details

Plain MD set-up

- FF: amber99sb-*ildn-parmbsc0-χ_{OL} corrections + ATP/ADP + Mg[a] + explicit water(TIP3P).
- Velocity Rescaling Thermostat, T_{ref}: 300 K.
- Berendsen Barostat, P_{ref}: 1.0 bar.
- Protease domain not included.
- Simulation time: 1 μs x 6 systems

[a]Best and Hummer. J. Phys. Chem. B. (2009); Lindorff-Larsen et al. Proteins (2010); Banas et al. JCTC (2010); Meagher, Redman and Carlson. J. Comp. Chem. (2003); Allnér, Nilsson and Villa. JCTC (2012)

GROMACS-4.6.x Software[a]

- Hybrid acceleration,
  - GPU accelerated: Non bonded force calculations.
  - CPU: Bonded and PME electrostatics.
- Cut-off scheme: Verlet.

Plumed2.0 plugin[b]

- CPU code.
- Analysis and enhanced sampling (e.g. metadynamics).

[a]www.gromacs.org
[b]www.plumed-code.org

Load balancing and PME

- Higher parallelization (at > 16 processes): 1/4 of the nodes do PME, 3/4 of the nodes do non-bonded (with domain decomposition).
- Lower parallelization (at < 16 processes).
- With GPU+CPU: 1 GPU per domain. Non-bonded on GPU, bonded + PME on CPU.
- Load balancing: cutoff of non-bonded is adapted on relative CPU/GPU load.

Load balancing and PME

aperez@login2:/scratch/aperez/largerRNA

File Edit View Search Terminal Help

```
-[no]hrex    bool   no      Enable hamiltonian replica exchange
-[no]ionize  bool   no      Do a simulation including the effect of an X-Ray
                            bombardment on your system


Number of CPUs detected (20) does not match the number reported by OpenMP (1).
Consider setting the launch configuration manually!
Reading file topol0.tpr, VERSION 4.6.7 (single precision)
Changing nstlist from 10 to 40, rlist from 1 to 1.09

The number of OpenMP threads was set by environment variable OMP_NUM_THREADS to 10
Using 4 MPI processes
Using 10 OpenMP threads per MPI process

2 GPUs detected on host gn05-09:
  #0: NVIDIA Tesla K20m, compute cap.: 3.5, ECC: yes, stat: compatible
  #1: NVIDIA Tesla K20m, compute cap.: 3.5, ECC: yes, stat: compatible

2 GPUs auto-selected for this run.
Mapping of GPUs to the 2 PP ranks in this node: #0, #1


Overriding thread affinity set out[illegible]n_mpi


WARNING: This run will generate roughly 139533 Mb of data

starting mdrun 'Protein in water'
100000000 steps, 200000.0 ps.
^Mstep   80: timed with pme grid 96 96 96, coulomb cutoff 1.000: 495.6 M-cycles
^Mstep  160: timed with pme grid 80 80 80, coulomb cutoff 1.[illegible]: 416.9 M-cycles
^Mstep  240: timed with pme grid 72 72 72, coulomb cutoff 1.[illegible]: 441.0 M-cycles
^Mstep  320: timed with pme grid 64 64 64, coulomb cutoff 1.476: 506.6 M-cycles
^Mstep  400: timed with pme grid 84 84 84, coulomb cutoff 1.125: 434.9 M-cycles
^Mstep  480: timed with pme grid 80 80 80, coulomb cutoff 1.181: 414.9 M-cycles
^Mstep  560: timed with pme grid 72 72 72, coulomb cutoff 1.312: 412.4 M-cycles
^Mstep  640: timed with pme grid 84 84 84, coulomb cutoff 1.125: 435.0 M-cycles
^Mstep  720: timed with pme grid 80 80 80, coulomb cutoff 1.181: 413.0 M-cycles
^Mstep  800: timed with pme grid 72 72 72, coulomb cutoff 1.312: 412.7 M-cycles
^M              optimal pme grid 72 72 72, coulomb cutoff 1.312
step 900, will finish Mon Jun  1 16:52:42 2015
imb F 24% step 1000, will finish Mon Jun  1 03:12:54 2015
step 1100, will finish Sun May 31 15:52:51 2015
imb F 25% step 1200, will finish Sun May 31 06:34:40 2015
step 1300, will finish Sat May 30 22:54:37 2015
                                                              178,1         0%
```

PME

Coulomb cut-off

OpenMP

- OpenMP multithreading: Exploit multicore machines!
- Multi-core parallelization,
  - Each domain runs on a separate node
  - Intra-domain particle decomposition with OpenMP
- OpenMP multithreading faster than MPI-based parallelization.

Computational details

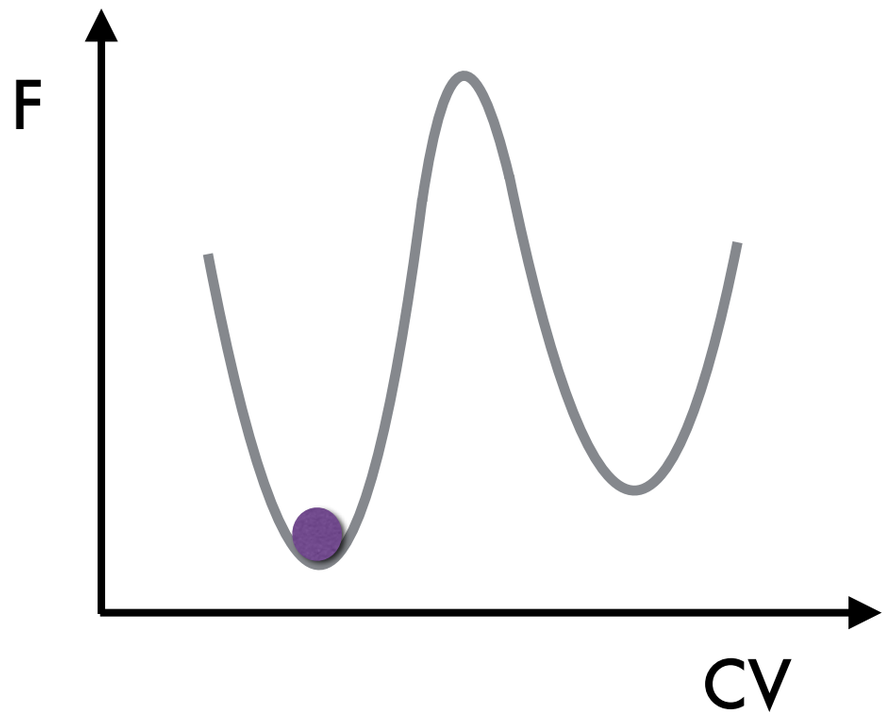
System	Peptide	ssRNA	Ligand	Mg	Na	Cl	Water	Total atoms
Apo	6528	179	---	---	70	62	93174	100012
ADP·Mg	6528	179	39	1	70	61	93174	100051
ATP·Mg	6528	179	43	1	70	60	93174	100054

Keeping same protonation for all peptides.

Total water molecules: 31058
Solute far 20 Å from border

Performance and bias sampling

Sampling rare events

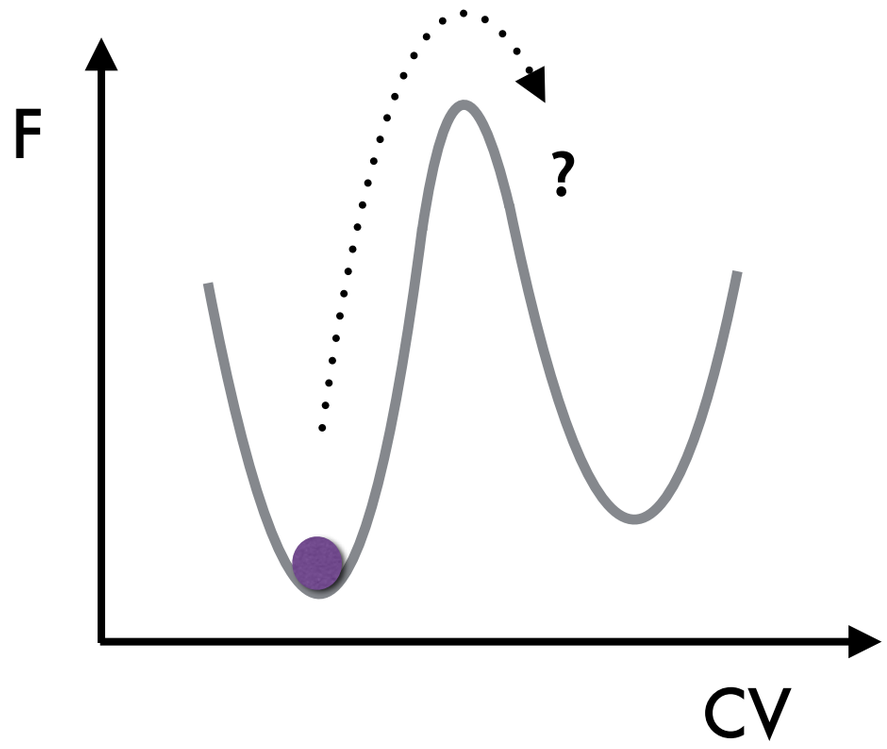


- Very long MD simulations (Brute force!)
- Simulations based on annealing.
- Simulations based on *a priori* physical knowledge (biased sampling): e.g. Metadynamics, Umbrella Sampling, etc…

★ Use of collective variables (CVs)

(PLUMED is a plugin implemented to perform bias sampling)

Sampling rare events



- Very long MD simulations (Brute force!)
- Simulations based on annealing.
- Simulations based on *a priori* physical knowledge (biased sampling): e.g. Metadynamics, Umbrella Sampling, etc…

★ Use of collective variables (CVs)

(PLUMED is a plugin implemented to perform bias sampling)

Performance and bias sampling

★ Biased sampling: Allows a significant speedup analyzing rare events and need good CVs.

Sometimes CVs are expensive:

- Steinhardt order parameters[a]
- Path/Property maps[b]
- Secondary structure CVs[c]
- SPRINT[d]
- Sketch maps[e]
- DH Energy[f]

[a]Steinhardt, Nelson, and Ronchetti, PRB (1983); Trudu, Donadio, and Parrinello, PRL (2006);
[b]Branduardi, Gervasio, and Parrinello, JCP (2007); Spiwok and Králová, JCP(2011);
[c]Pietrucci and Laio, JCTC (2009); [d]Pietrucci and Andreoni, PRL (2011);
[e]Tribello, Ceriotti, and Parrinello, PNAS (2012); [f]Do, Carloni, Varani, and Bussi JCTC (2013)

Performance and bias sampling

★ Biased sampling: Allows a significant speedup analyzing rare events and need good CVs.

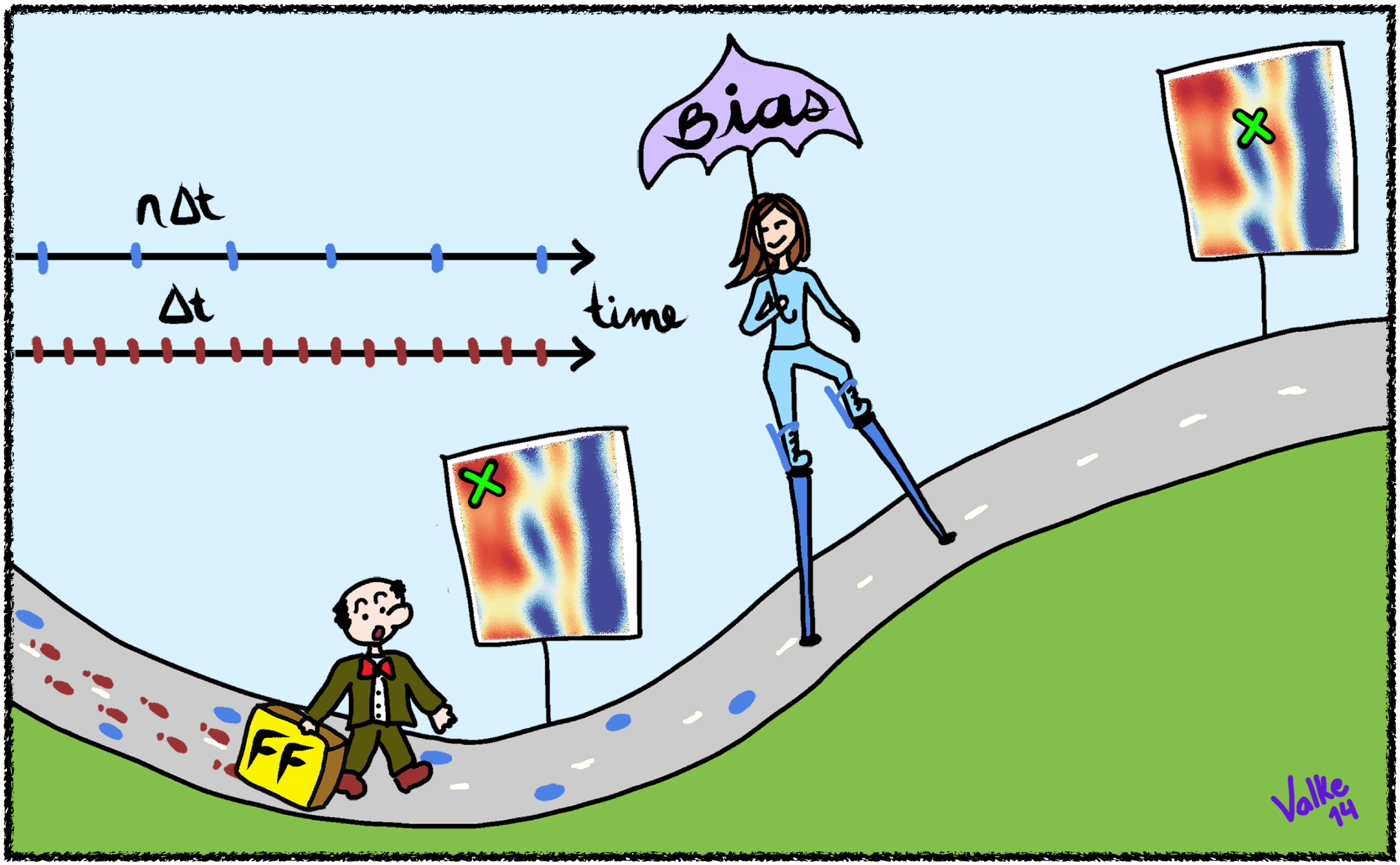
Sometimes CVs are expensive:

- Steinhardt order parameters[a]
- Path/Property maps[b]
- Secondary structure CVs[c]
- SPRINT[d]
- Sketch maps[e]
- DH Energy[f]

We need a strategy to speedup!

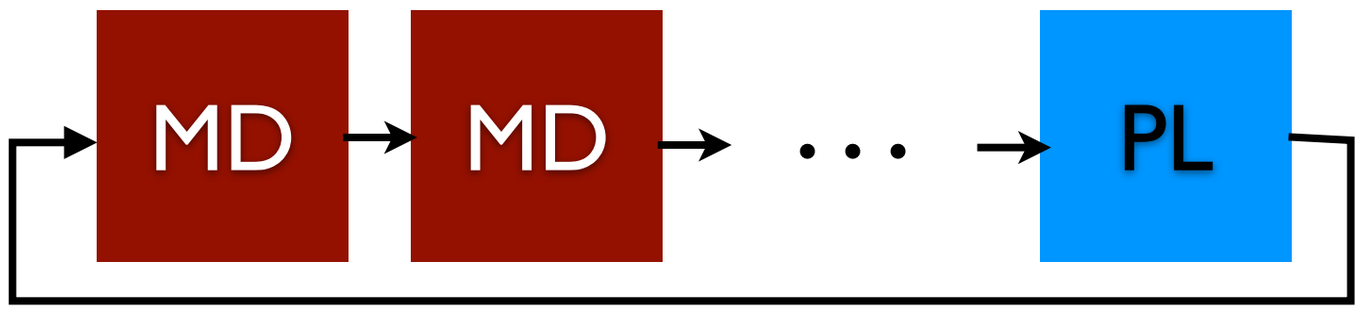
[a]Steinhardt, Nelson, and Ronchetti, PRB (1983); Trudu, Donadio, and Parrinello, PRL (2006);
[b]Branduardi, Gervasio, and Parrinello, JCP (2007); Spiwok and Králová, JCP(2011);
[c]Pietrucci and Laio, JCTC (2009); [d]Pietrucci and Andreoni, PRL (2011);
[e]Tribello, Ceriotti, and Parrinello, PNAS (2012); [f]Do, Carloni, Varani, and Bussi JCTC (2013)

Multiple time step



Multiple time step

Compute PLUMED forces every *n* steps:



$$t_{tot}=t_{MD}+t_{PL}/n$$

$$e^{A+B} \approx e^{\frac{A}{2}} e^{B} e^{\frac{A}{2}}$$

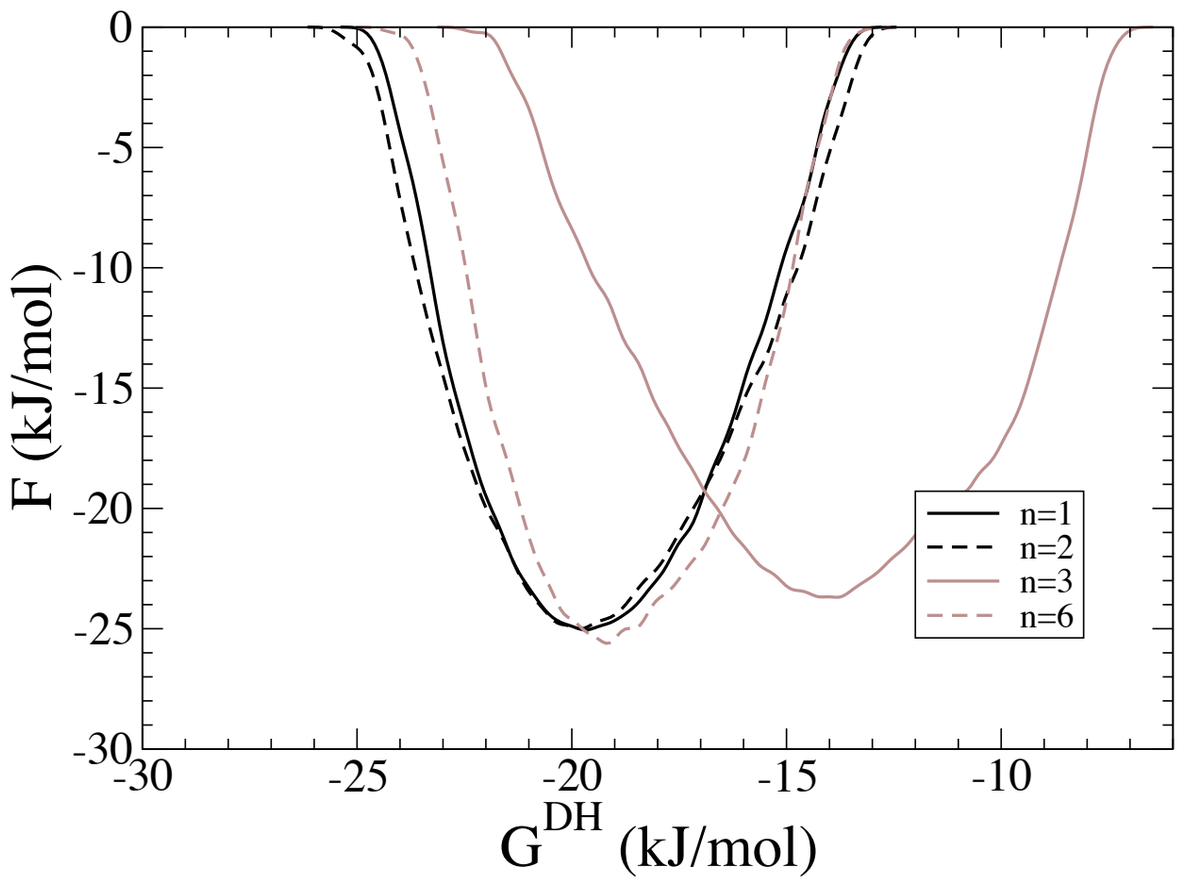
- Forces from PLUMED scaled up by a factor n[a]
- Reversible trajectories

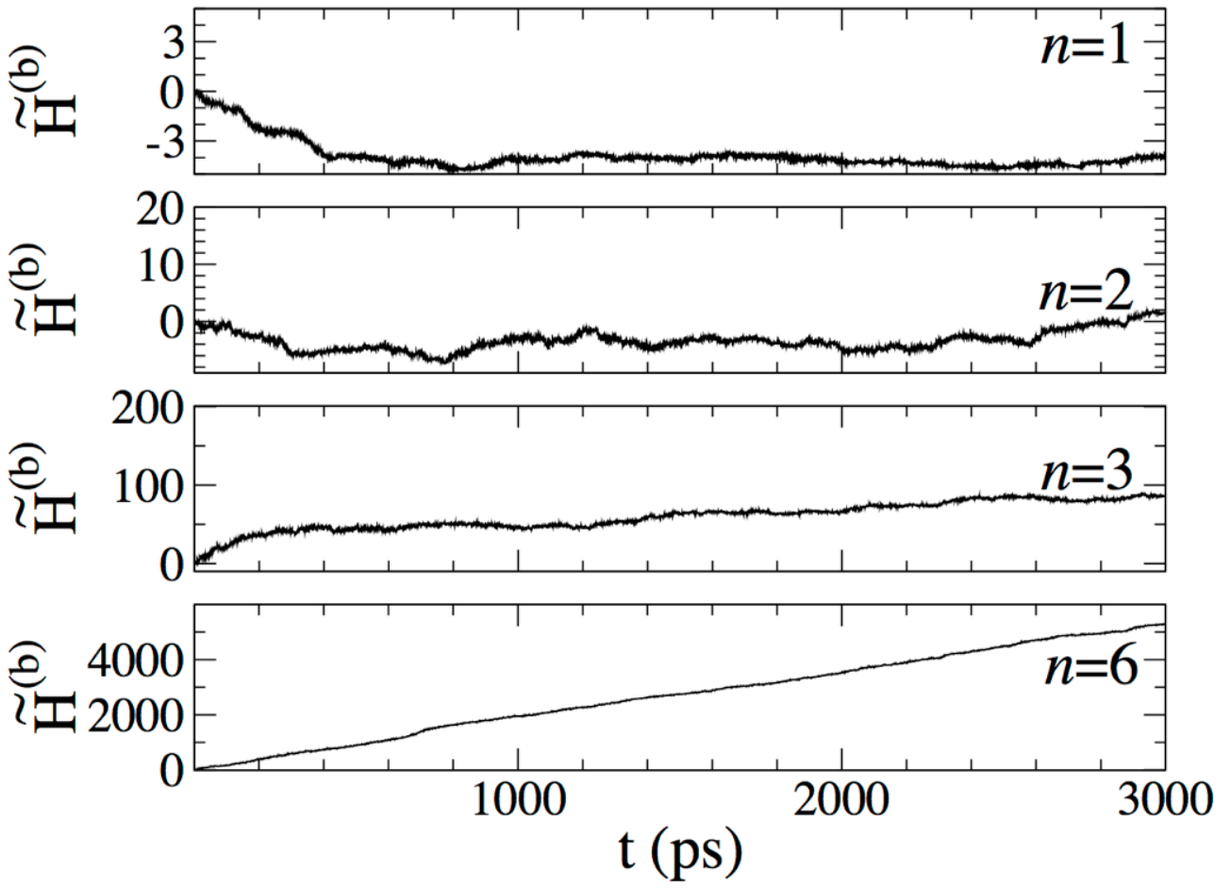
[a]Tuckerman, Berne, and Martyna, JCP (1992); Sexton and Weingarten, Nucl. Phys. B (1992)

Multiple time step

Metadyn RNA/NS3h interaction:

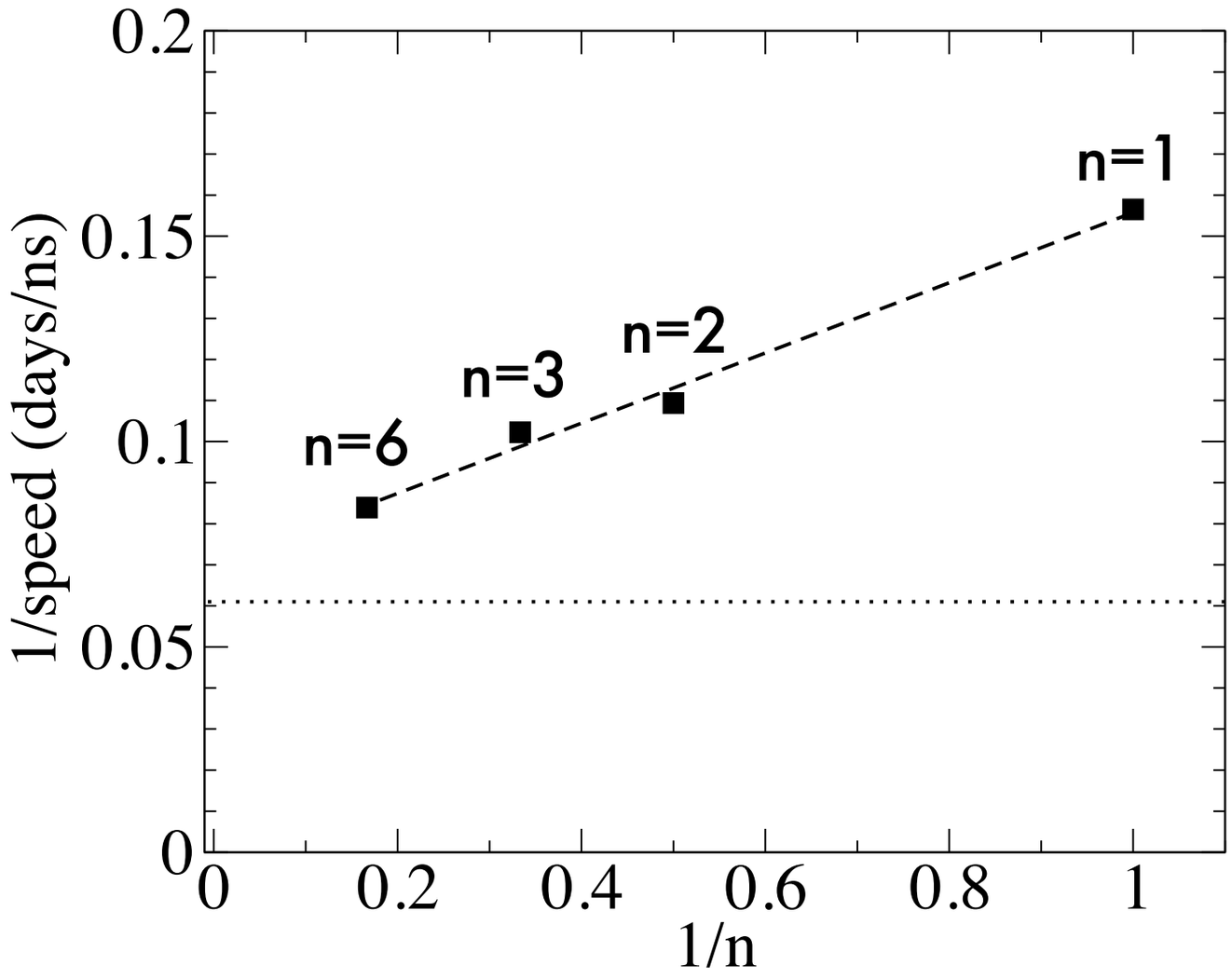
$$G^{DH} = \frac{1}{k_B T \epsilon_w} \sum_{i \in \text{prot}} \sum_{j \in \text{RNA}} q_i q_j \frac{e^{-\kappa |\mathbf{r}_{ij}|}}{|\mathbf{r}_{ij}|}$$





DH Energy as introduced in Do, Carloni, Varani, and Bussi JCTC (2013)
Ferrarotti, Bottaro, Pérez-Villa, and Bussi, JCTC (2015)

Speed up



GROMACS-only speed

PLUMED overhead can be decreased by a factor *n*
Even *n*=2 can be interesting!

Ferrarotti, Bottaro, Pérez-Villa, and Bussi, JCTC (2015)

Conclusions II

- Bias enhanced-sampling useful to tackle rare events.
- When CVs are expensive → Bottleneck in enhanced simulations ☹
- Multiple time step algorithm: A way to speedup simulations when CVs are expensive ☺
    - Splitting integration of biasing forces and physical forces.

Summary

- Translocation of NS3 helicase along RNA: Mechanism still not very clear (A way to tackle: MD simulations ☺)
- 1μs Plain MD, 6 systems. No significant conformational change.
- Analysis of HB network and contacts between RNA, peptide and ligand.
- GROMACS hybrid parallelization: faster simulations.
  - GPU accelerated non-bonded force calculations!!
- Multiple time step scheme: Strategy to speedup bias enhanced-sampling when selected CVs are expensive.

Acknowledgements

- Prof. Giovanni Bussi - SRNAS, SISSA

Thanks for your attention!