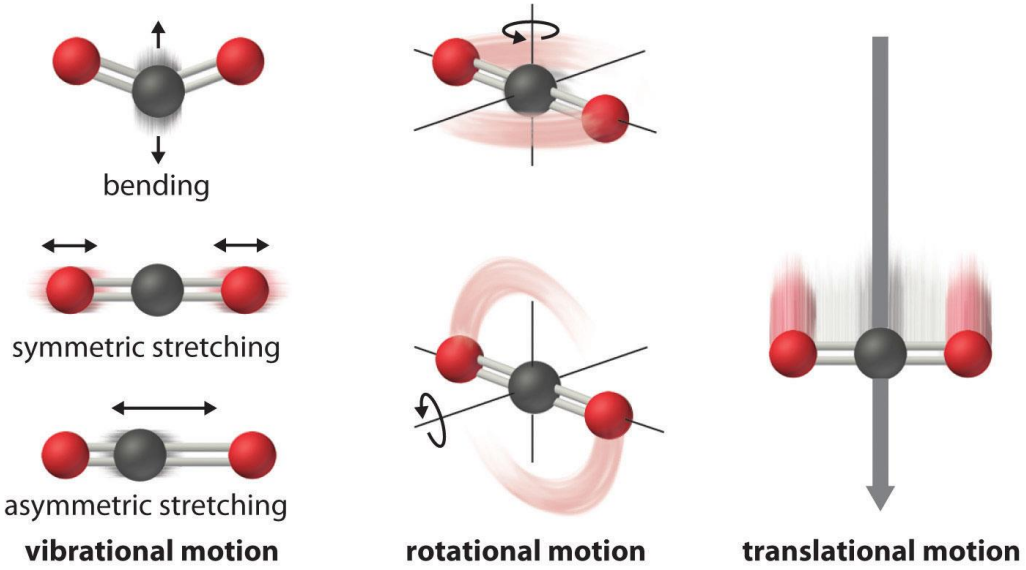
Molecular Dynamics

Molecular Dynamics: Definition

Molecular dynamics provide an alternative approach to determine the preferred conformers and the global minimum of a molecule. This is achieved by the simulation of the dynamical motions of the molecule as it vibrates and undergoes internal rotation.



Molecular Dynamics: Definition

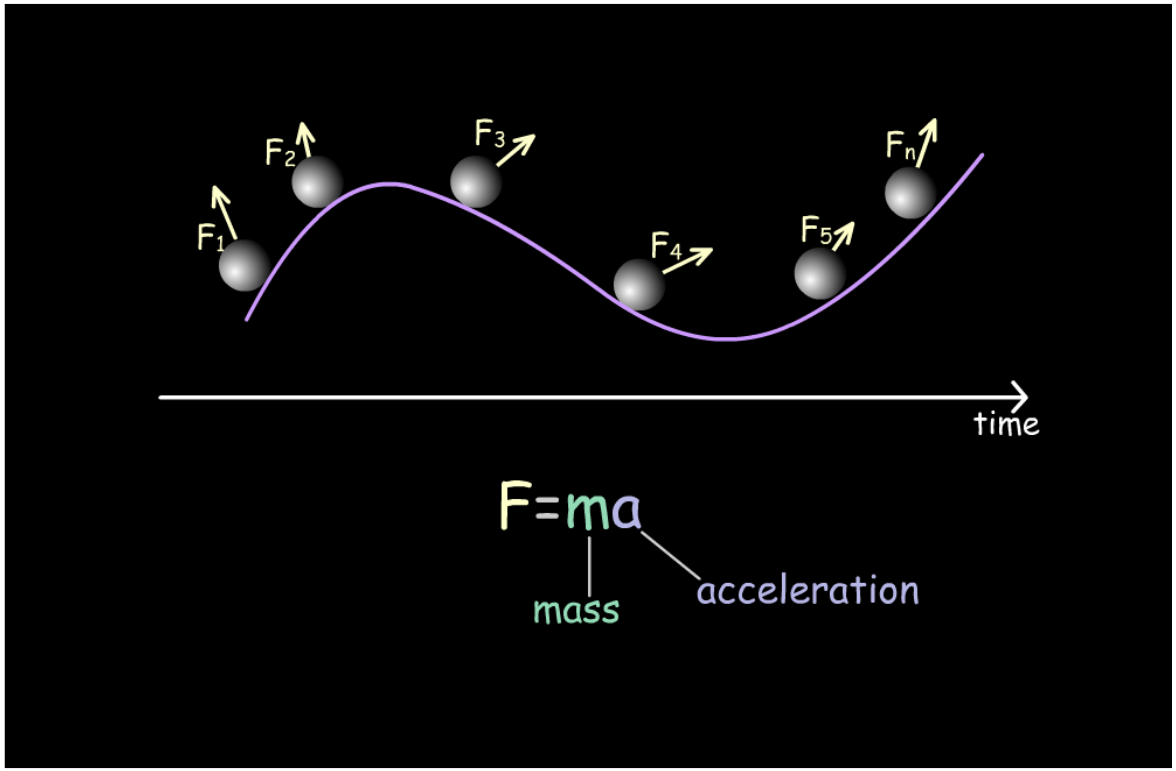
MD is an extension of the molecular mechanics approach and it is based on the idea that the atoms of the molecule feel forces and want to move.

Each atom is treated as a particle responding to Newton's equation of motion:

$$F = m \times a$$

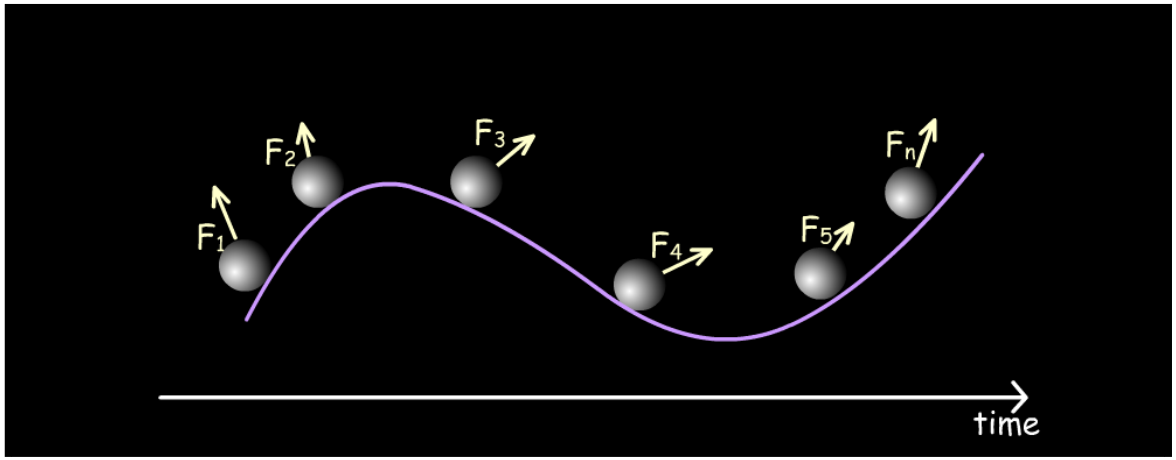
Molecular Dynamic: Definition

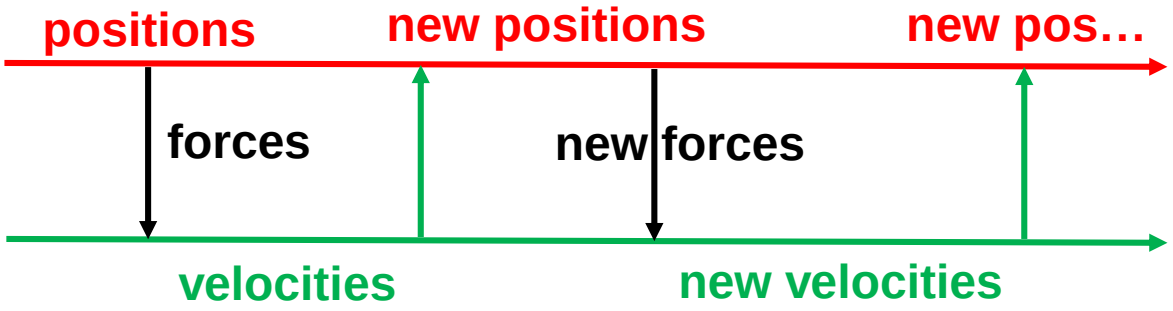
Integration of these equations with successive time steps leads to the trajectory of the atoms over time in the form of a list of positions and velocities.



Molecular Dynamic: Definition

Integration of these equations with successive time steps leads to the trajectory of the atoms over time in the form of a list of positions and velocities.





Molecular Dynamics Simulations

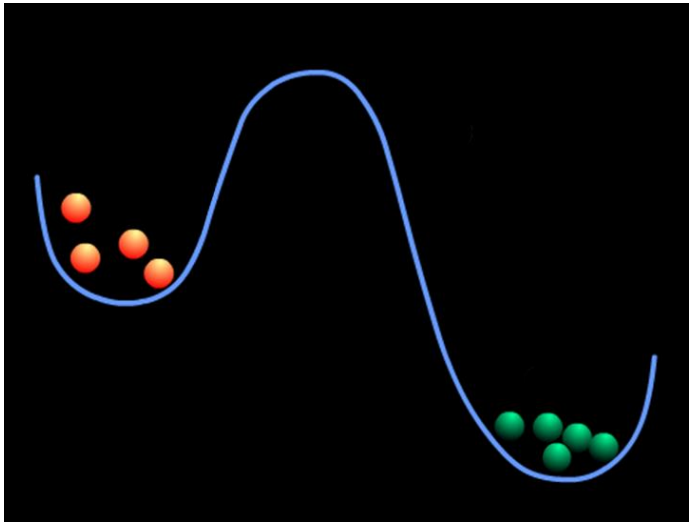
- One of the principal tools for modeling proteins, nucleic acids and their complexes.
- Stability of proteins
- Folding of proteins
- Molecular recognition by proteins, DNA, RNA, lipid membranes, etc.
- Enzyme reactions
- Rational design of biologically active molecules (drug design)
- Small and large-scale conformational changes.
- Determination and construction of 3D structures (homology modeling, X-ray diffraction, NMR, Cryo-EM)
- Dynamic processes such as ion transport in biological systems

…and much more

MD for conformational search

Determining preferred or low-energy conformations of biological molecules can be a very difficult task.

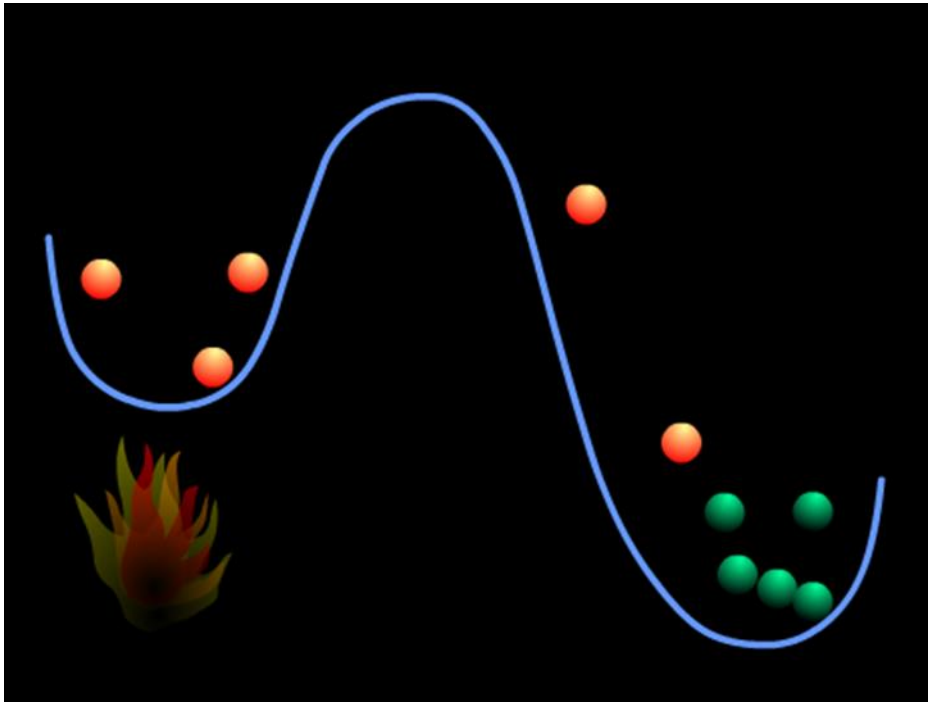
Local minima of conformational energy are separated by barriers that minimization algorithms are unable to overcome.



MD can solve this problem

MD for conformational search

The available energy of the molecule is distributed between potential and kinetic energy, and molecules are thus able to overcome barriers separating minima if the barrier height is less than the kinetic energy.



MD for conformational search

Given a high enough energy (which is closely related to the simulation temperature),

$$E_{\mathrm{K}} = \frac{1}{2} m \langle v^2 \rangle = \frac{f}{2} k_{\mathrm{B}} T$$

the dynamics can virtually sample the whole surface, but this would require an impractically long simulation time due to the nature of the MD calculation.

Molecular Dynamics Simulations

The potential energy is a function of the atomic positions of all the atoms in the system. Due to the complicated nature of this function, there is no analytical solution to the equations of motion; they must be solved numerically.

Various numerical algorithms have been developed for integrating the equations of motion:

- Verlet
- Velocity Verlet
- LeapFrog
- Beeman

With these algorithms the molecule motion is divided into time steps, Δt.

Molecular Dynamics Simulations

This implies a strong limit: the Δt step must be short enough so that during Δt the variation of V is negligible.

Therefore Δt cannot be longer than the most common motions of the atoms.

Δt cannot be longer than the fastest atomic motion and this means:

$$\Delta t = 10^{-15} \text{ sec (1 femtosecond)}$$

Consequently, a simulation of a microsecond needs one billion steps.

Molecular Dynamics Simulations

Since quite small time steps must be used, the simulation time is short (usually tens to some hundreds of nanoseconds).

This means that, combined with the use of reasonable temperatures (few hundreds of degrees), only the local area around the starting point can be sampled, and only relatively small barriers (few kcal/mol) can be overcome.

Different local minima may be generated by selecting configurations at suitable intervals during the simulation and subsequently minimizing these structures.

Treatment of the solvent

If we want to simulate and study the behavior of protein, DNA, RNA and small-molecule structures in a realistic environment, we cannot avoid considering the solvent as a part of the system.

For simulating a physiological-like system, the solvent to be considered is water, which constitutes the medium of intracellular and extracellular environments. Ions must be considered too, at least in the proper amount to ensure system neutrality.

The question is: how should we consider solvent within our system and balance accuracy with computation time?

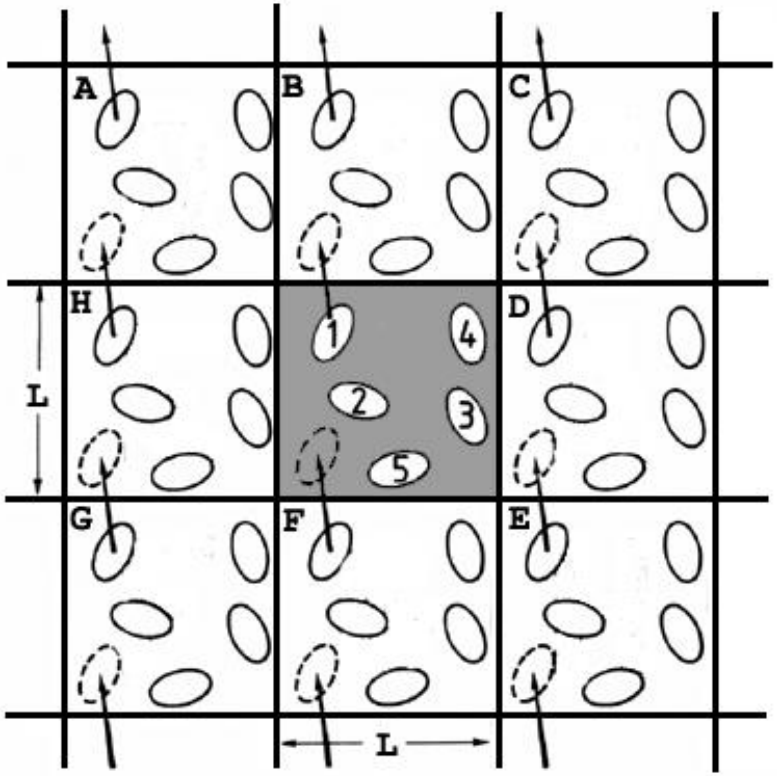
Treatment of the solvent

- **Implicit:** The macromolecule interacts only with itself, but the electrostatic interactions are modified to account for the solvent.
  - Choice of dielectric constant (Vacuum ε=1, Solute ε=2-20, Water ε=80)
  - Advanced treatment of solvent: Generalized Born models, Poisson-Boltzmann model (continuum models)
- **Explicit:** The macromolecule is surrounded by solvent molecules (water and ions) with which the macromolecule interacts. Specific nonbonded interactions are calculated. More reliable (fewer approximations) but more computationally expensive.

Periodic Boundaries

A realistic model of a solution requires a very large number of solvent molecules to be included along with the solute. Simply placing the solute in a box of solvent is not enough: while some solvent molecules will be at the boundary between solute and solvent, and others within the bulk of the solvent, a large number will be at the edge between the solvent and the surrounding vacuum. This is obviously a non-realistic picture of a bulk fluid. In order to prevent the outer solvent molecules from boiling off into space, and to allow a relatively small number of solvent molecules to reproduce the properties of the bulk, periodic boundary conditions are employed. In this method, the particles being simulated (the system) are enclosed in a box that is replicated in all three spatial dimensions to give a periodic array.

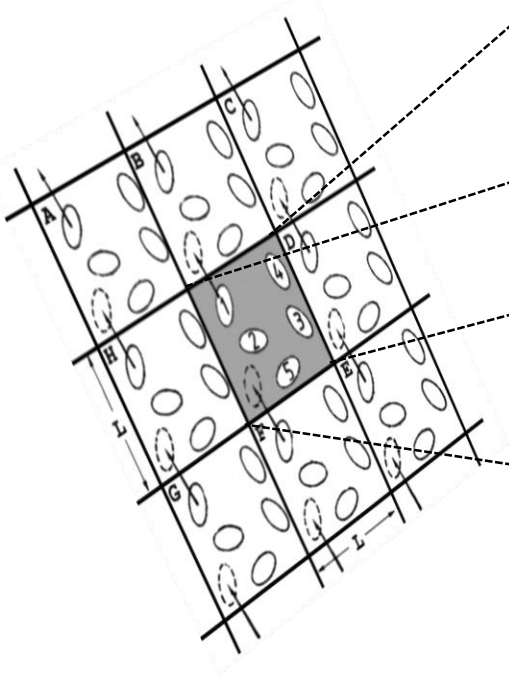
Periodic Boundaries



Periodic boundary scheme

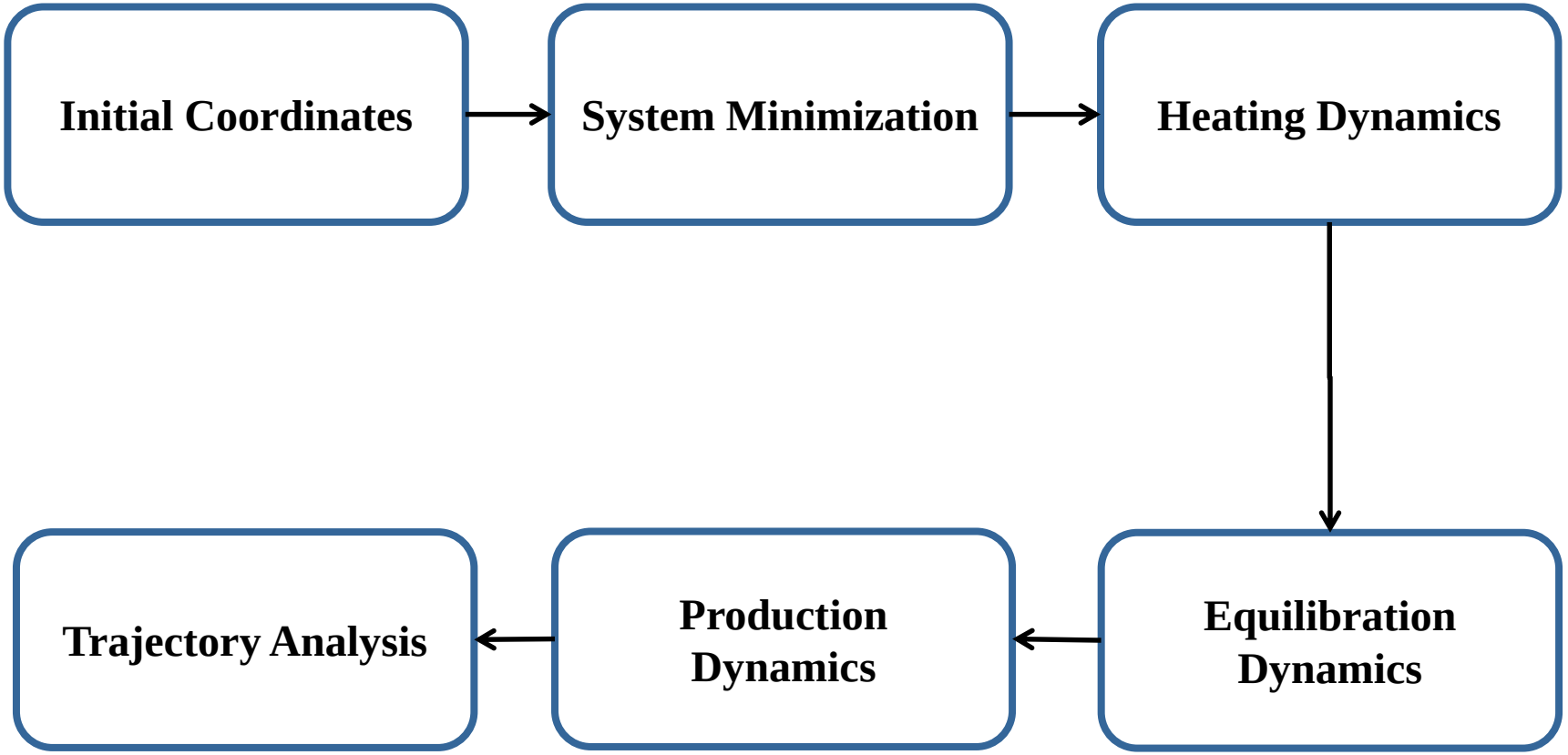
During the simulation only one of the infinite systems (in only one box) is represented, but the effects are reproduced over all system images, with the particles not only interacting with the other particles within the same box, but also with the images in the neighboring boxes. Particles that leave one side of the box re-enter from the opposite side as their image. In this way, the total number of particles in the central box remains constant.

Periodic Boundaries



Central box replicated in all directions

Steps of a MD simulation



Steps of a MD simulation

The motions of the atoms and chemical groups obtained by molecular dynamics simulation reveal a complex underlying molecular machinery.

In the beginning of the simulation, the motions are frequently interrupted by collisions with neighboring groups and each group seems to have an erratic trajectory.

However, over longer periods of time, coherent and collective motions start to develop, revealing how some groups can fluctuate somewhat more than others.

Heating MD

```
First MD
 &cntrl
  imin   = 0,
  irest  = 0,
  ntx    = 1,
  ntb    = 1,
  cut    = 10,
  ntr    = 1,
  ntc    = 2,
  ntf    = 2,
  tempi  = 0.0,
  temp0  = 300.0,
  ntt    = 3, gamma_ln = 5.0,
  nstlim = 250000, dt = 0.002,
  ntpr = 1000, ntwx = 5000,
  ntwr = 5000
  /
```

imin = 0 ::: Flag to run minimization, = 0 No minimization (only do molecular dynamics), = 1 Perform minimization (and no molecular dynamics) = 5 Read in a trajectory for analysis.

irest = 0 ::: Flag to restart the run, = 0 Noeffect (default), = 1 restart calculation. Requires velocities in coordinate input file, so you also may need to reset NTX if restarting MD

ntx = 1 ::: Option to read the initial coordinates, velocities and box size, = 1 X is read formatted with no initial velocity information, = 2 X is read unformatted with no initial velocity information, = 4 X and V are read unformatted, = 5 X and V are read formatted; box information will be read if ntb>0, the velocity information will only be used if *irest*=1, = 6 X, V and BOX(1..3) are read unformatted; in other respects, this is the same as option "5".

ntb = 1 ::: Periodic boundary, = 0 no periodicity is applied and PME is off, = 1 constant volume (default), = 2 constant pressure

cut = 10 ::: nonbonded cutoff, in Angstroms.

ntr = 1 ::: Flag for restraining specified atoms in Cartesian space using a harmonic potential, = 0 No position restraints, = 1 restraint of specified atoms

ntc = 2 ::: Flag for SHAKE to perform bond length constraints, = 1 SHAKE is not performed (default), = 2 bonds involving hydrogen are constrained, = 3 all bonds are constrained (not available for parallel or qmmm runs in *sander*)

ntf = 2 ::: Force evaluation. Note: If SHAKE is used (see NTC), it is not necessary to calculate forces for the constrained bonds., = 1 complete interaction is calculated (default), = 2 bond interactions involving H-atoms omitted (use with NTC=2), = 3 all the bond interactions are omitted (use with NTC=3), = 4 angle involving H-atoms and all bonds are omitted, = 5 all bond and angle interactions are omitted, = 6 dihedrals involving H-atoms and all bonds and all angle interactions, are omitted, = 7 all bond, angle and dihedral interactions are omitted, = 8 all bond, angle, dihedral and non-bonded interactions are omitted

tempi = 0 and temp0 = 300 ::: Initial temperature and reference temperature at which the system is to be kept.

ntt = 3 ::: Switch for temperature scaling, = 0 Constant total energy classical dynamics (assuming that *ntb*<2, as should probably always be the case when *ntt=0*), = 1 Constant temperature, using the weak-coupling algorithm, = 2 Andersen temperature coupling scheme, = 3 Use Langevin dynamics with the collision frequency γ given by *gamma_ln*

gamma_ln = 5.0 ::: The collision frequency γ, in ps-1, when *ntt* = 3.

nstlim = 100000 ::: Number of MD-steps to be performed.

dt = 0.002 ::: The time step (psec).

ntpr = 100 ::: Every NTPR steps energy information will be printed in human-readable form to files "mdout" and "mdinfo

ntwx = 1000 ::: Every NTWX steps the coordinates will be written to file "mdcrd".

ntwr = 1000 ::: Every NTWR steps during dynamics, the "restrt" file will be written.

Equilibration/Production MD

```
Second MD
 &cntrl
  imin = 0, irest = 1,
  ntx = 7, ntp = 2,
  ntb = 2, pres0 = 1.0,
  taup = 2.0,
  cut = 10, ntr = 0,
  ntc = 2, ntf = 2,
  tempi = 300.0,
  temp0 = 300.0,
  ntt = 3, gamma_ln = 5.0,
  nstlim = 5000000, dt = 0.002,
  ntpr = 5000, ntwx = 5000,
  ntwr = 5000
  /
```

imin = 0 ::: Flag to run minimization, = 0 No minimization (only do molecular dynamics), = 1 Perform minimization (and no molecular dynamics) = 5 Read in a trajectory for analysis.

irest = 1::: Flag to restart the run, = 0 Noeffect (default), = 1 restart calculation. Requires velocities in coordinate input file, so you also may need to reset NTX if restarting MD

ntx = 7 ::: Option to read the initial coordinates, velocities and box size, = 1 X is read formatted with no initial velocity information, = 2 X is read unformatted with no initial velocity information, = 4 X and V are read unformatted, = 5 X and V are read formatted; box information will be read if ntb>0, the velocity information will only be used if *irest*=1, = 6 X, V and BOX(1..3) are read unformatted; in other respects, this is the same as option "5".

ntb = 2 ::: Periodic boundary, = 0 no periodicity is applied and PME is off, = 1 constant volume (default), = 2 constant pressure

pres0 = 1.0 ::: Reference pressure (atm)

ntp = 2 ::: Flag for constant pressure dynamics, = 0 Used with NTB not = 2 no pressure scaling, =1 md with isotropic position scaling, = 2 md with anisotropic (x-,y-,z-) pressure scaling.

taup = 2.0 ::: Pressure relaxation time (in ps)

cut = 10 ::: nonbonded cutoff, in Angstroms

ntr = 0 ::: Flag for restraining specified atoms in Cartesian space using a harmonic potential, = 0 No position restraints, = 1 restraint of specified atoms

ntc = 2 ::: Flag for SHAKE to perform bond length constraints, = 1 SHAKE is not performed (default), = 2 bonds involving hydrogen are constrained, = 3 all bonds are constrained (not available for parallel or qmmm runs in *sander*)

ntf = 2 ::: Force evaluation. Note: If SHAKE is used (see NTC), it is not necessary to calculate forces for the constrained bonds., = 1 complete interaction is calculated (default), = 2 bond interactions involving H-atoms omitted (use with NTC=2), = 3 all the bond interactions are omitted (use with NTC=3), = 4 angle involving H-atoms and all bonds are omitted, = 5 all bond and angle interactions are omitted, = 6 dihedrals involving H-atoms and all bonds and all angle interactions, are omitted, = 7 all bond, angle and dihedral interactions are omitted, = 8 all bond, angle, dihedral and non-bonded interactions are omitted

tempi = 300 and temp0 = 300 ::: Initial temperature and reference temperature at which the system is to be kept.

ntt = 3 ::: Switch for temperature scaling, = 0 Constant total energy classical dynamics (assuming that *ntb*<2, as should probably always be the case when *ntt=0*), = 1 Constant temperature, using the weak-coupling algorithm, = 2 Andersen temperature coupling scheme, = 3 Use Langevin dynamics with the collision frequency γ given by *gamma_ln*

gamma_ln = 5.0 ::: The collision frequency γ, in ps-1, when *ntt* = 3.

nstlim = 500000 ::: Number of MD-steps to be performed.

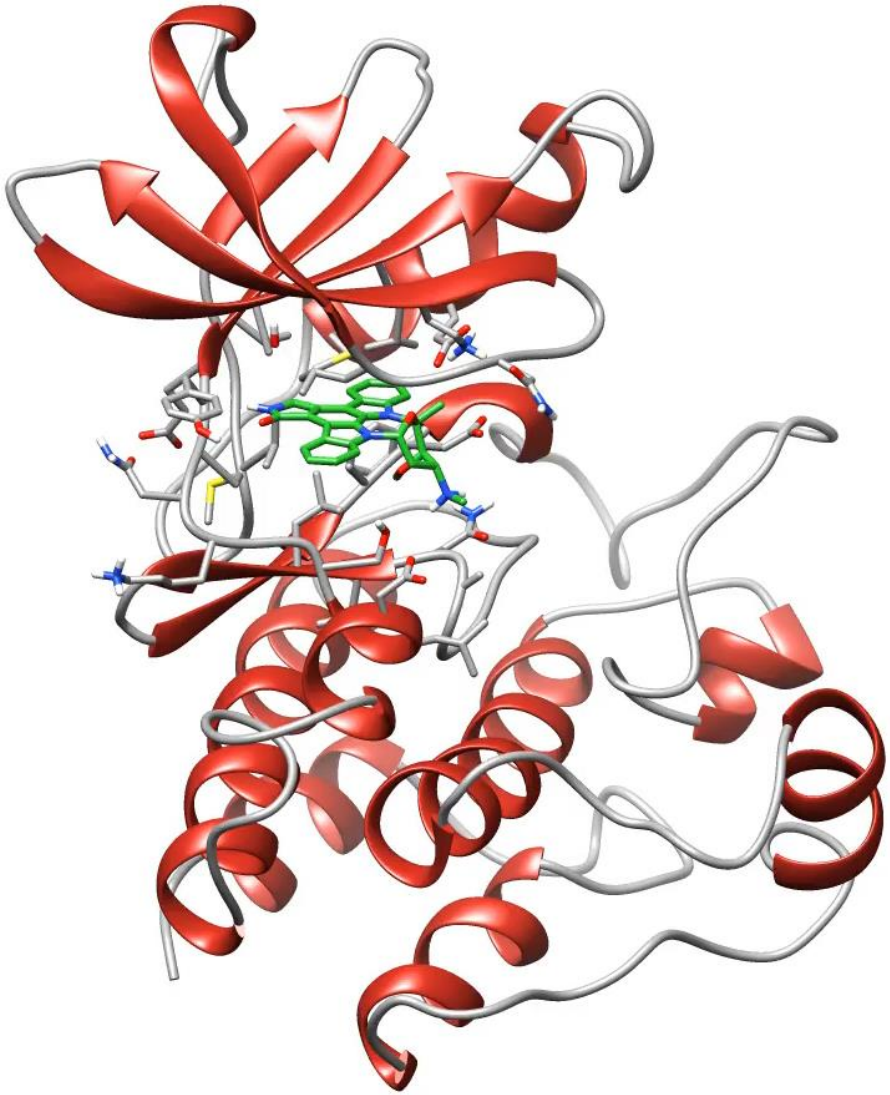
dt = 0.002 ::: The time step (psec).

ntpr = 100 ::: Every NTPR steps energy information will be printed in human-readable form to files "mdout" and "mdinfo

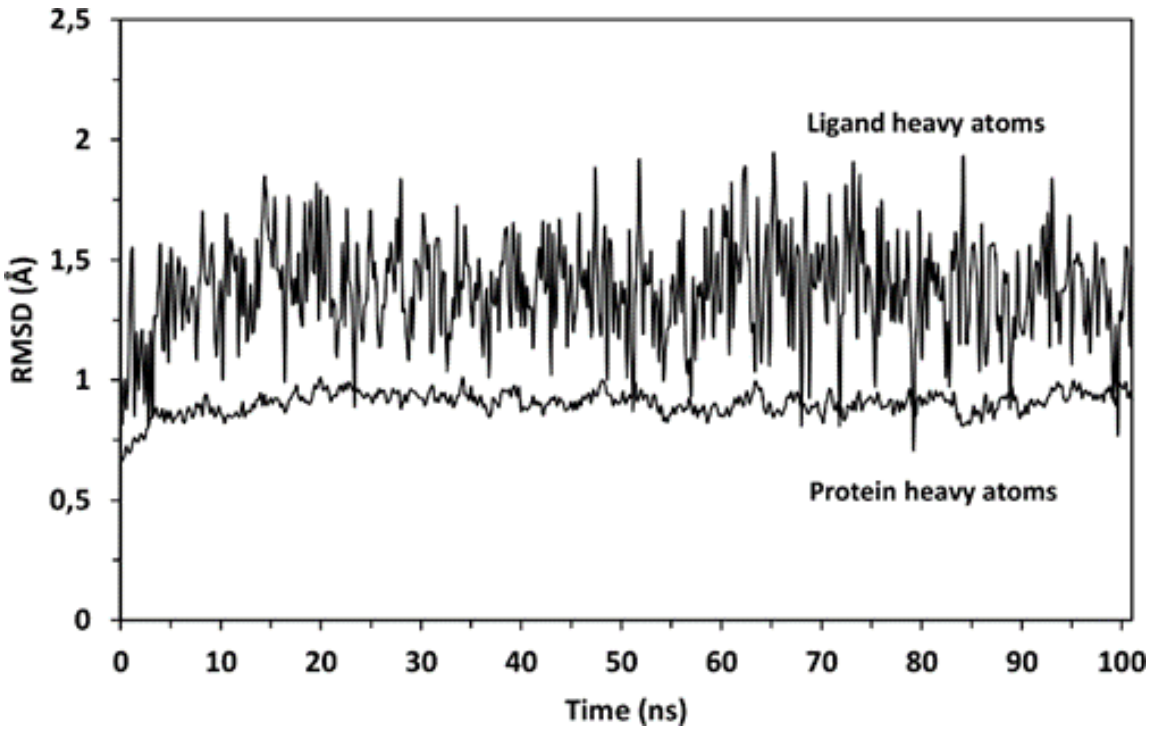
ntwx = 1000 ::: Every NTWX steps the coordinates will be written to file "mdcrd".

ntwr = 1000 ::: Every NTWR steps during dynamics, the "restrt" file will be written.

Trajectory Analysis

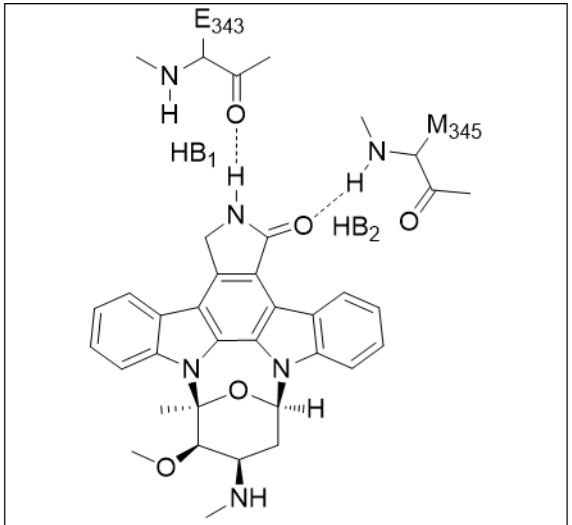


Trajectory Analysis



The MD plot shows the RMSD during the simulation of the heavy atoms of the protein and the heavy atoms of the ligand with respect to their initial coordinates.

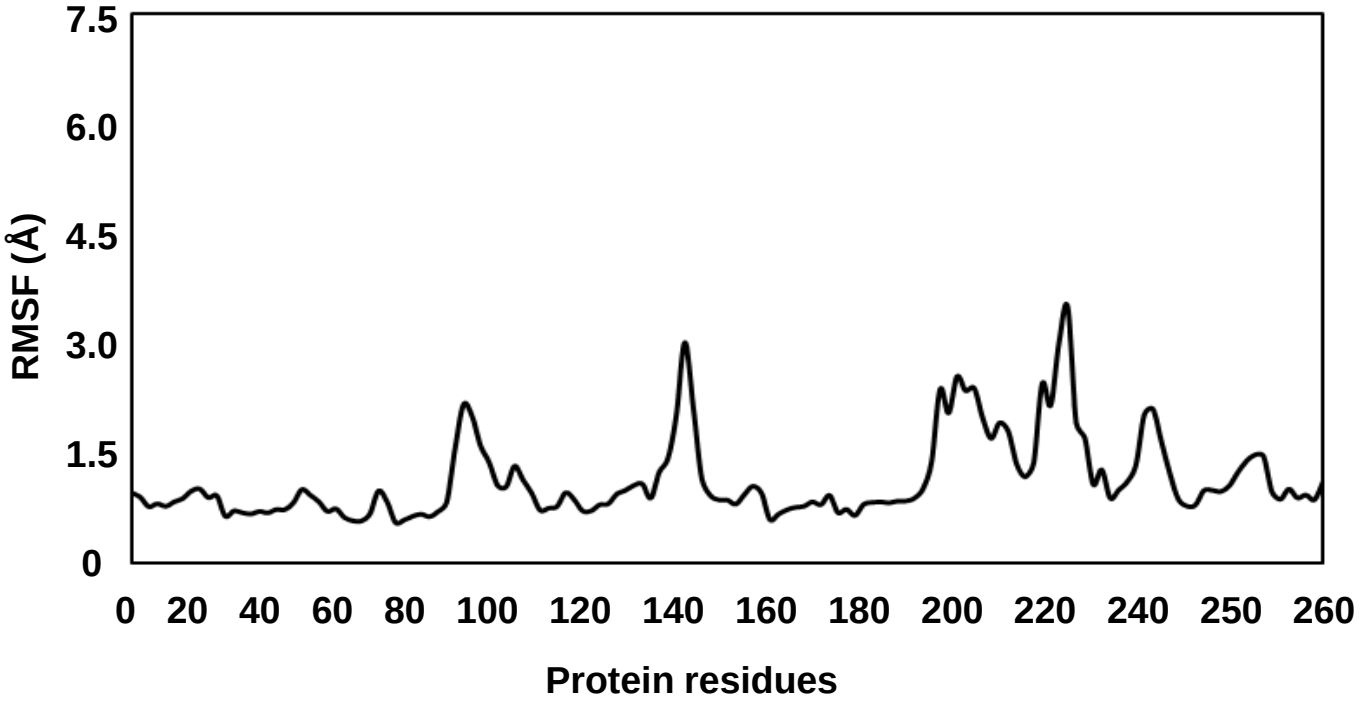
Trajectory Analysis



H-bond	Distance (Å)	% occupied
HB1	2.8	100.0
HB2	2.8	95.5

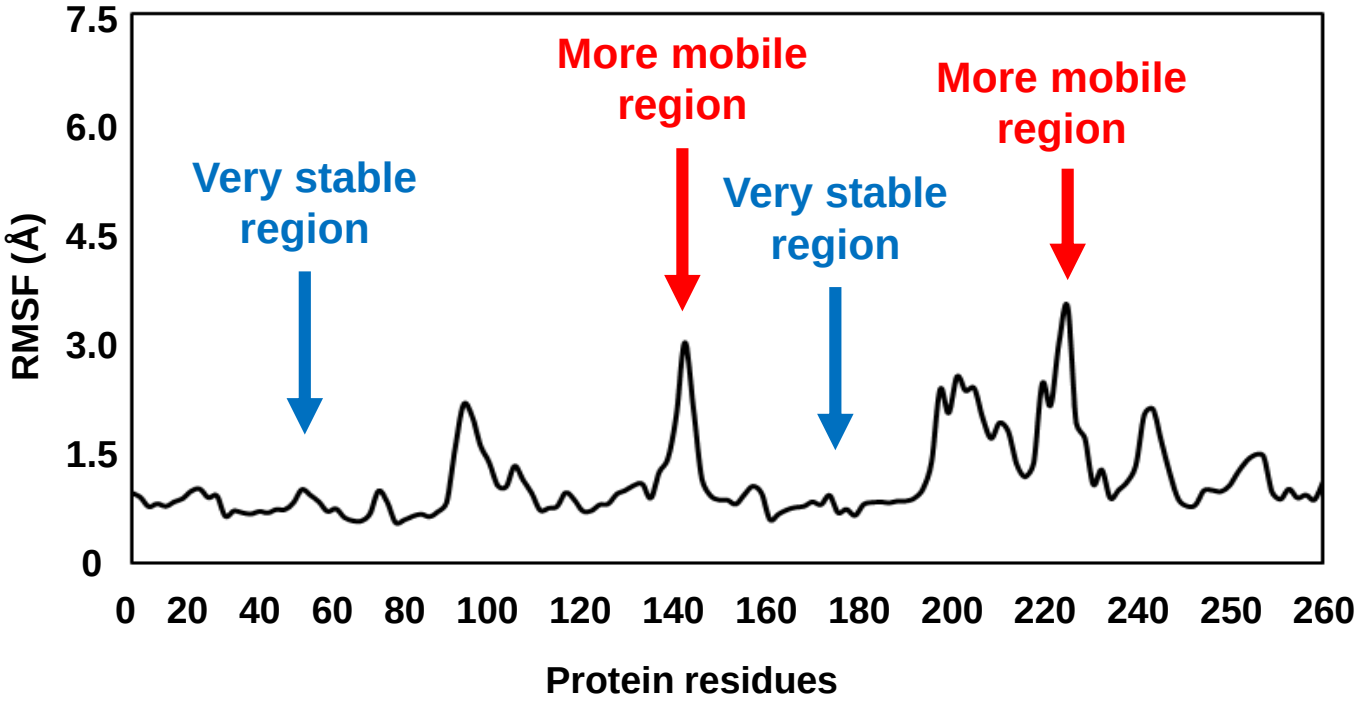
H-Bond analysis during 100 ns MD simulation of the staurosporine-Fyn complex

Trajectory Analysis



The plot shows the root-mean-square fluctuations (RMSF) calculated for the Cα of each protein residue with respect to the initial coordinates. It basically shows the average displacement of the Cα of each protein residue during the simulation.

Trajectory Analysis

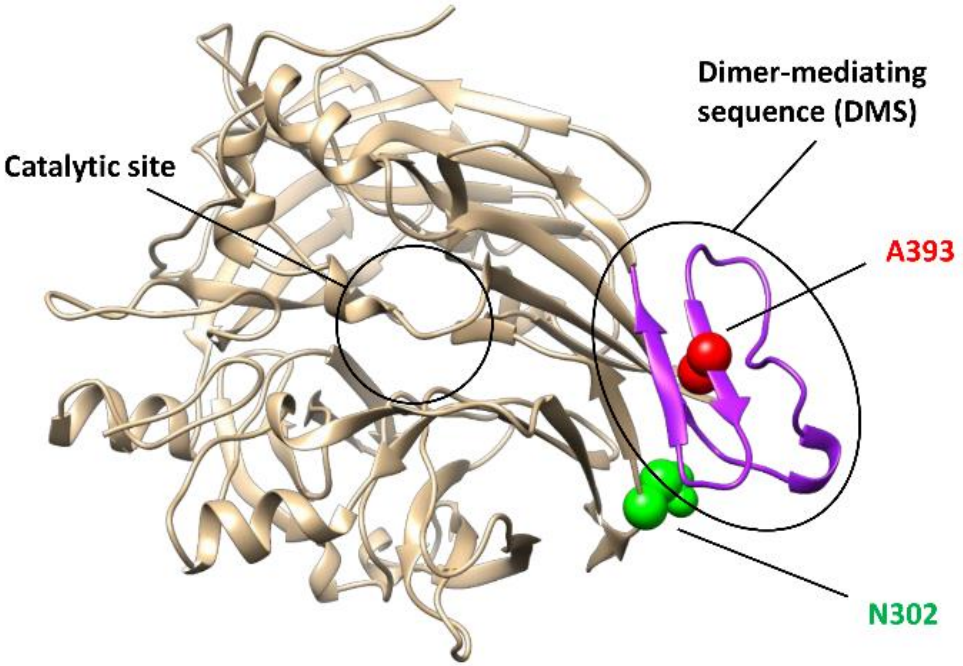


The plot shows the root-mean-square fluctuations (RMSF) calculated for the Cα of each protein residue with respect to the initial coordinates. It basically shows the average displacement of the Cα of each protein residue during the simulation.

Advanced MD studies: Evaluating the impact of protein mutations

Protocol development

The human retinal pigment epithelium-specific 65-kDa protein (*h*RPE65) is an essential enzyme of the visual cycle. Several missense mutations of *h*RPE65 are associated with inherited retinal diseases. MD simulations can be used to evaluate the impact of missense mutations on protein folding and conformation stability.

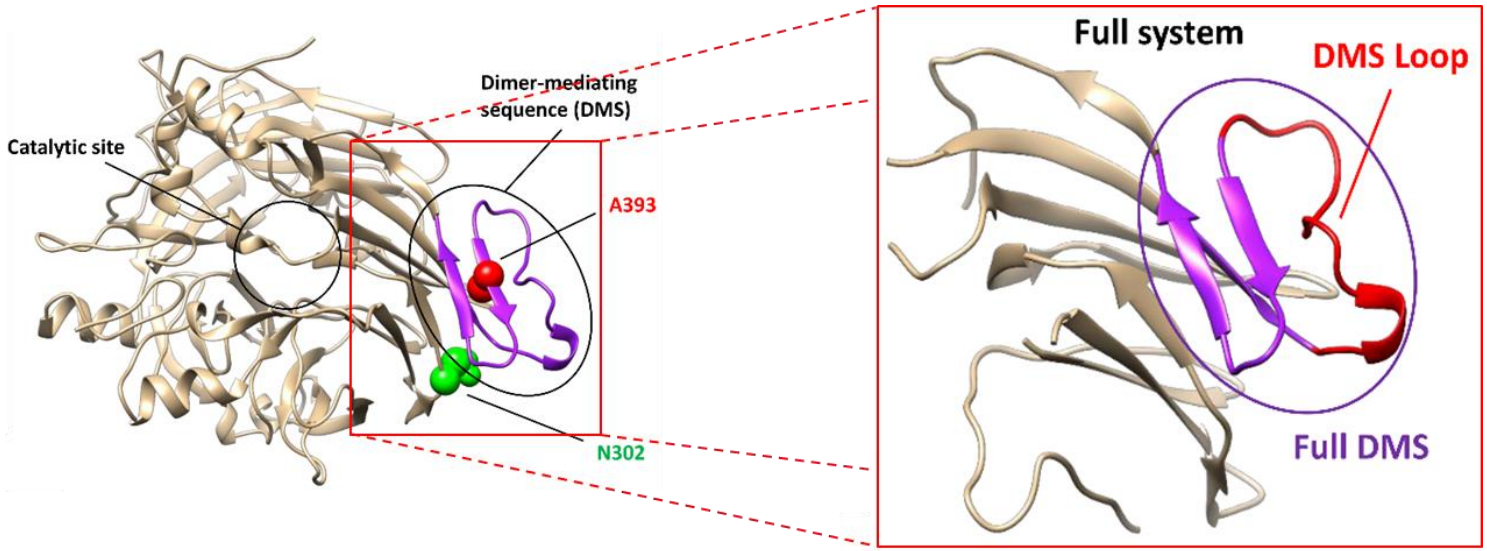


Residue associated to a reportedly pathogenic mutation (A393E)

Residue associated to a non-pathogenic mutation (N302I)

Protocol development

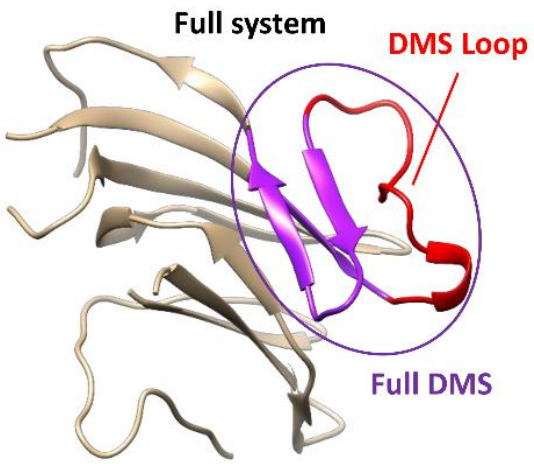
Reducing the dimensions of the system is a trick to increase the speed of MD simulations, which often scales linearly with the number of atoms within the system. Position restrains can be applied to anchor the terminal sequences of the "cut" protein and make the system stable. For small systems, long simulation times become more affordable and allow increasing the chance of observing long-range conformational motions. Additional tricks can be applied for achieving a further speed up.



Poli et al. *J Enzyme Inhib Med Chem*. 2022; *37(1)*: 1765–1772.

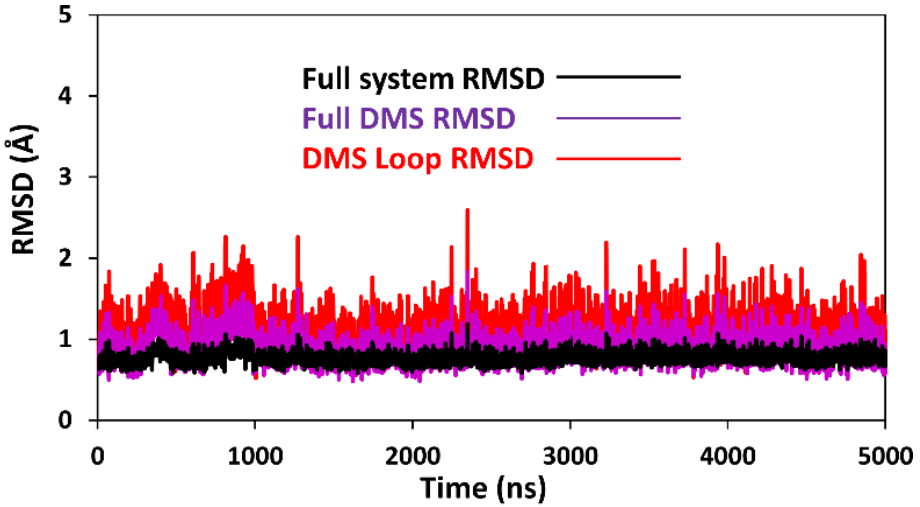
Protocol evaluation

The first step to be performed is a protocol evaluation. The preliminary assessment of the approach is necessary for ensuring its reliability. It also provides reference results to which those obtained through the subsequent analyses can be compared.



Stability analysis

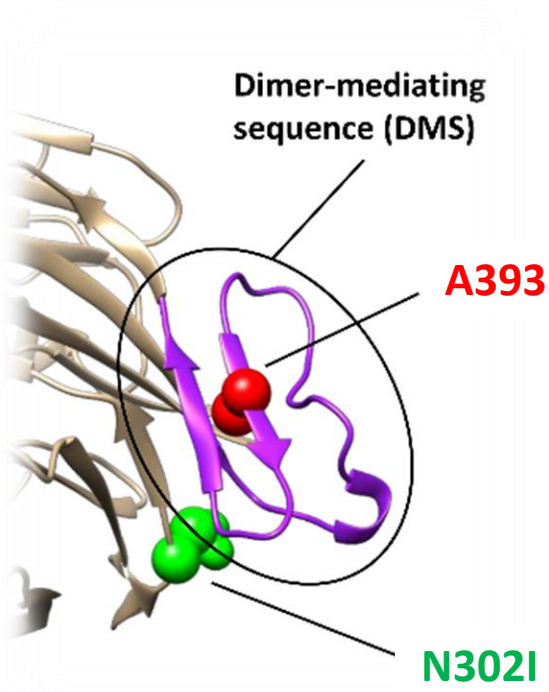
over a 5 µs MD



Poli et al. *J Enzyme Inhib Med Chem*. 2022; *37(1)*: 1765–1772.

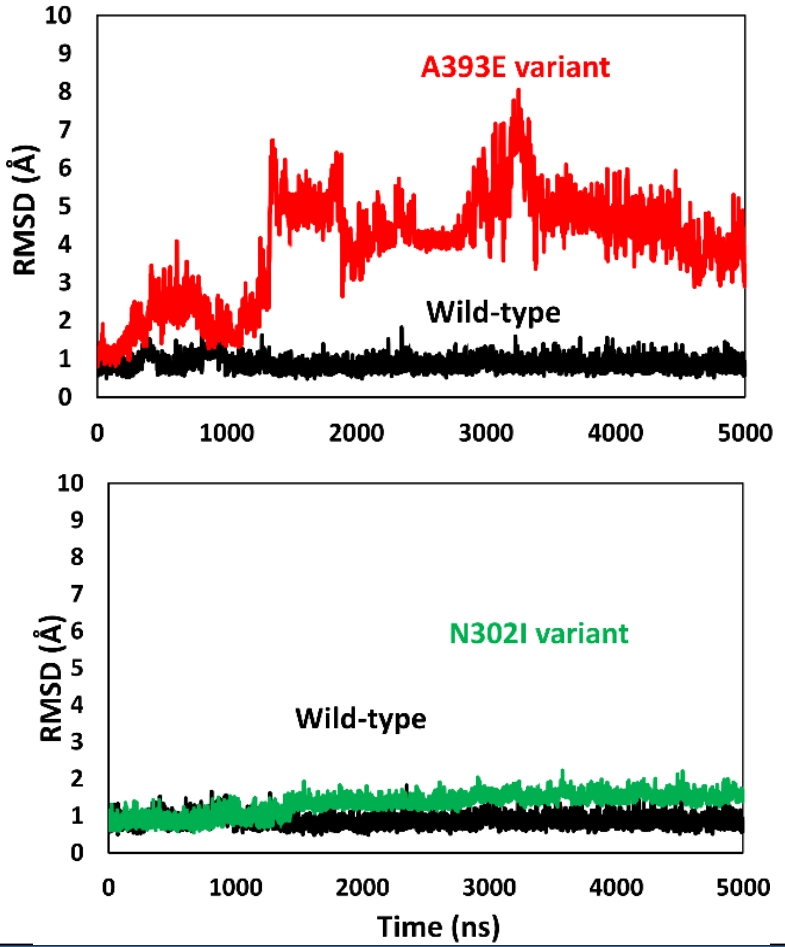
Evaluating the impact of mutations

Once the reliability of the protocol is confirmed, it can be applied for studying the target variants and analyzing the impact on local folding.



Stability analysis over a 5 µs MD

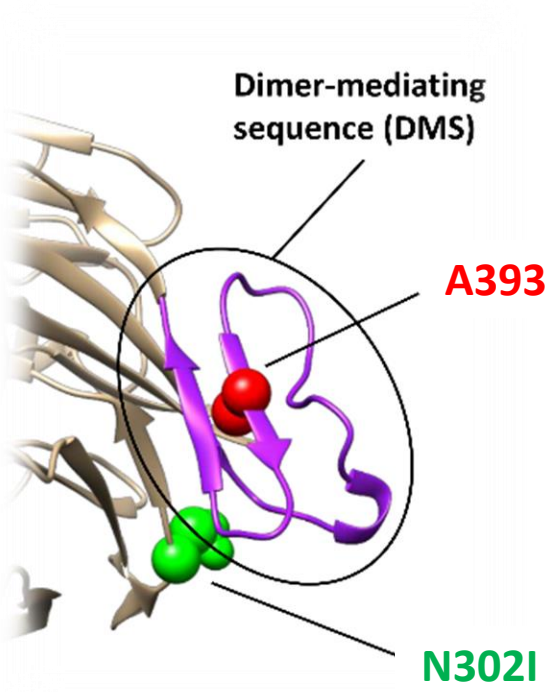
Stability analysis over a 5 µs MD



Poli et al. *J Enzyme Inhib Med Chem*. 2022; *37(1)*: 1765–1772.

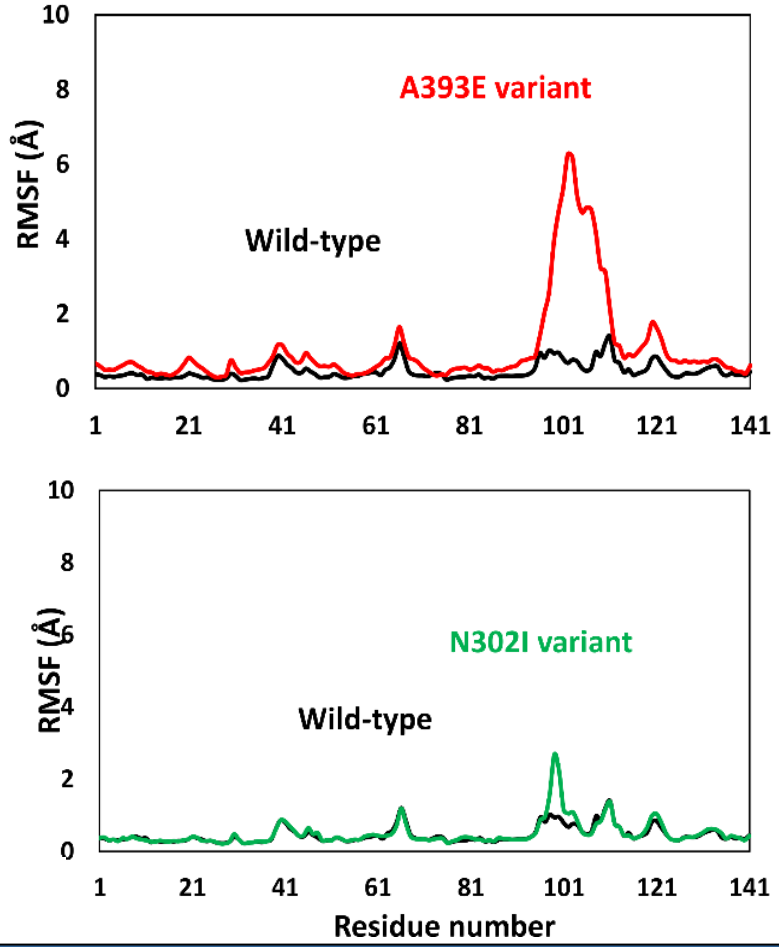
Evaluating the impact of mutations

Once the reliability of the protocol is confirmed, it can be applied for studying the target variants and analyzing the impact on local folding.



Stability analysis
over a 5 µs MD

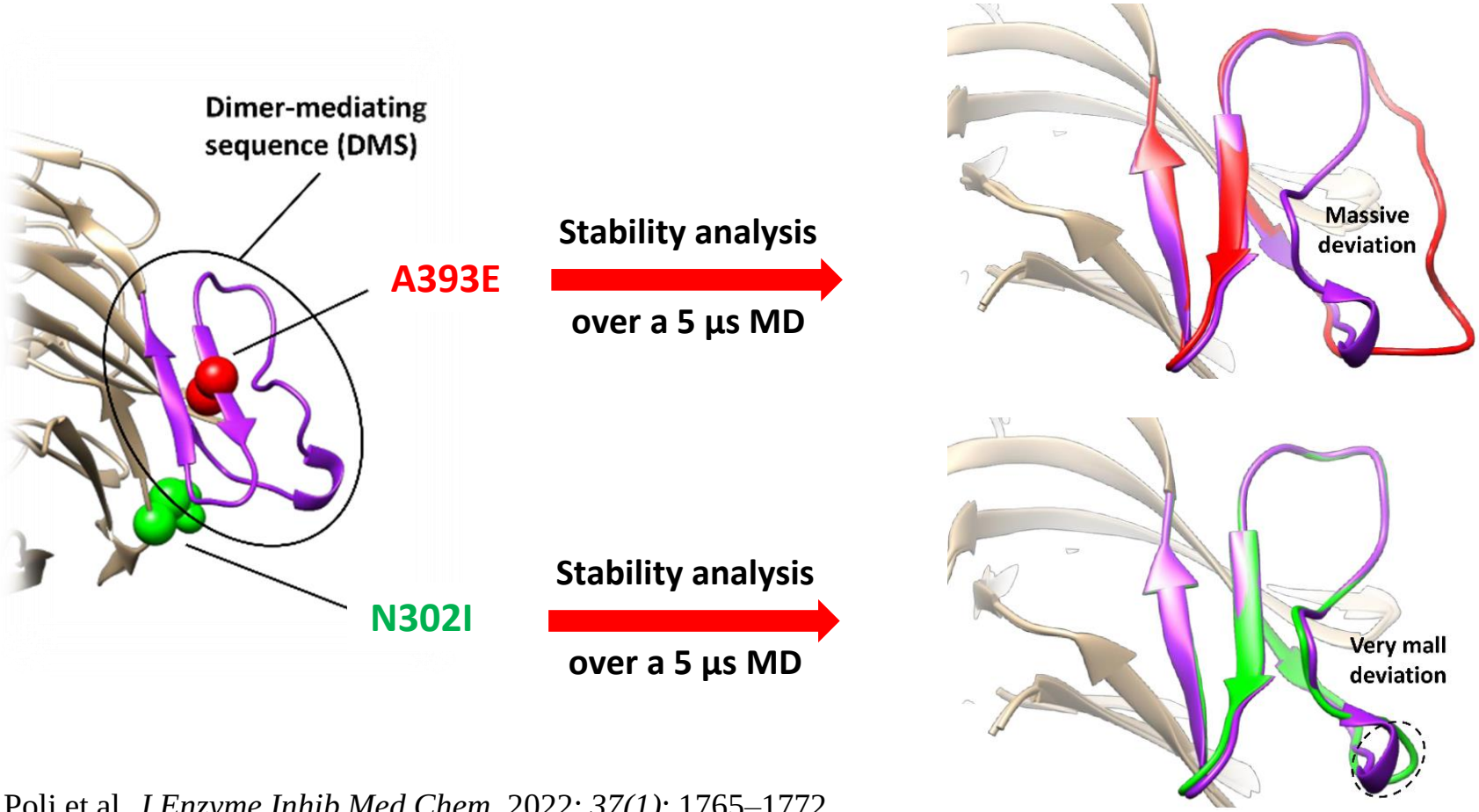
Stability analysis
over a 5 µs MD



Poli et al. *J Enzyme Inhib Med Chem*. 2022; *37(1)*: 1765–1772.

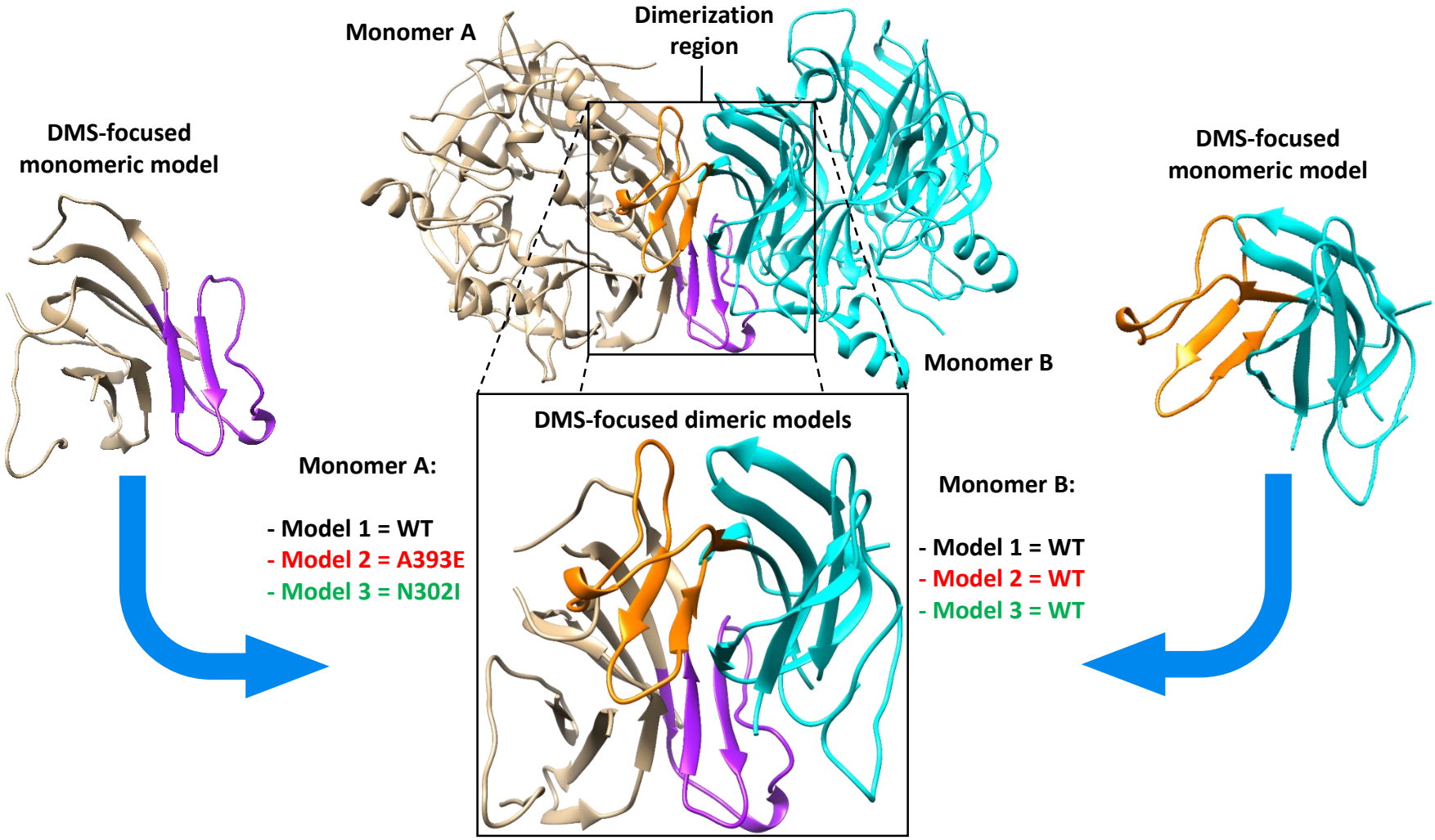
Evaluating the impact of mutations

Once the reliability of the protocol is confirmed, it can be applied for studying the target variants and analyzing the impact on local folding.



Poli et al. *J Enzyme Inhib Med Chem*. 2022; *37(1)*: 1765–1772.

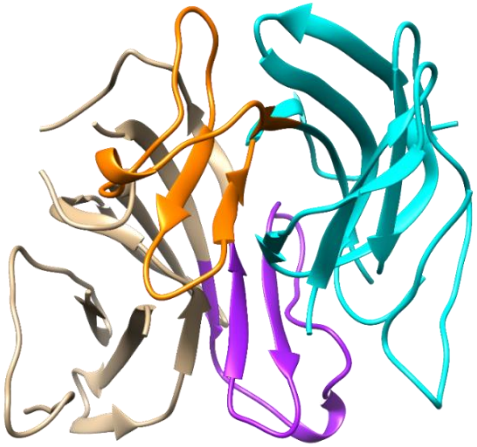
Protocol optimization



Poli et al. *J Enzyme Inhib Med Chem*. 2022; *37(1)*: 1765–1772.

Evaluating the impact of mutations

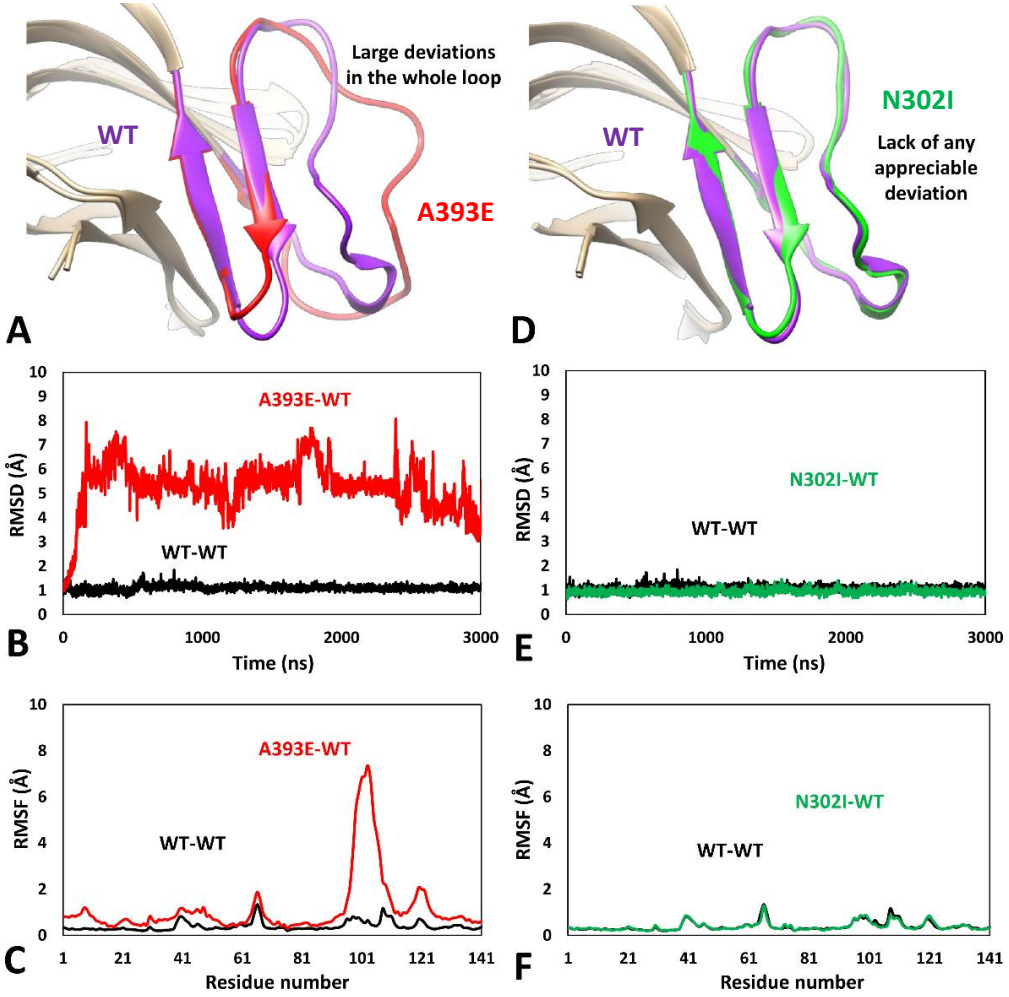
Repeating the MD simulations and analyses to evaluate the benefits of the optimized *in silico* protocol



DMS-focused dimeric models

Stability analysis

over a 3 µs MD



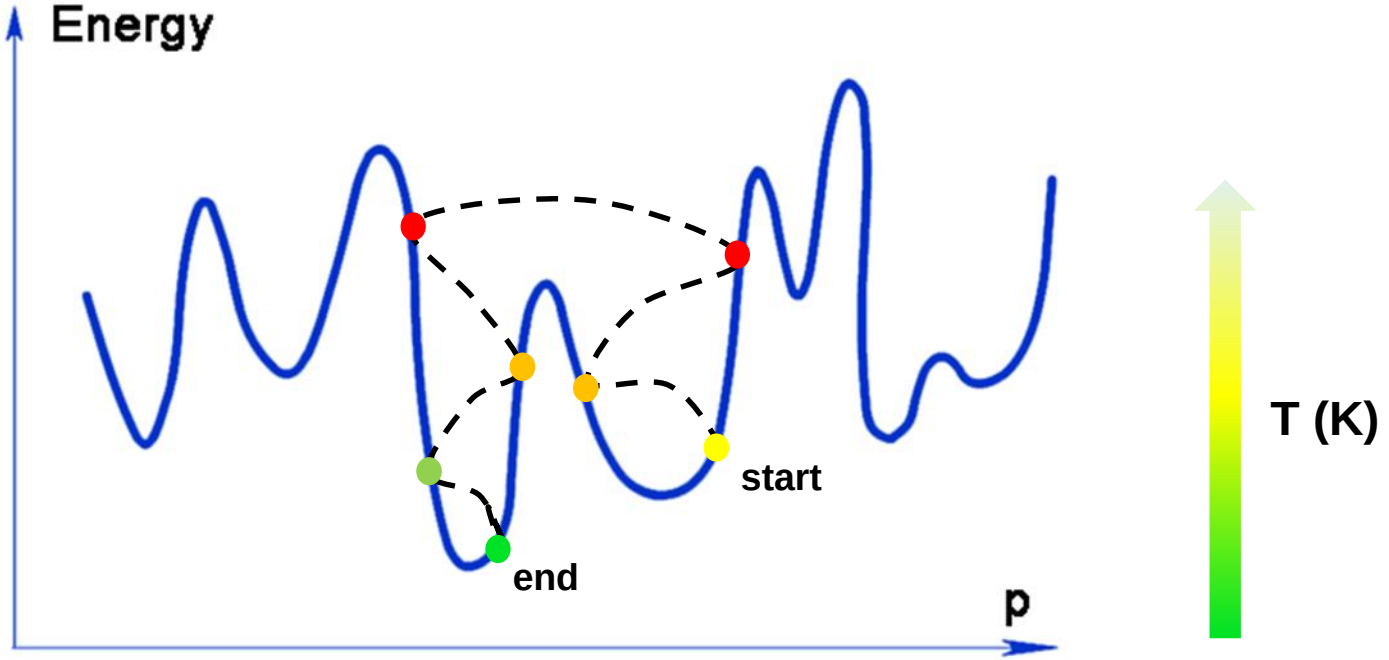
Poli et al. *J Enzyme Inhib Med Chem*. 2022; *37(1)*: 1765–1772.

MD algorithms

- Classic MD simulation
- Simulated annealing
- Coarse-grained MD
- Steered MD

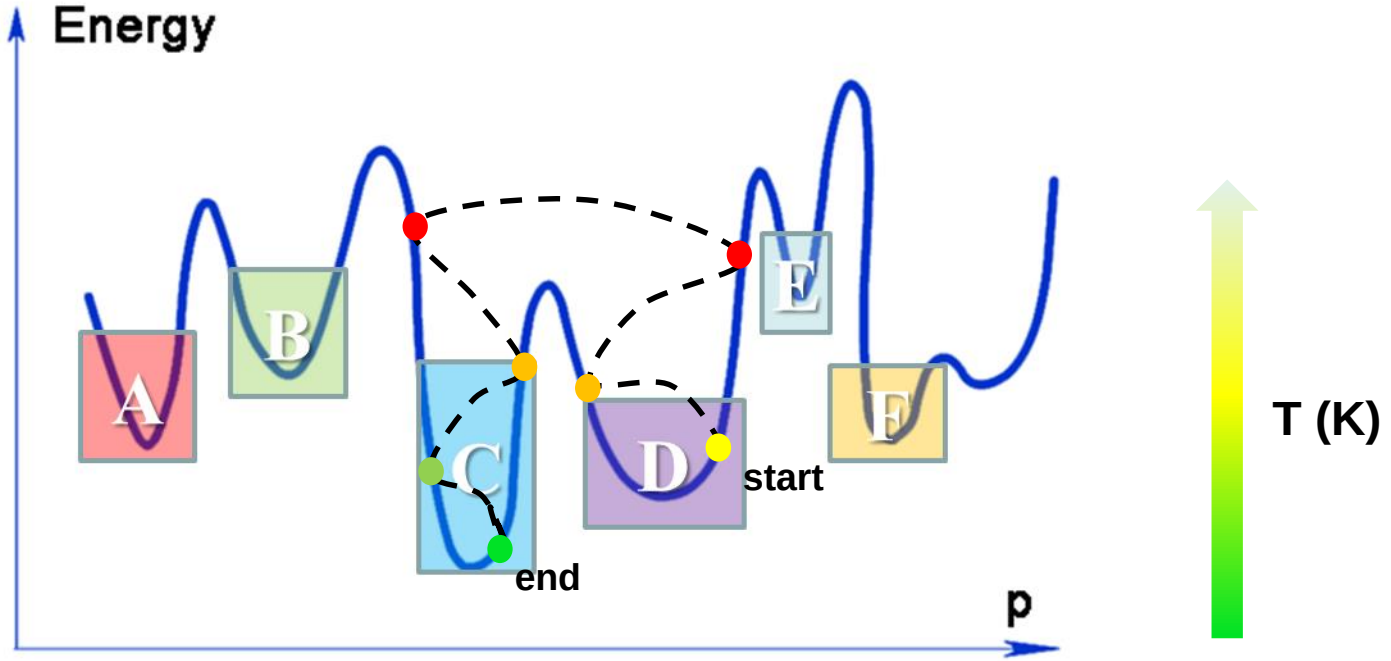
Simulated Annealing

Simulated annealing (SA) is a special type of dynamics where the system is initially heated at high temperature (usually 2000-3000 K). An MD run is then initiated, during which the temperature is slowly reduced. Initially the molecule is allowed to move over a large area, but as the temperature decreases, it ends up being trapped in a minimum.



Simulated Annealing

This process is generally repeated many times until several low energy conformations are obtained. The name, simulated annealing, comes from the analogy of growing crystals or hardening steel. **It is commonly used for obtaining a thorough conformational analysis of large systems.**

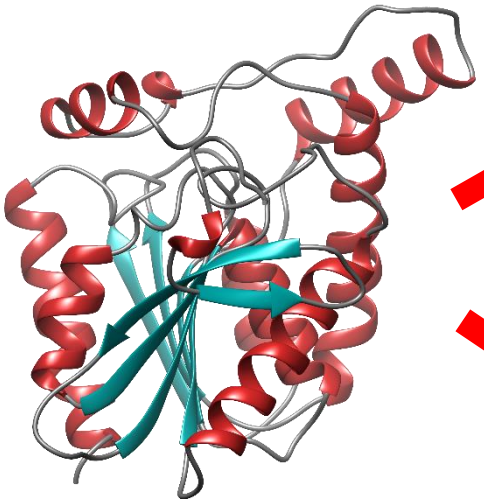


Coarse-grained MD

Generally, classic MD approaches are limited to simulation times of hundreds of nanoseconds and to systems including a few hundred thousand atoms. A possible way to extend molecular modeling and bridge it with experimental techniques is to use coarse-graining, allowing **to represent a system by a reduced in number of degrees of freedom with respect to an all-atom description.**

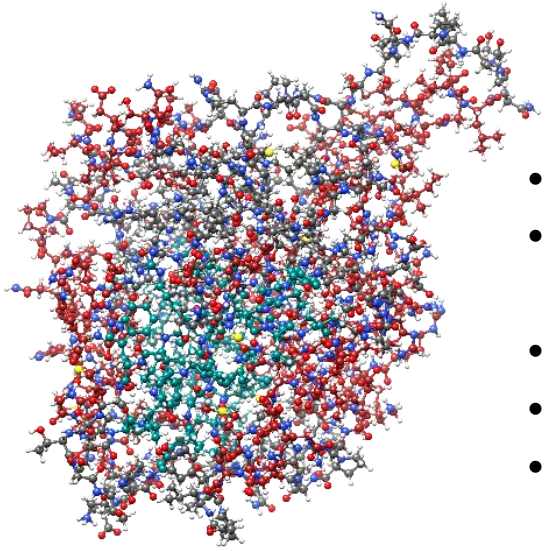
Due to the reduction in the degrees of freedom and elimination of fine interaction details, the simulation of a coarse-grained (CG) system requires less resources and goes faster than the MD of the same system in all-atom representation. As a result, an increase of orders of magnitude in the simulated time and length scales can be achieved. Instead of explicitly representing every atom of the system, CG uses "pseudo-atoms“ (beads) to represent groups of atoms. **Coarse-grained calculations are often used for analyzing protein folding.**

Coarse-grained MD



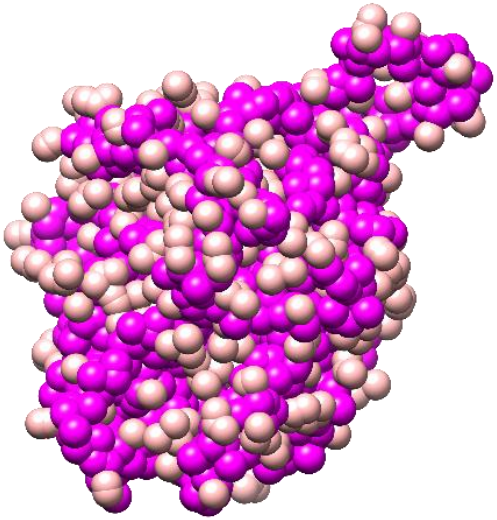
Protein structure of 300 residues used as input for a solvated system (15 Å cubic water box)

All-atom system



- 4500 protein atoms
- 55000 solvent atoms (15 Å cubic water box)
- 59500 total atoms
- dt = 0.002 ps
- MD speed = x ns/day

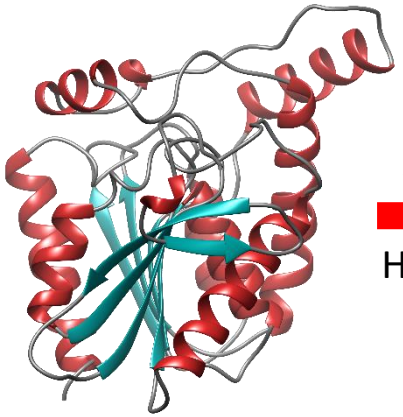
Coarse-grained system



- 1370 protein beads
- 6400 solvent beads (15 Å cubic water box)
- 7770 total beads
- dt = **0.020** ps
- MD speed = **17x** ns/day

Coarse-grained MD

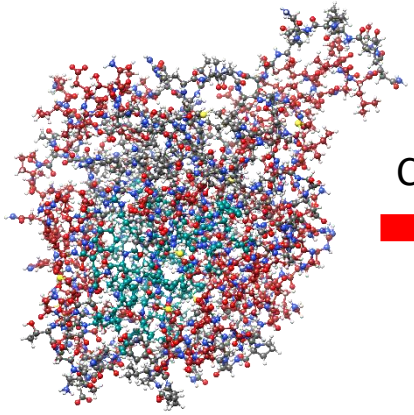
Target protein



X-ray structure

Homology model

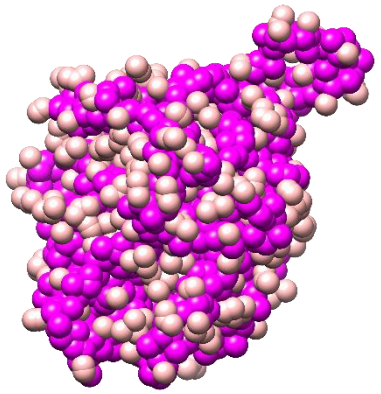
All-atom structure



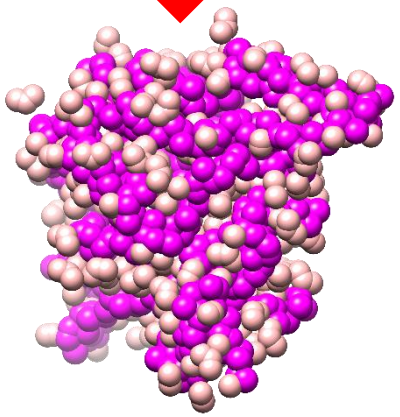
Coarse-grained

mapping

CG system for MD

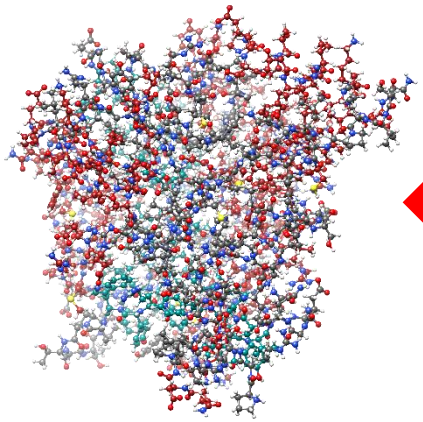


MD run



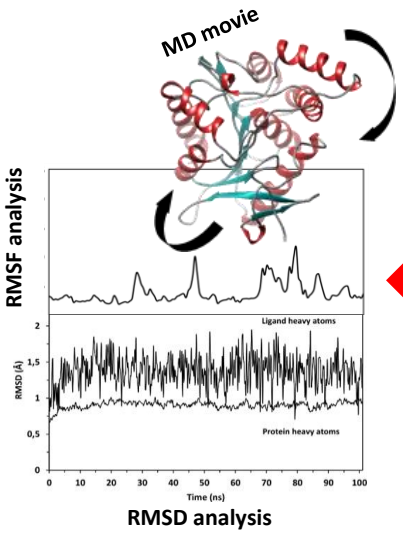
backmapping

to all-atom



MD

analysis



Steered MD

Knowledge of the mechanism of association and dissociation of macromolecules is important for many biological structures and processes.

Among possible examples are the binding and dissociation of substrates of enzyme reactions, the recognition of ligands by their receptors, the recognition of DNA sequences by their regulatory protein binding domains.

These processes have in common a transition from one equilibrium state to another, which is often a rare event on the time scale of molecular dynamics simulations of a few hundred nanoseconds.

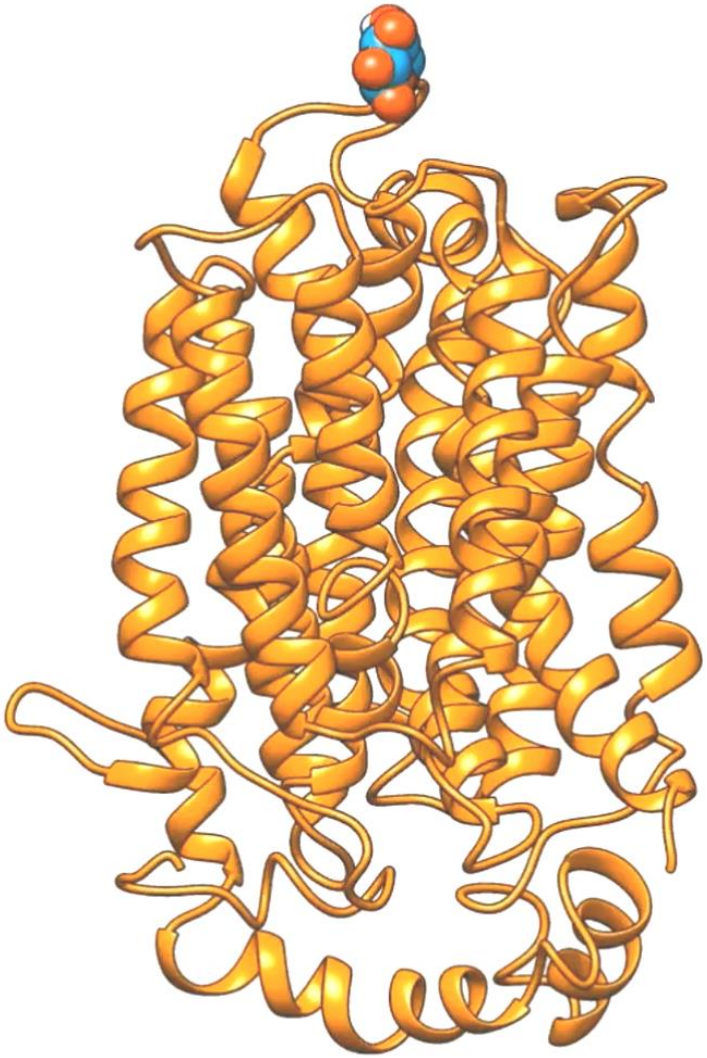
Steered MD

A methodologically related avenue to characterize rare events through molecular dynamics simulation is the addition of external forces to overcome the energy barriers.

This approach has the advantage that it closely corresponds to micromanipulation through atomic force microscopy or optical tweezers.

The external force techniques can be applied to study many processes, including dissociation of protein-protein complexes, dissociation of ligands from their target, stretching of DNA, etc.

Steered MD



Steered MD