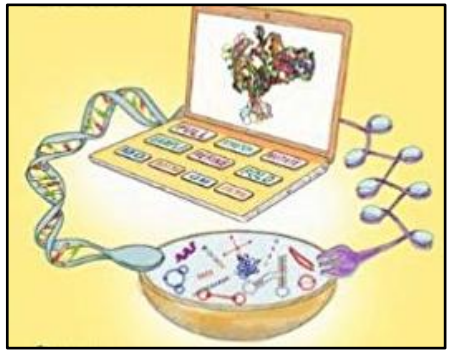
Computational *Microscopy* of Biomolecular Processes using High Performance Computing

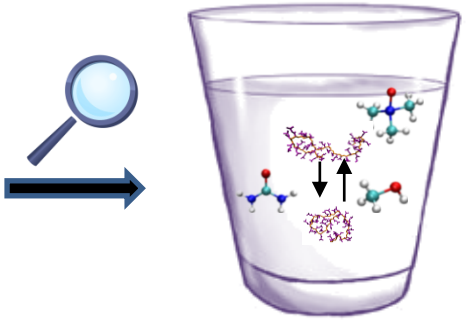
Challenges and Perspectives

Divya Nayar

Centre for Computational and Data Sciences,
Indian Institute of Technology Kharagpur



Book cover: T. Schlick



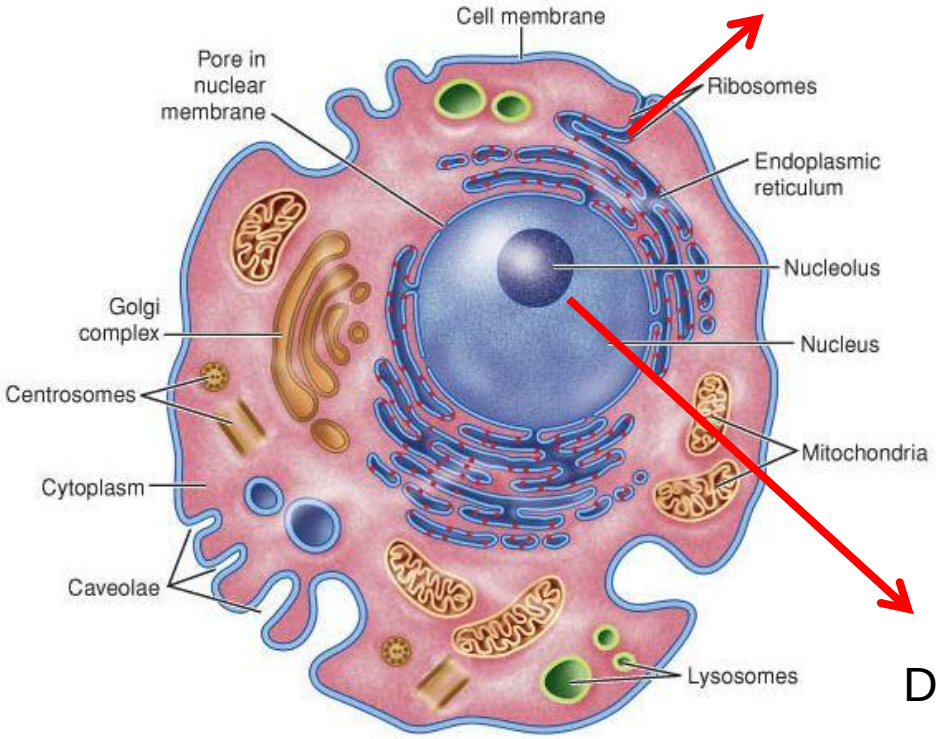
Workshop on Software Challenges to Exascale Computing
13-14 December 2018, Delhi

A living cell environment: Macromolecular crowding

~10-100 µm

Protein folding-unfolding

Exascale



DNA condensation

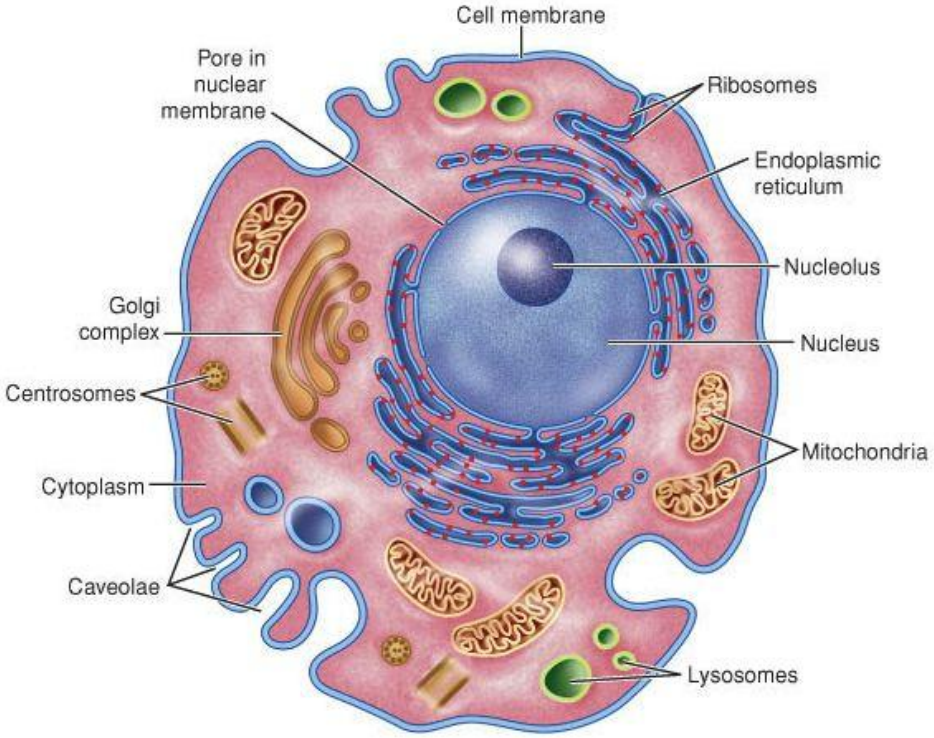
Representation of a living cell

- Steric interactions
- Water behaves differently
- Dynamics affected

Large system sizes !

A living cell environment: Macromolecular crowding

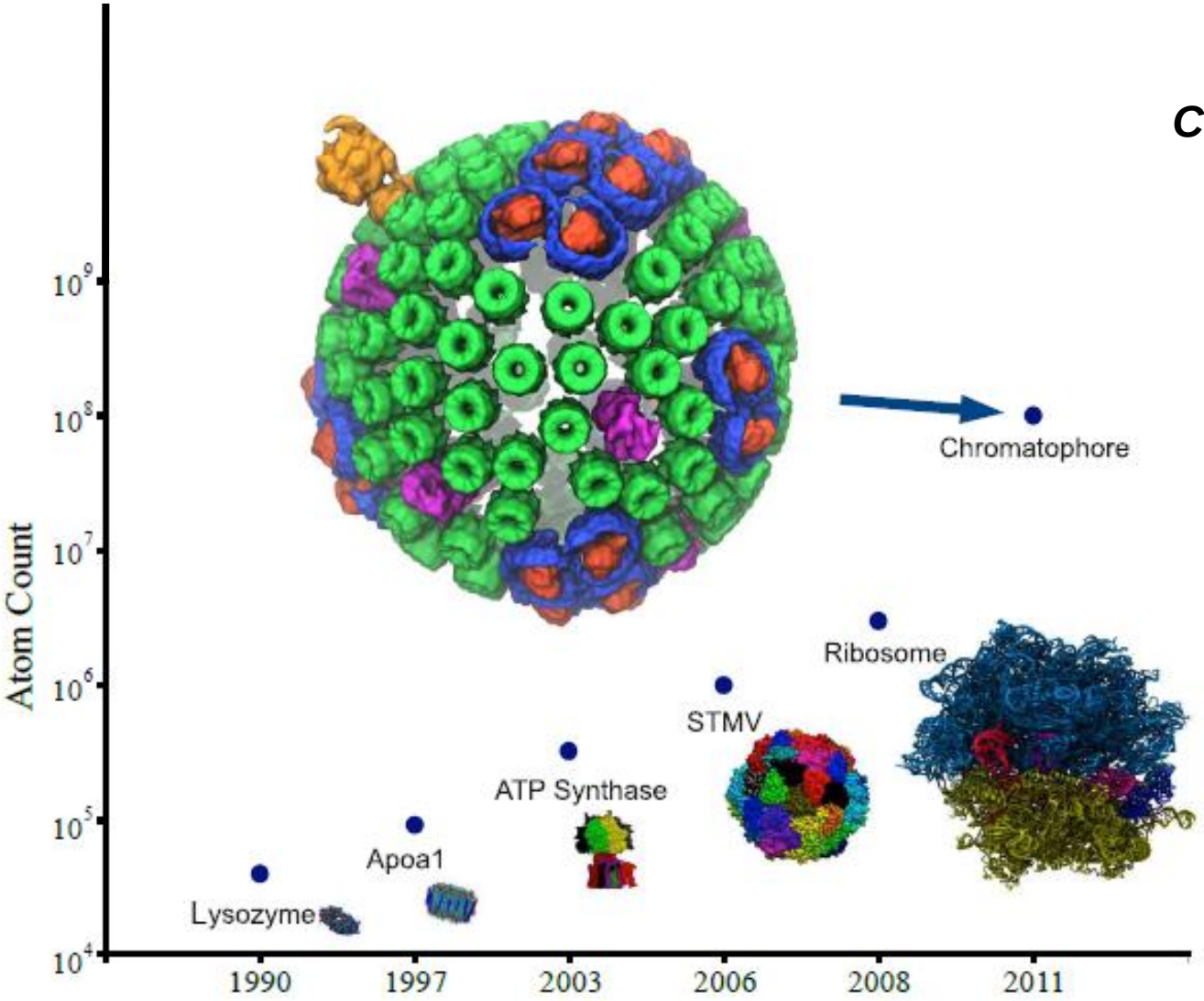
Current simulation stage: **Tera/Petascale**



Representation of a living cell

Macromolecular crowding needs to be accounted for !

Breakthroughs: Molecular-level understanding

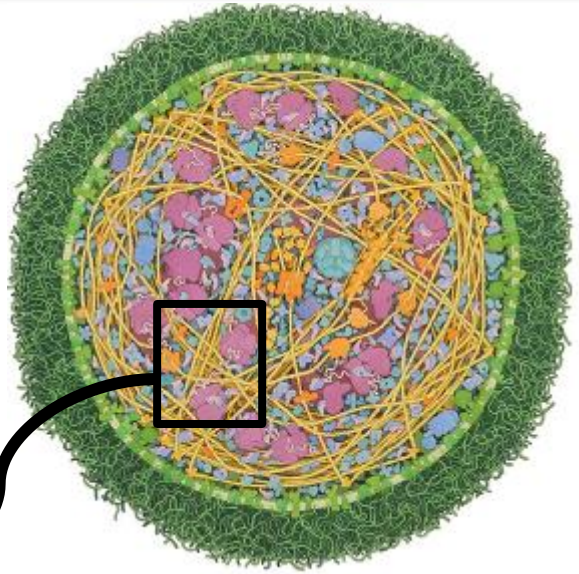


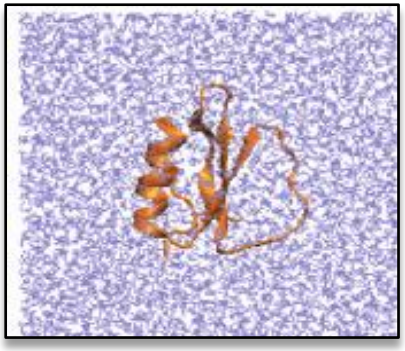
Cellular-level systems ?

Sanbonmatsu et al. *J Struct Biol.* **2007,** *157,* 470–480

Computational Challenges

- Accurate modelling
- Large system sizes: N~*10 million atoms*
- Long simulation times needed: ~ 100 *μsec*
- Large data size generated: ~ *50 TB*

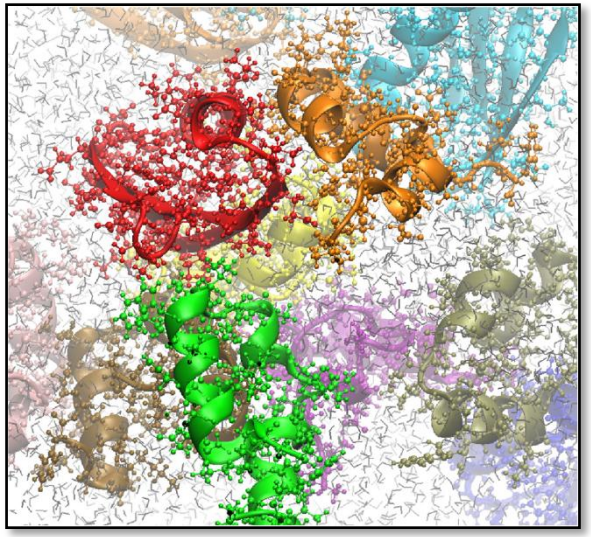




Dilute ~ $5X10^4$ atoms

Current understanding

Tera/Petascale



Crowded ~ 10^7 atoms

For complete understanding

Exascale

Needed:
-Efficient parallel simulations
-GPU acceleration
-Making MD packages efficient

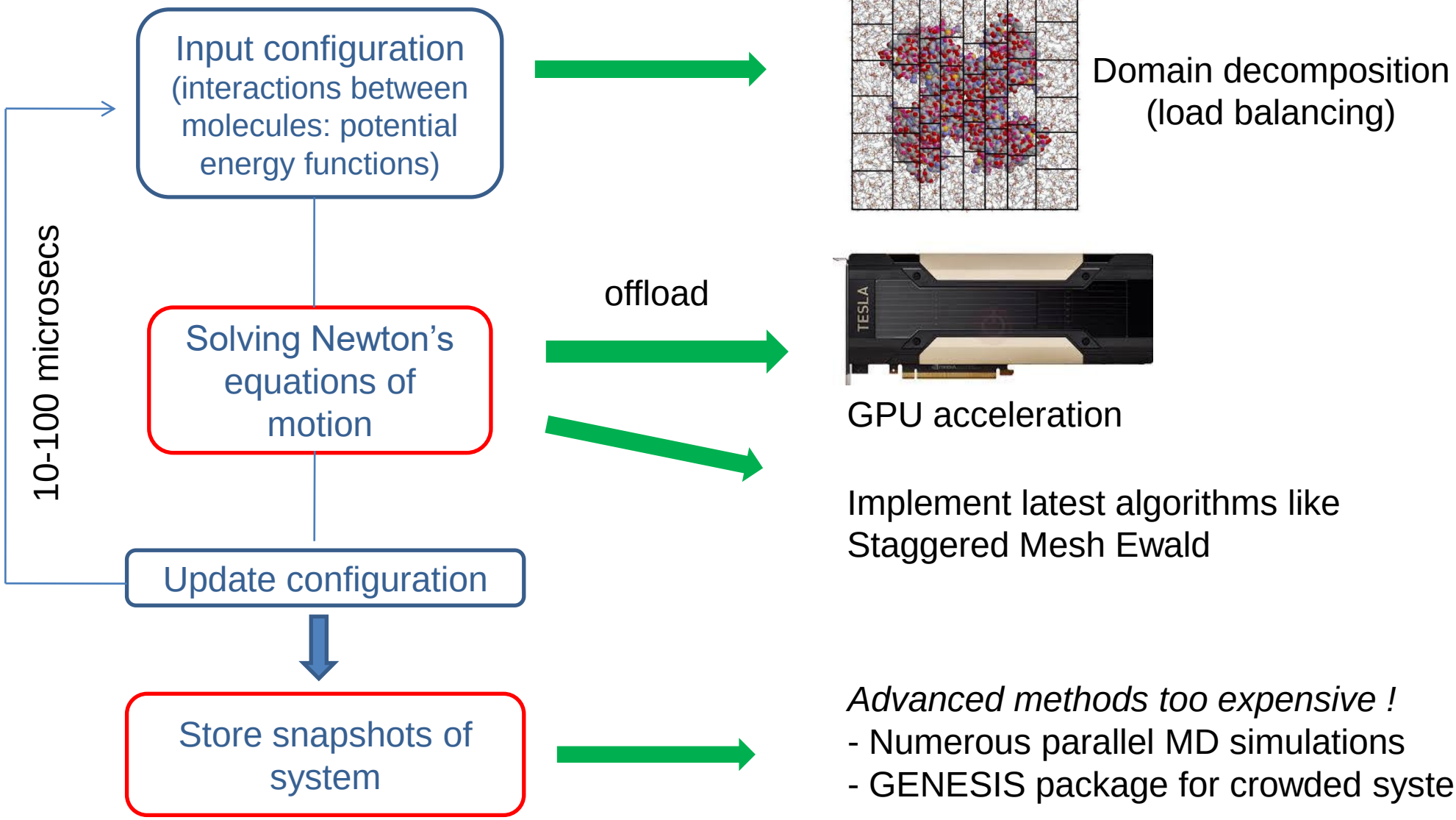
Feig et al. *J. Phys. Chem. B* 2012, 116, 599
Feig et al. *J. Mol. Graph. Model*. 2013, 45, 144

David S. Goodsell, the Scripps Research Institute (2016).
Http://mgl.scripps.edu/people/goodsell/illustration/mycoplasma

Molecular dynamics algorithm: *Make it efficient !*

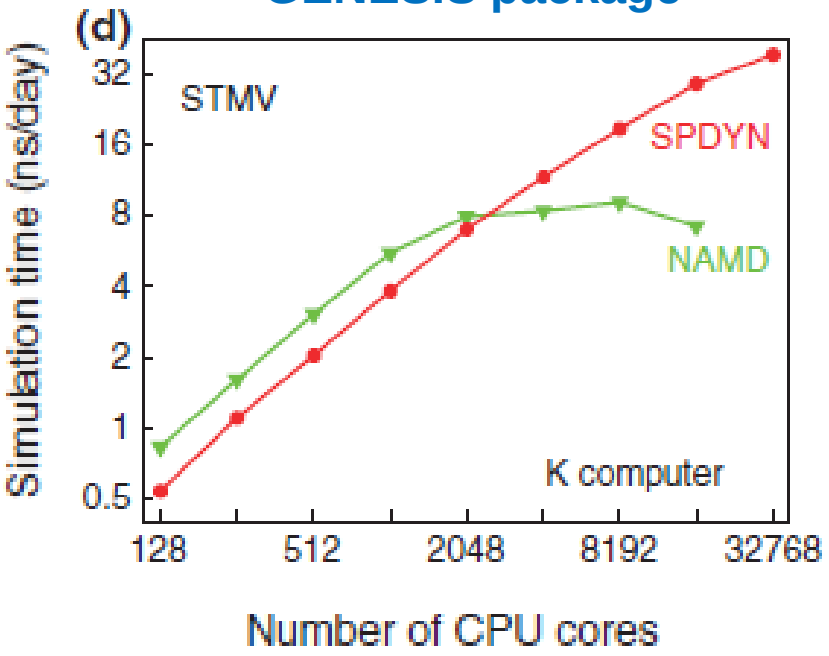
MD packages (open-source): GROMACS, NAMD, LAMMPS

Parallelization schemes: MPI, MPI+OpenMP multi-threading

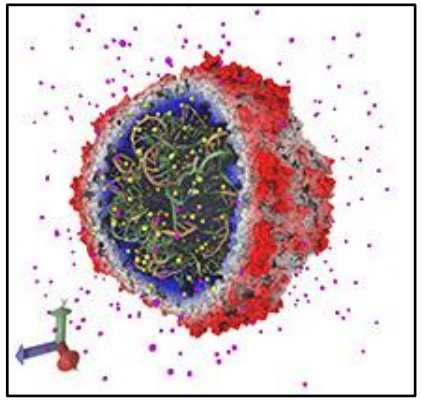


Benchmark performance of MD simulations

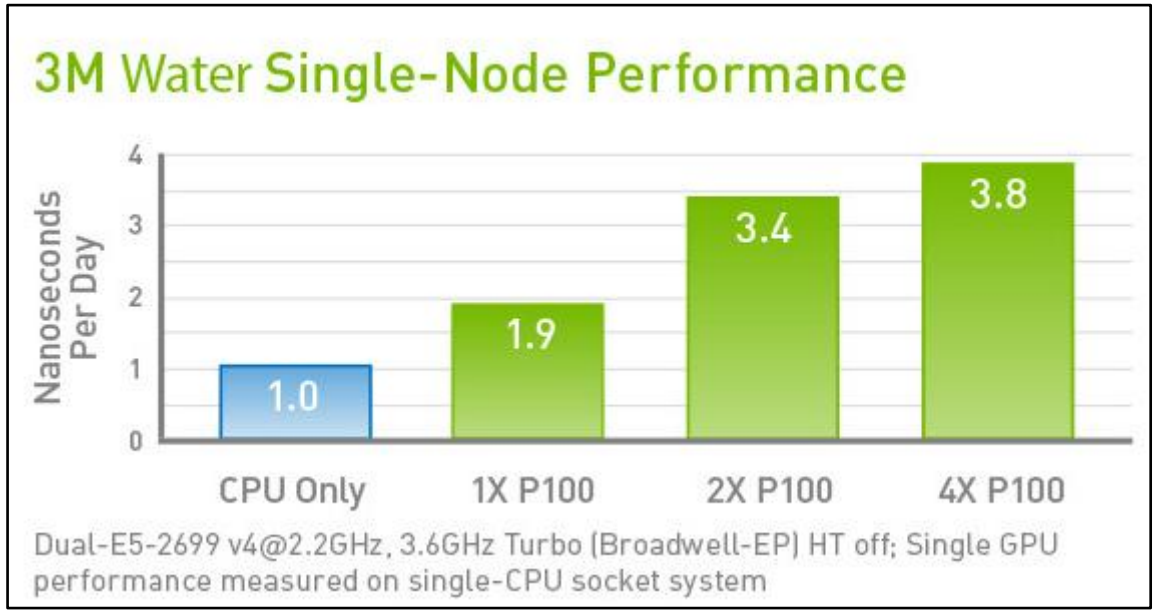
GENESIS package



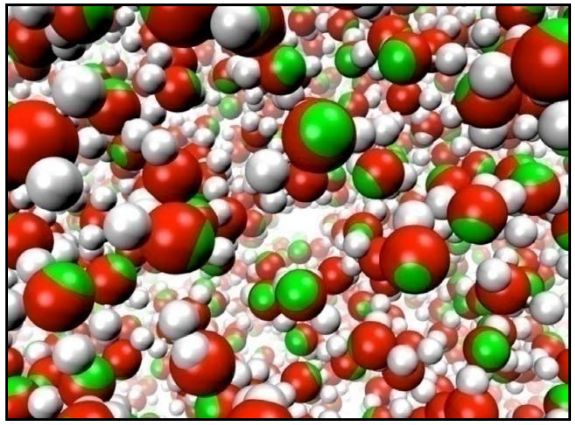
Intel Xeon E5-2690 CPUs, each with eight 2.9GHz cores ; System size: ~1 million atoms



GROMACS 5.1.2 package



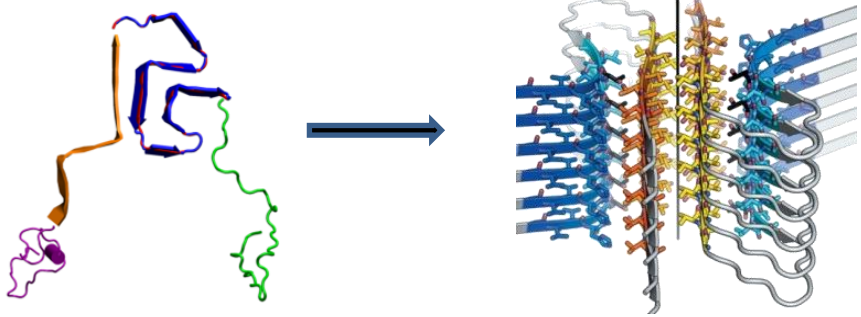
https://www.nvidia.com/en-us/data-center/gpu-accelerated-applications/gromacs



Jung et al. *WIREs Comput . Mol. Sci.* **2015**, 5, 310

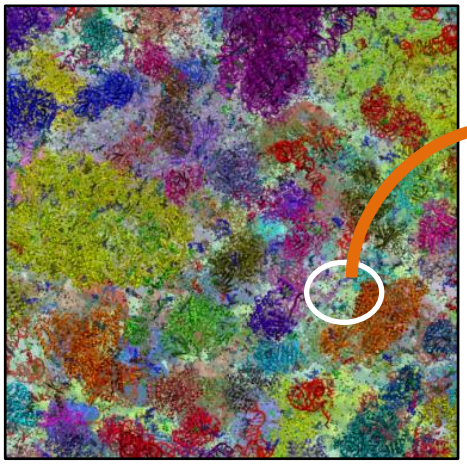
Example: Aggregation of α-synuclein protein-Parkinson's disease

Enabled predicting binding free energy to form Amyloid



α –synuclein monomer

Amyloid aggregate (Parkinson's disease)

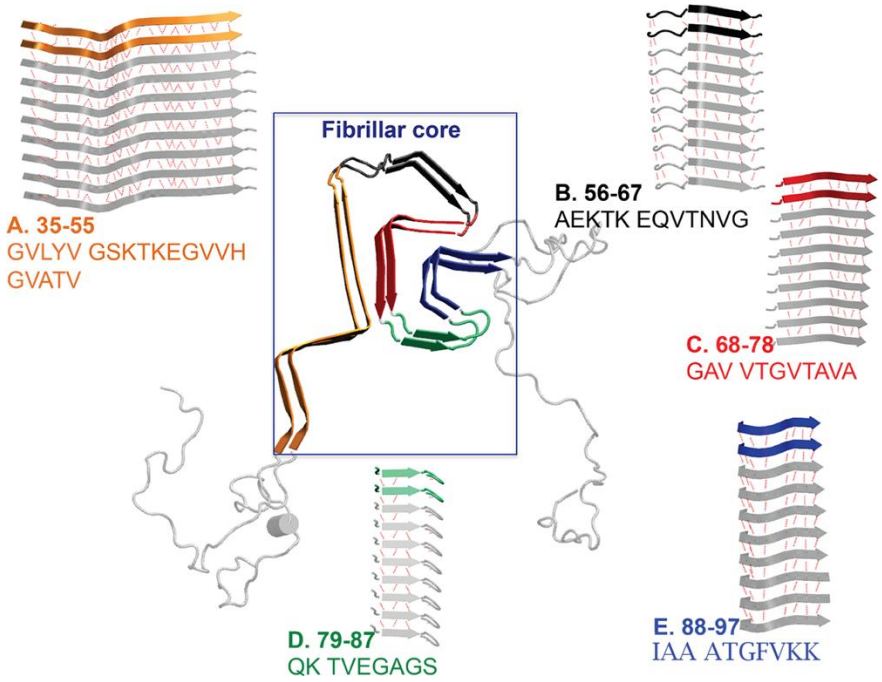


Next step:

Realistic cellular environment

Challenges to be addressed !

Parallel MD simulations of dimers using HPC



- **Dilute solutions !!**
- **Only dimers studied**
- **~ 2 µsec (Petascale)**
- **GROMACS 2016**

Ilie, I.M.; Nayar, D. et al. *J. Chem. Theory Comput.* **2018**, 14, 3298

Centre for Computational and Data Sciences (CCDS)
IIT Kharagpur

(Estd. March, 2017)

- **1.3 Peta-Flop** Supercomputing facility : *National Supercomputing Mission (NSM)*.
- IIT Kharagpur: Nodal Centre for the HR-development activities
- Interdisciplinary Centre
- *Faculty working in different HPC application domains*: Computational Chemistry/Biology, Material science, Atmospheric Modeling, Computational Fluid Dynamics, Geo-Scientific Computations, Modeling and Mining of Heterogeneous Information Network, Computational Physics, Cryptanalysis, Numerical Mathematics, Computational Mechanics, Non-equilibrium Molecular Dynamics
- Interdisciplinary teaching for Ph.D./ Master's students

Acknowledgements

Prof. Nico van der Vegt: Technische Universitaet (TU) Darmstadt, Germany

Prof. Wim Briels: University of Twente, Netherlands

Dr. Ioana M. Ilie: University of Zurich

Dr. Wouter K. den Otter: University of Twente, Netherlands

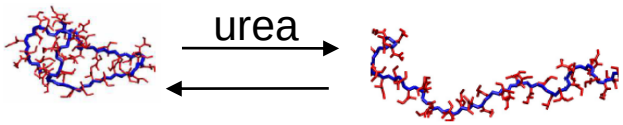
Computational facility: Lichtenberg HPC Cluster, TU Darmstadt

Organizers of SCEC 2018

Thank you for your attention !

Example 1: Parallel MD simulations protocol

Polymer in aqueous urea solution



Big question: How do cosolvents protect proteins in the cell under extreme conditions ?

Polymer System	System size (atoms)	No. of parallel simulations	Total simulation time per concentration	Total CPU time (core-hours)	Wall clock time per run of 20 ns (hrs)	CPU memory per core
PNiPAM	26000	1800	4 µs	648000	9	200 MB
PDEA	72000	2000	4 µs	3456000	20	400 MB
Total		3800	8 µs	~4.1 million		

MD package: GROMACS 4.6.7 (MPI enabled, 64-bit)

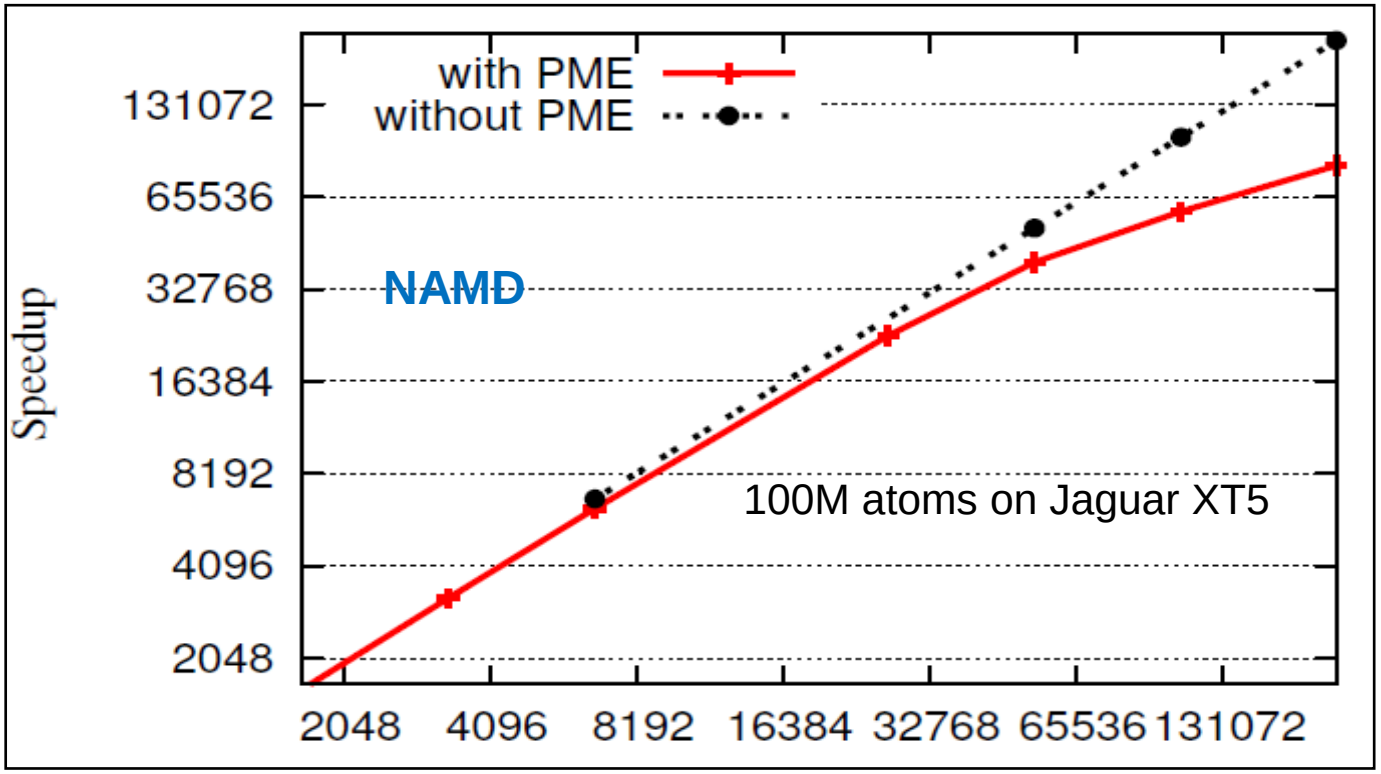
Hardware: Intel(R) Xeon(R) CPU E5-4650 @ 2.70GHz

CPU accelerator: avx2

- Particle Mesh Ewald: electrostatics
- Domain decomposition

Computational resources: Lichtenberg High Performance Computing Cluster, TU Darmstadt

Nayar et al. *Phys. Chem. Chem. Phys*., 2017, 19, 18156.

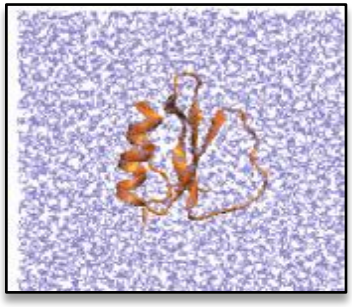


http://www.ks.uiuc.edu/Training/Workshop/Bremen/lectures/day1/Day1b_MD_intro.key.pdf

Molecular dynamics (MD) simulations..

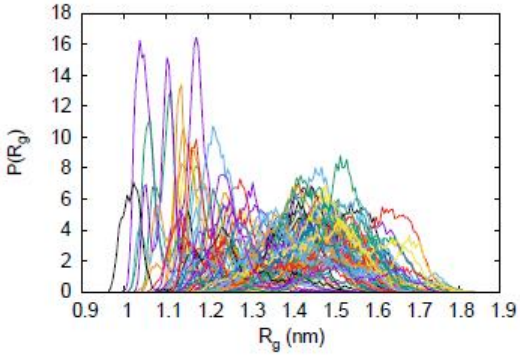
Protein structure

Build →



Simulate →

Analyze →

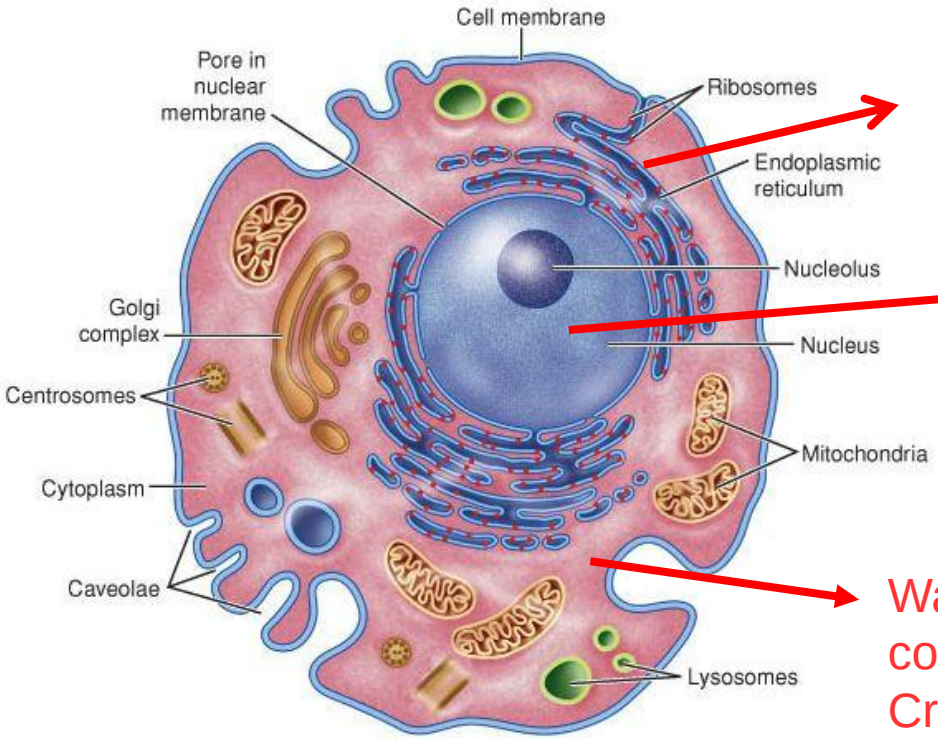


MD packages
(open-source):
GROMACS,
NAMD, LAMMPS

Parallelization schemes:
-MPI
-MPI+OpenMP multi-threading

A living cell: Crowded environment !

~10-100 µm



Protein folding-unfolding

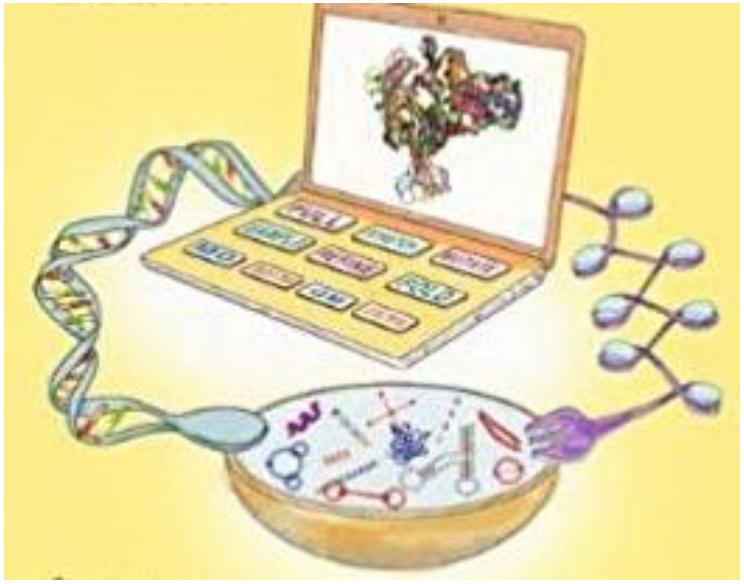
DNA condensation

Water, ions cosolvent, Crowders

Representation of a living cell

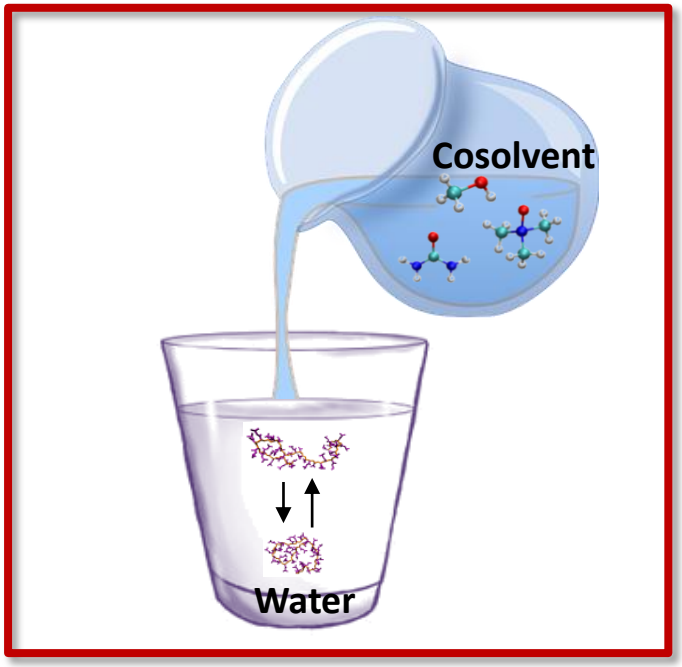
Microscope

Our Computational *Microscope:* Molecular dynamics simulations



Molecular simulations

High Performance Computing



Elucidating molecular mechanisms

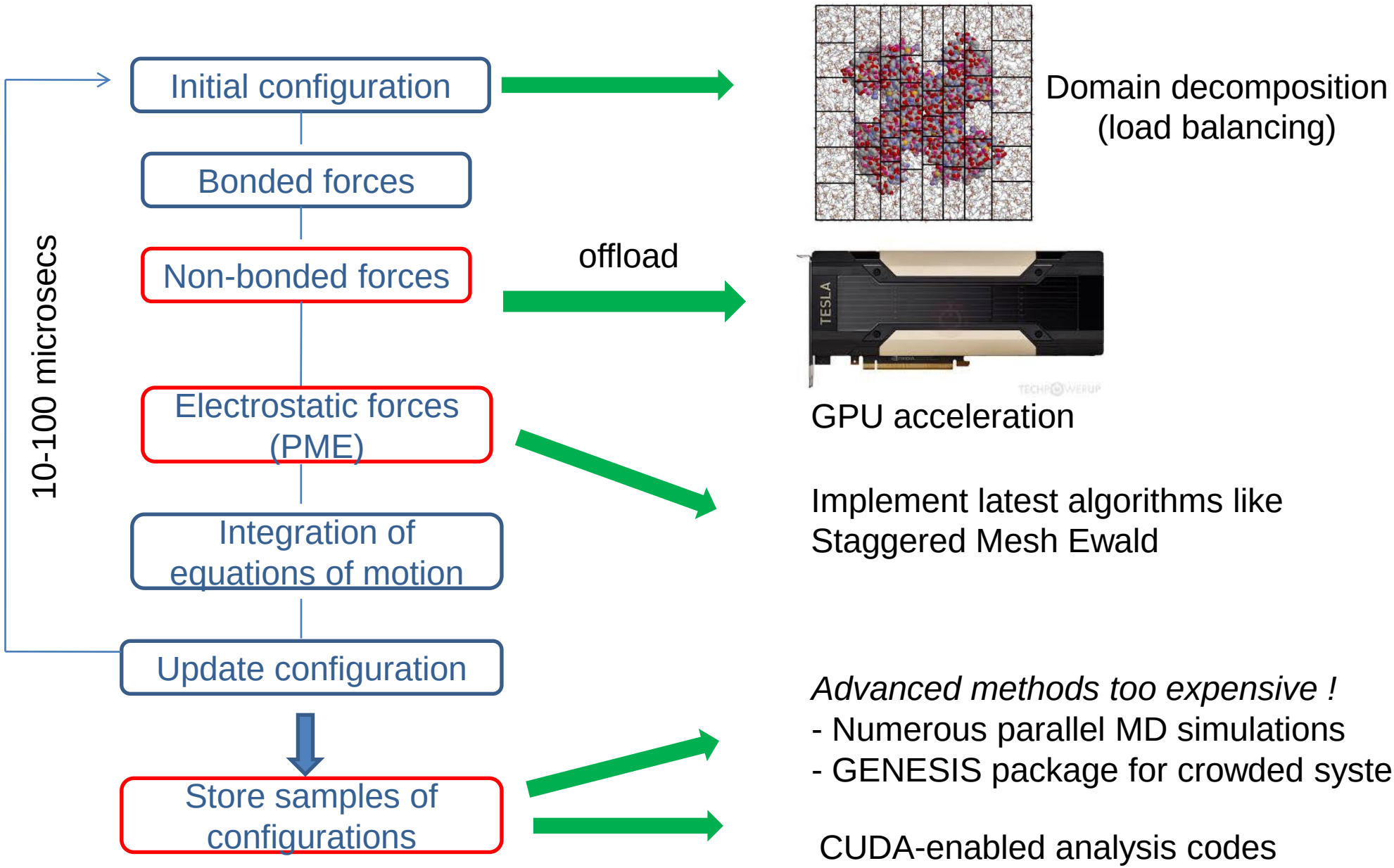
Book cover: Molecular Modeling and Simulation: An Interdisciplinary Guide; Tamar Schlick

van der Vegt, N.F.A.; Nayar, D. *J. Phys. Chem. B* **2017**, 121, 9986

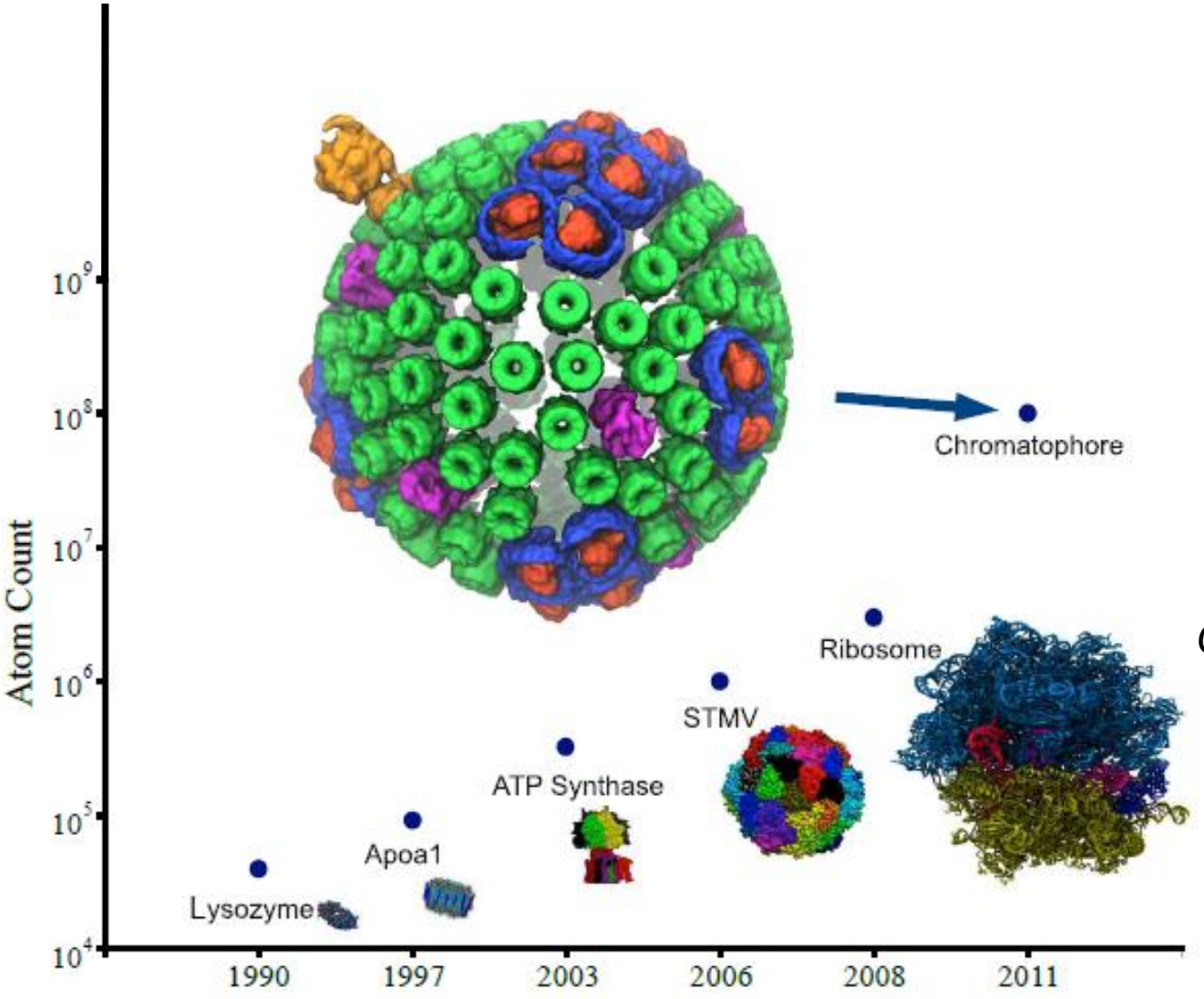
Molecular dynamics algorithm: *Make it efficient !*

MD packages (open-source): GROMACS, NAMD, LAMMPS

Parallelization schemes: MPI, MPI+OpenMP multi-threading

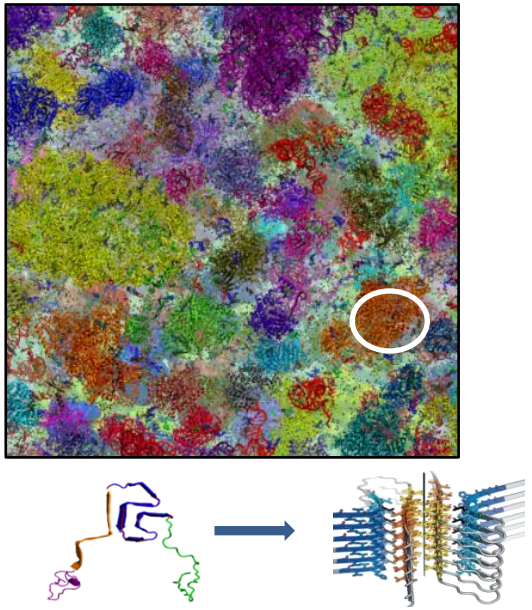


Other breakthroughs: Molecular-level understanding



Next step:

Realistic cellular environment



Exascale computing !!

Challenges to be addressed

Sanbonmatsu et al. *J Struct Biol.* **2007,** *157,* 470–480