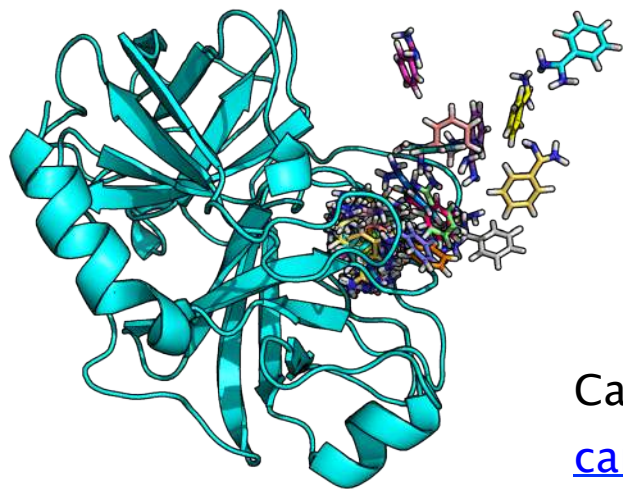


MOLECULAR MECHANICS AND MOLECULAR DYNAMICS SIMULATIONS



“The long range goal of molecular approaches to **biology** is to describe living systems in terms of **chemistry** and **physics**”
Dirac. *Proc. Roy. Soc.* **1929**, 123, 714

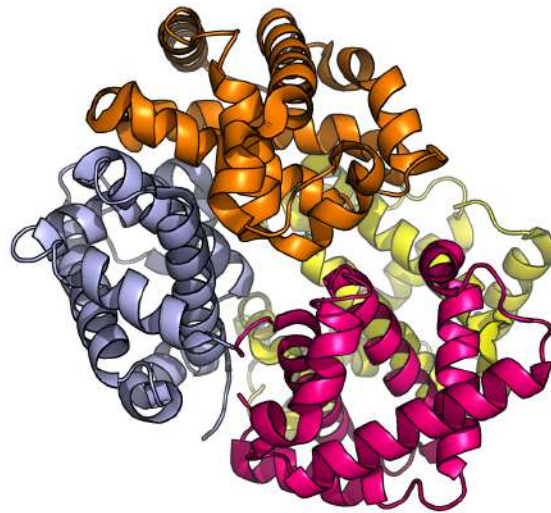
Carla Calvó-Tusell
carla.calvoa@udg.edu

Ferran Feixas
ferran.feixas@udg.edu

Institut de Química Computacional i Catàlisi (IQCC)
Universitat de Girona, Catalonia, Spain

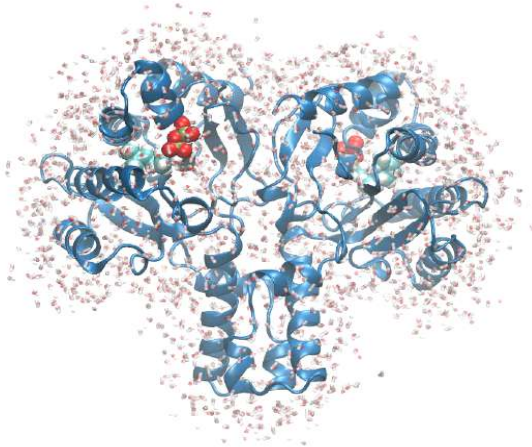
Topic 1. Fundamentals of Biomolecular Structure and Dynamics

- 1.1. The three-dimensional structure of biomolecules
- 1.2. Conformational dynamics and flexibility of biomolecules
- 1.3. Basic Concepts Molecular Modelling



Topic 2. Computational Techniques for Biomolecular Design.

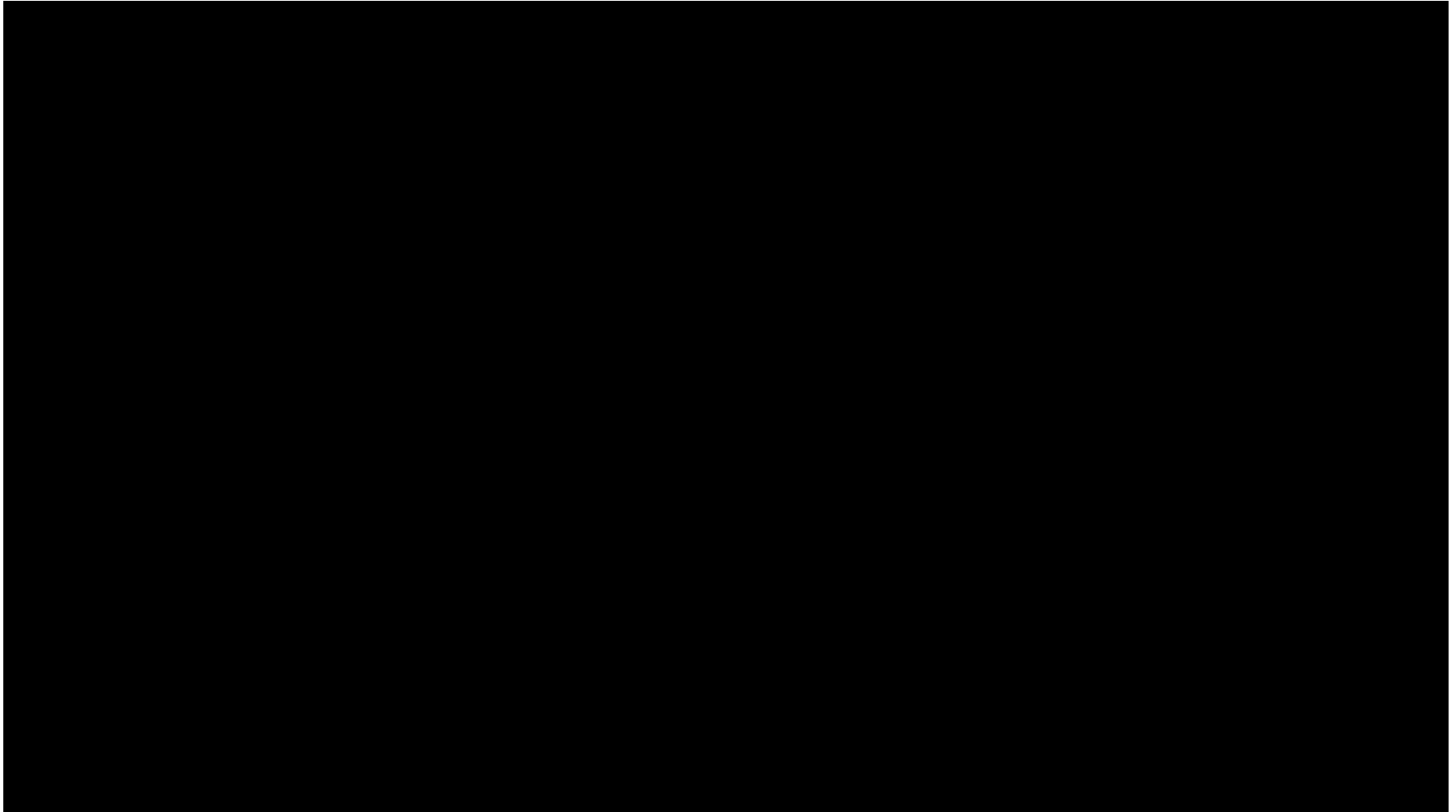
- 2.1. Molecular Mechanics: Force fields
- 2.2. Introduction to Molecular Dynamics
- 2.3. Enhanced-sampling techniques



$$\begin{aligned}
 U = & \sum_{i < j} \sum 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \\
 & + \sum_{i < j} \sum \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \\
 & + \sum_{\text{bonds}} \frac{1}{2} k_b (r - r_0)^2 \\
 & + \sum_{\text{angles}} \frac{1}{2} k_a (\theta - \theta_0)^2 \\
 & + \sum_{\text{torsions}} k_\phi [1 + \cos(n\phi - \delta)]
 \end{aligned}$$

Biomolecules

A biomolecule is any molecule that is present in living organisms



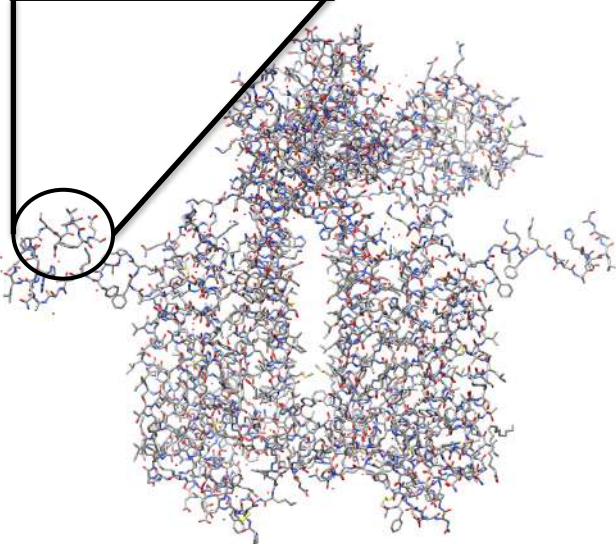
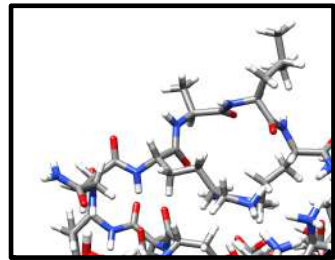
Schrödinger Movie: <https://www.youtube.com/watch?v=nh0jITjssnQ>

What is a protein

<http://pdb101.rcsb.org/learn/videos/what-is-a-protein-video>

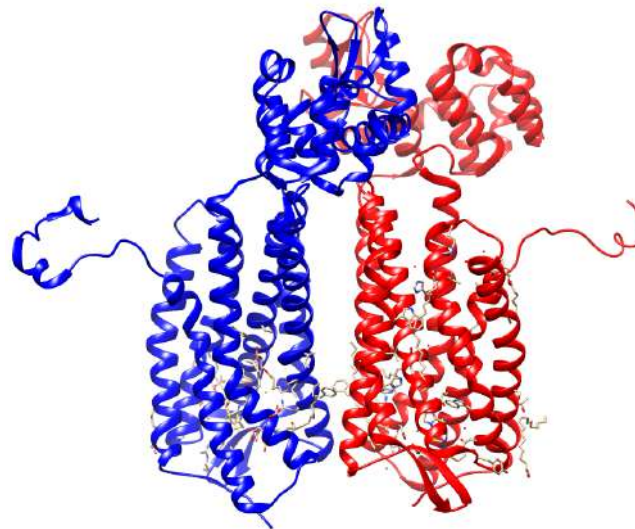
Concept 1. Protein

Proteins are central to **cellular function**.
Display a fascinating **variety of functions**.



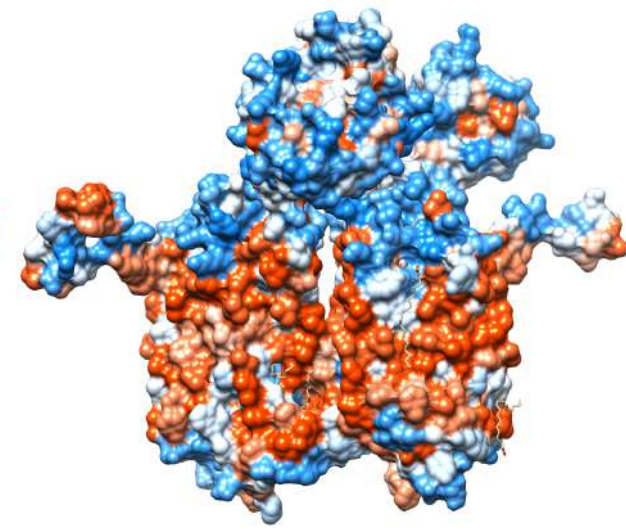
Chemistry

Enzymes and Receptors



Biology

GPCRs
Ion Channels
Nuclear Receptors

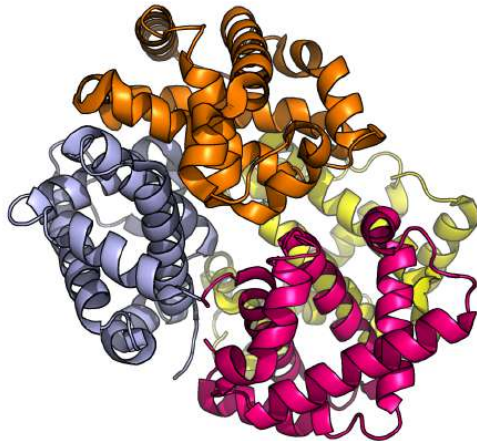


Physics

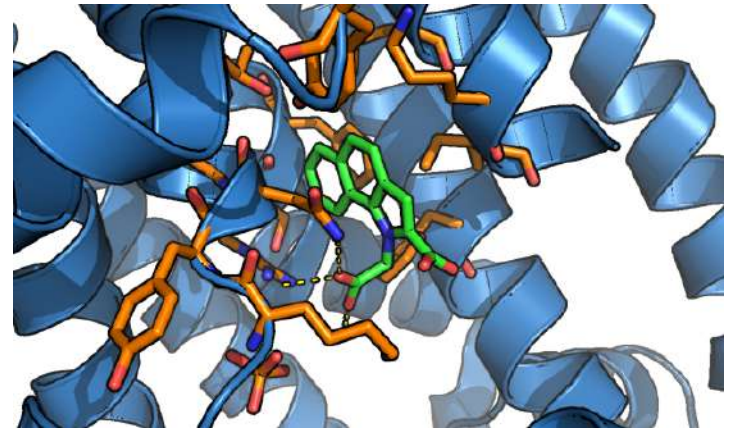
Concept 2. Protein Structure

Structural biology have provided **atomic-resolution models** of many of the molecules that are essential to life, including proteins and nucleic acids.

Structure-function relationship



Protein Structure

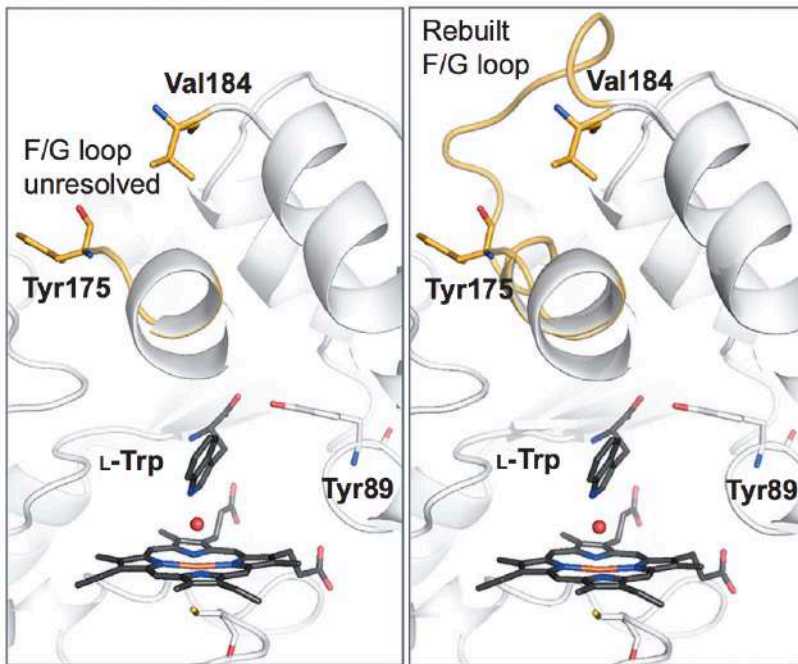


Structure Based-Drug
Design

Concept 2. Protein Structure

Enzyme Catalysis

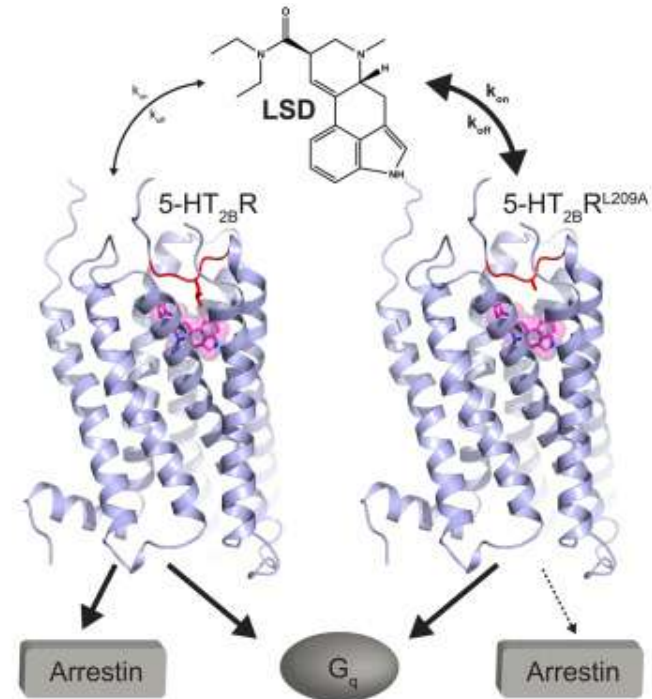
Orientation of catalytically important residues.



S. C. Dodani et al. *Nature Chemistry* 2016, 8, 419 ([article](#))

Structure-based Drug Design

Orientation of drug and biomolecular interactions ligand-receptor.

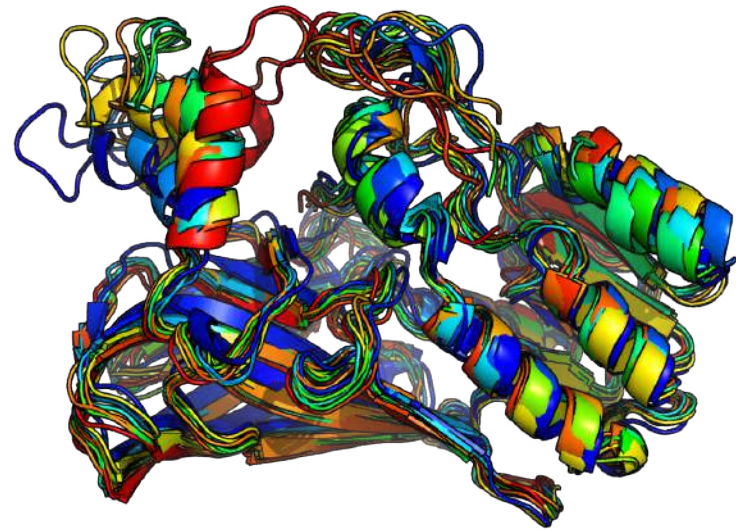


D. Wacker et al. *Cell* 2017, 168, 377 ([article](#))

Concept 3. Protein Dynamics

Proteins are **highly dynamic** and their motions are often critical to their function

Proteins sample a large **ensemble** of conformations



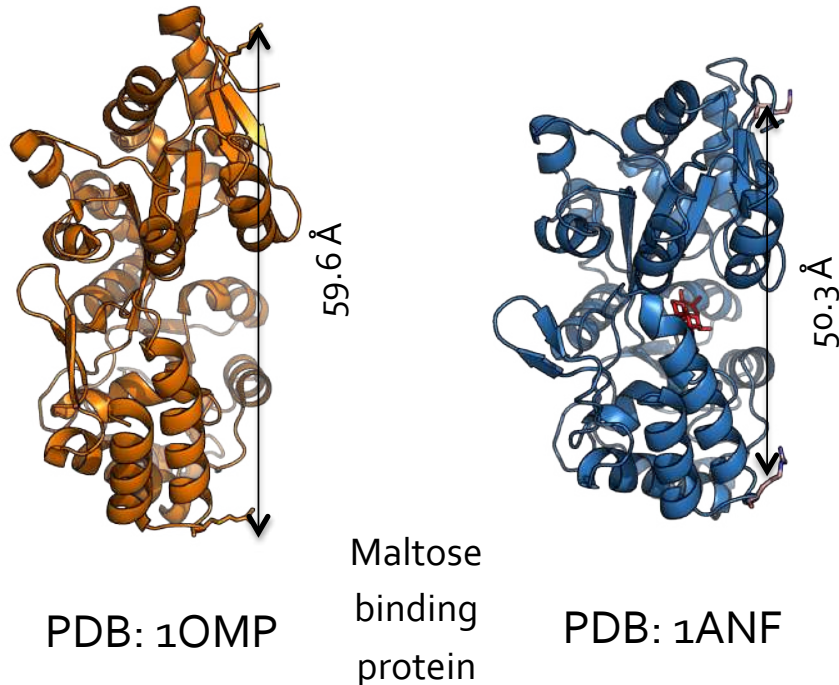
Extension of the **structure–function** paradigm to include **dynamics**

Concept 3. Protein Dynamics

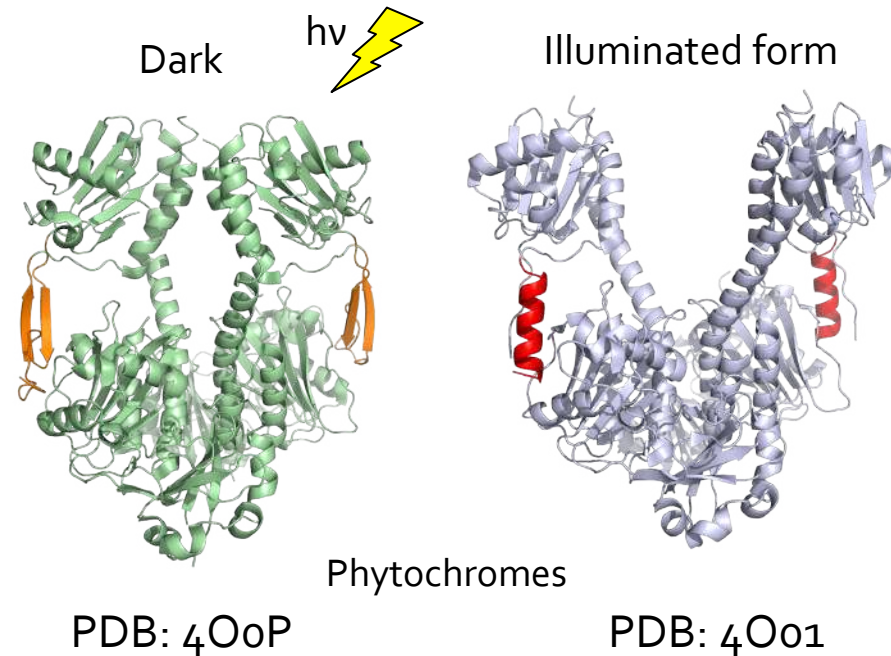
Same sequence, different protein **conformation**?



a) Ligands (substrate/drugs) can stabilize protein conformations

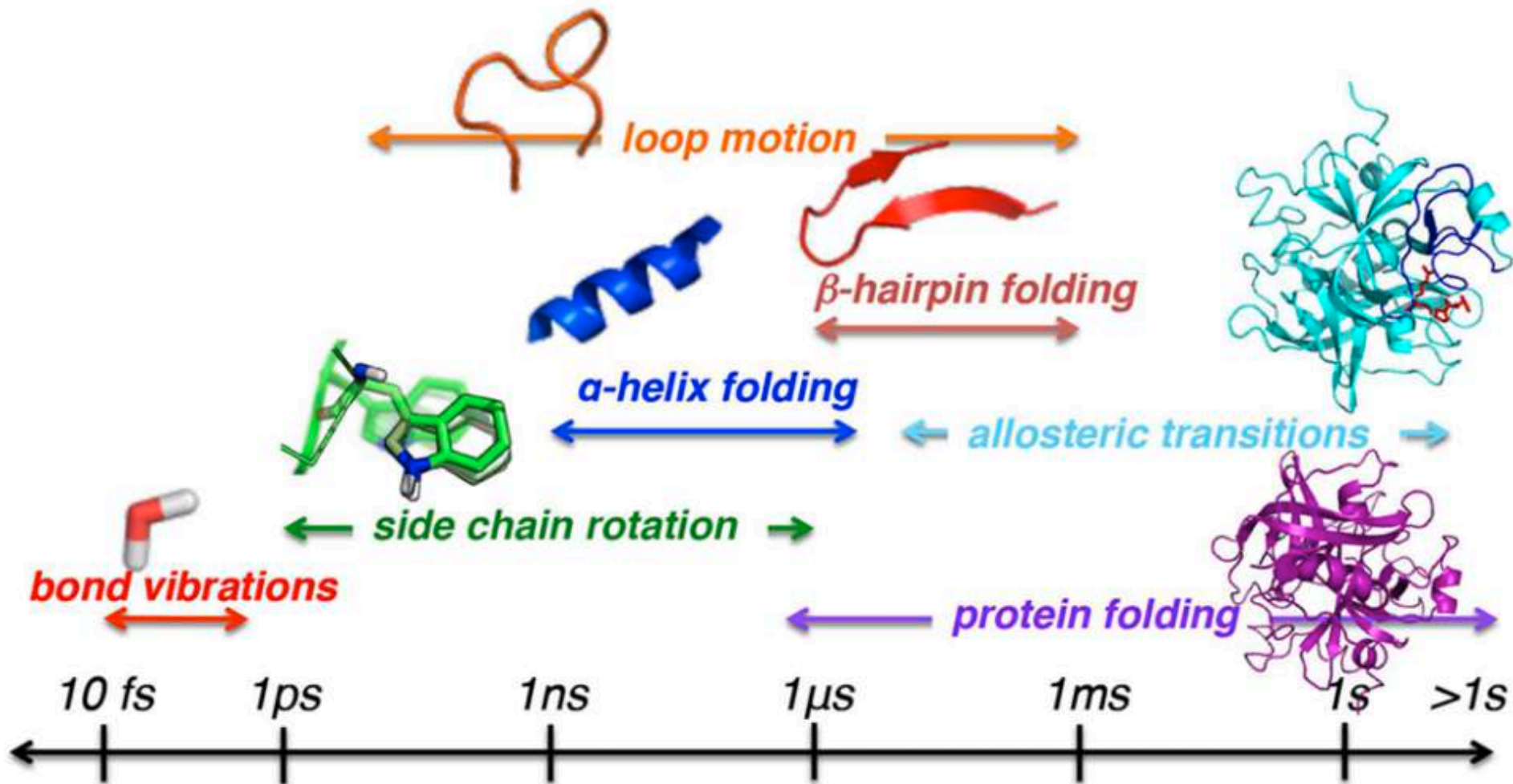


b) Light can control conformational changes



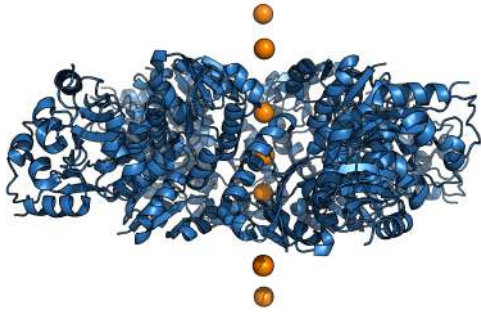
Concept 3. Protein Dynamics

Biological time scales of fundamental biomolecular processes

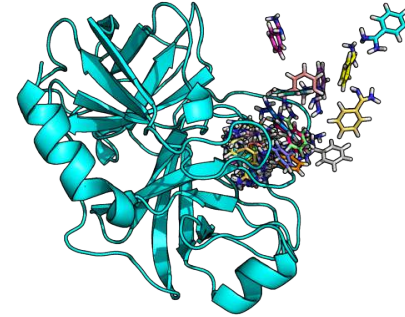


Concept 4. Protein Dynamics & Biomolecular Processes

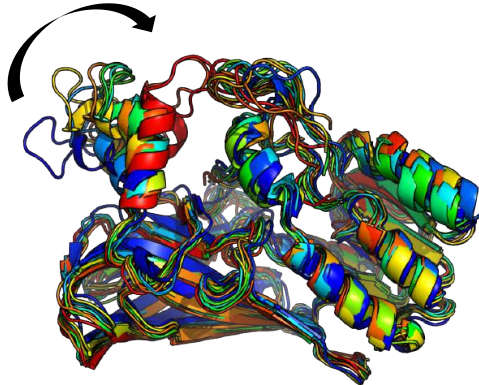
a) Transport Across Membrane



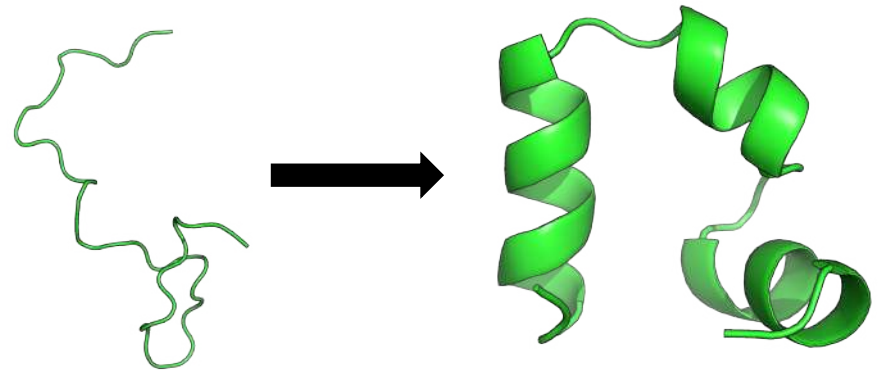
b) Ligand Binding



c) Conformational Changes



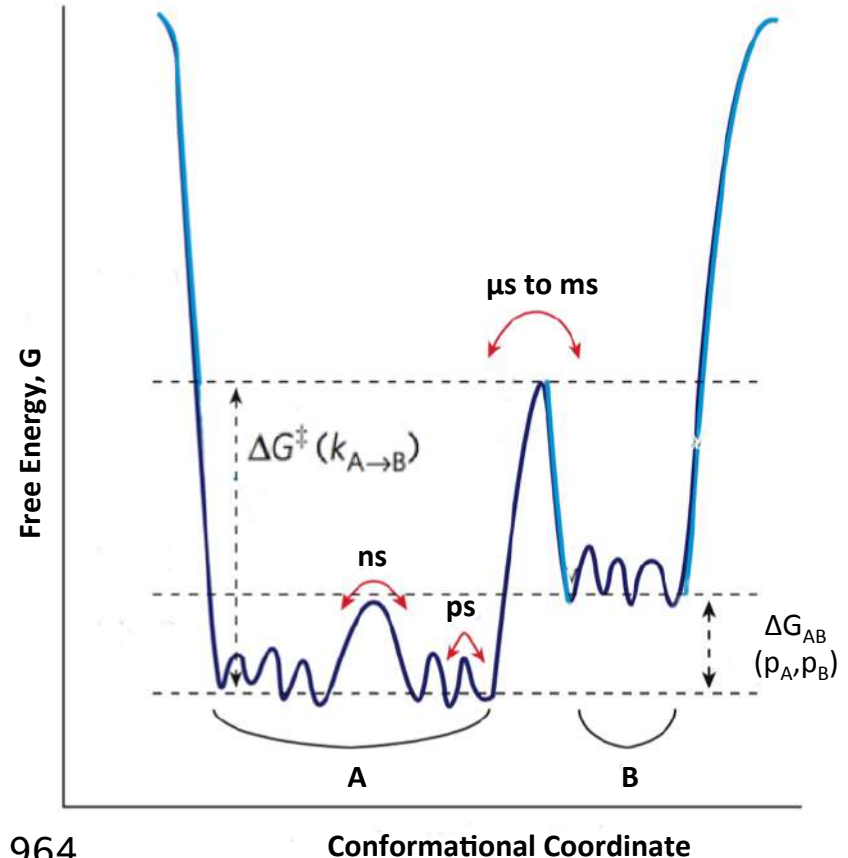
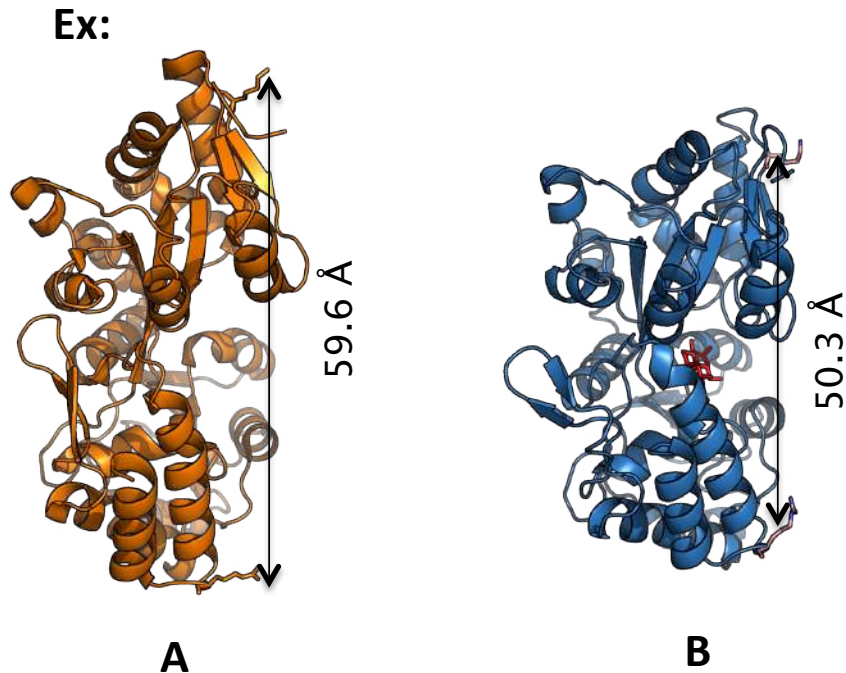
d) Protein Folding



Proteins undergo a wide variety of conformational changes that allow them to act as **transporters**, **catalysts**, **signaling molecules**. They interact dynamically with **hormones**, **drugs**, and **with other proteins**.

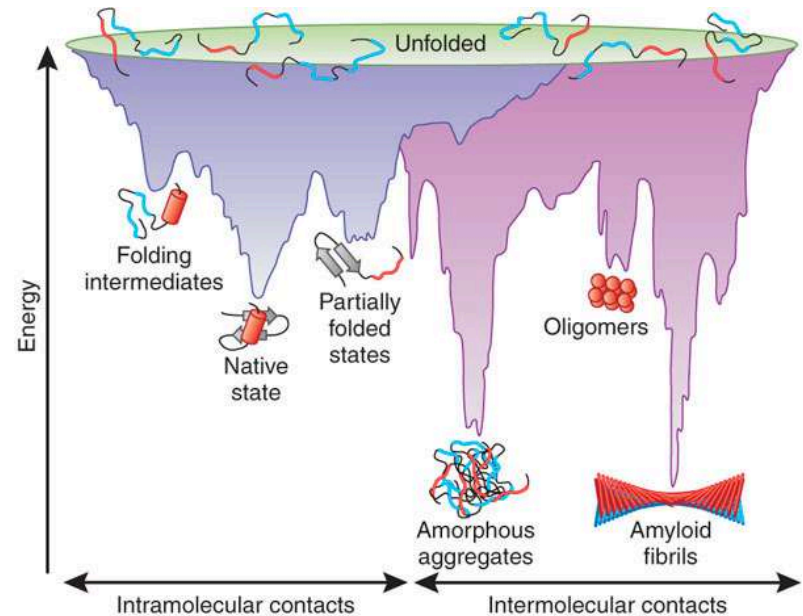
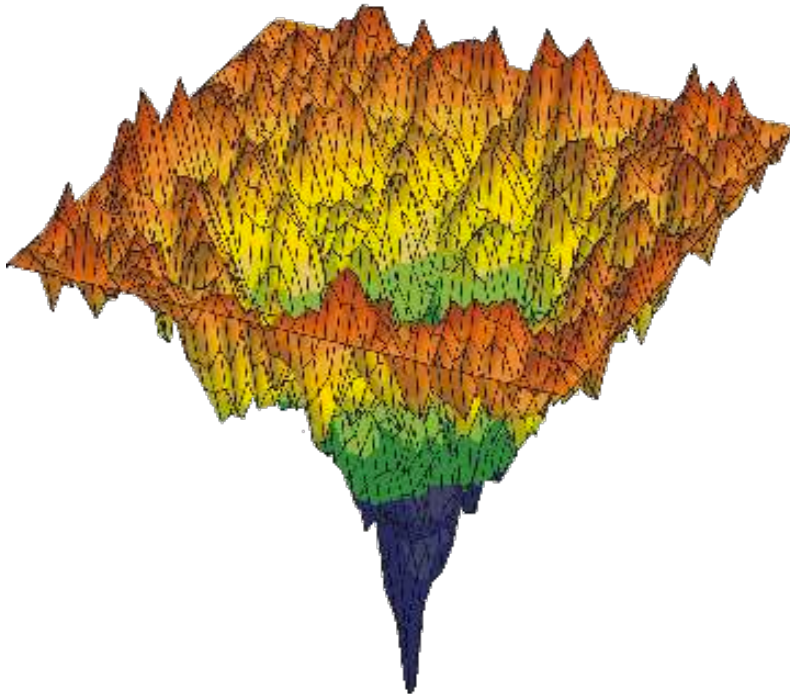
Concept 5. Timescales and Energy Landscape

A complete description of a protein requires a **multidimensional energy landscape** that provides **relative probabilities** of the conformational states (**thermodynamics**) and the **energy barriers** between them (**kinetics**).



Concept 5. Timescales and Energy Landscape

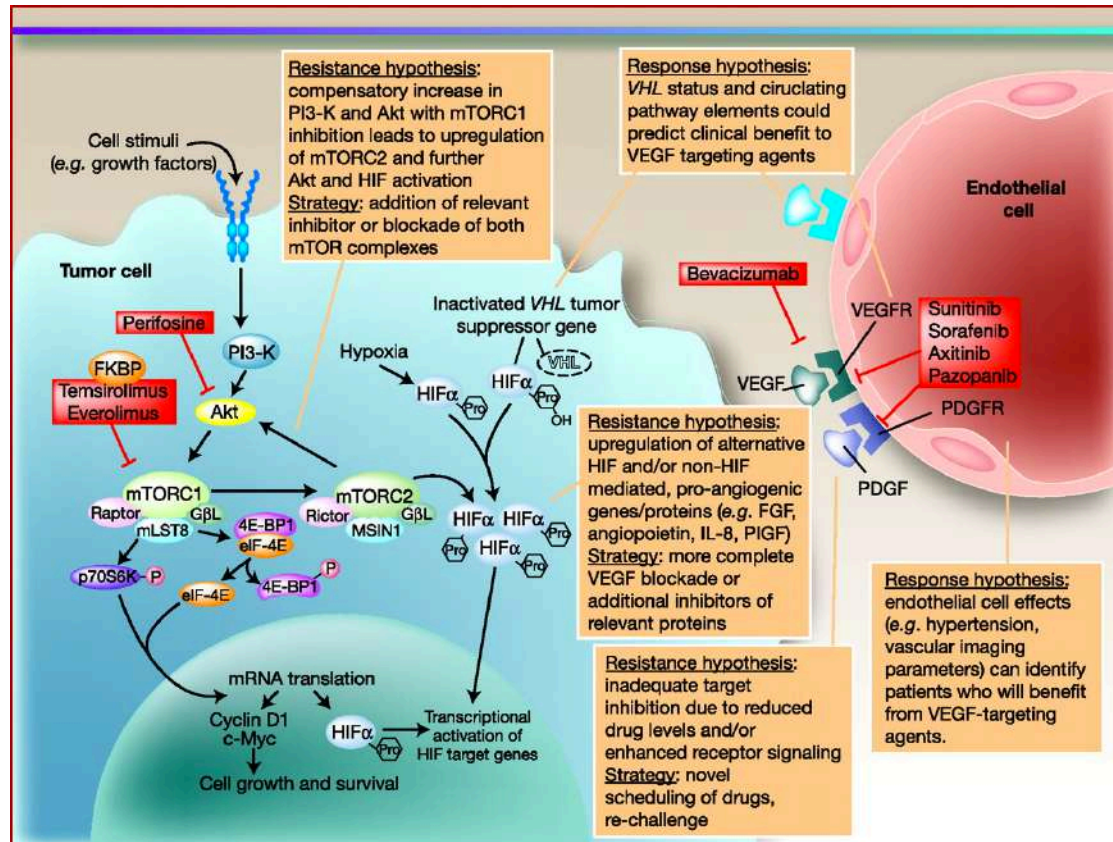
- Potential energy surface of proteins are extremely complex.
- Multidimensional. Multiple minima and transition states:



Concept 6. Molecular Basis of Disease

Proteins play a key role in disease. Importance of mutations.

Molecular basis of cancer:



Concept 7. Complexity

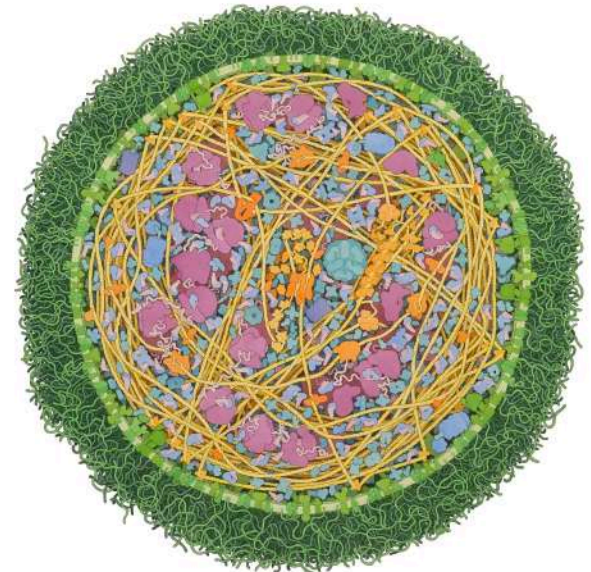
Intracellular environment is crowded (high concentration) and inhomogeneous.

Reactants + Solvent + Catalyst



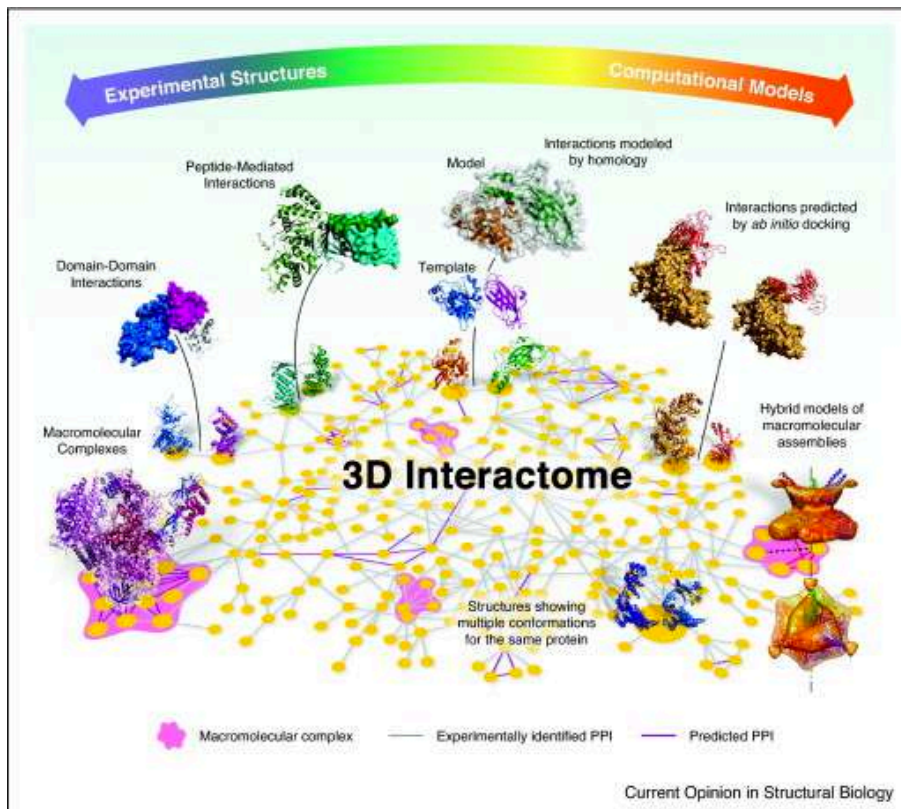
VS

Cell environment + proteins +
membrane + DNA + ...



Concept 8. Biomolecular Interactions

Concept: Proteins interact with other molecules including other proteins, small molecules, peptides, drugs, membrane, DNA, ...



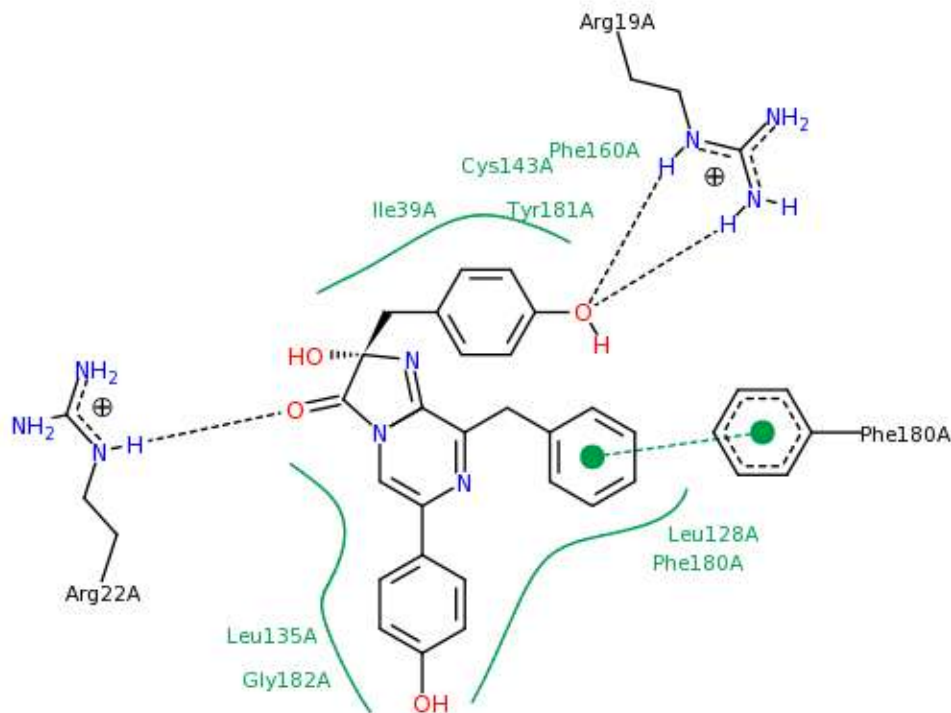
Proteins Don't Know Biology

Proteins Know Physics

Non-Covalent Interactions

Concept 8. Biomolecular Interactions

Concept: Proteins interact with other molecules including other proteins, small molecules, peptides, drugs, membrane, DNA, ...



Molecules Don't
Know Biology

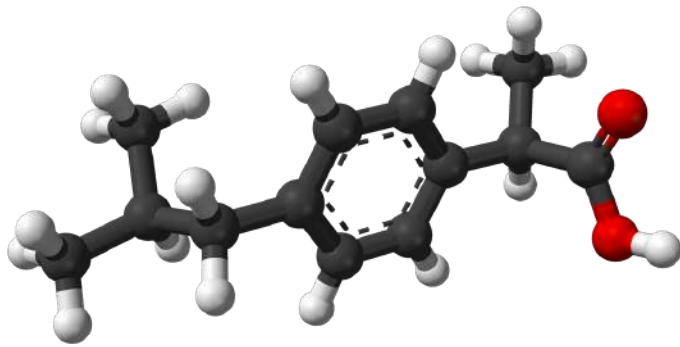
Molecules Know
Physics

Non-Covalent
Interactions

Concept 9. Molecular Modeling

9.1. Many Body Problem

A protein is a many-body system with a tremendous amount of degrees of freedom

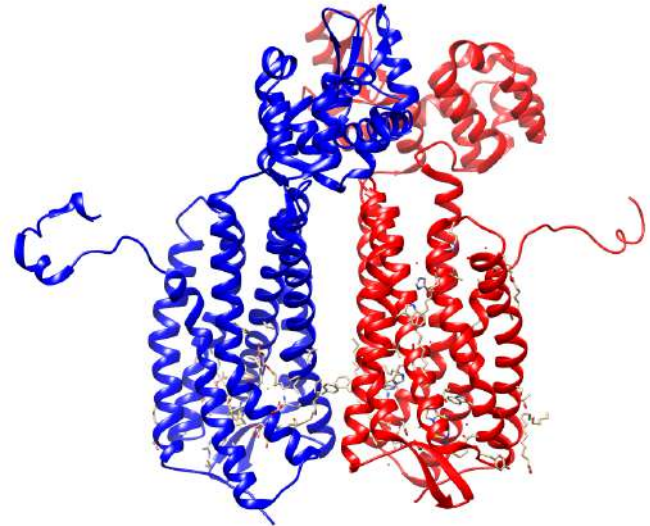


$$\mathbf{R} = (x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_N, y_N, z_N)$$

$N = 33$ atoms

$$3N = 99$$

Dimensions ($3N - 6$): 96



$$\mathbf{R} = (x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_N, y_N, z_N)$$

$N = 75000$ atoms

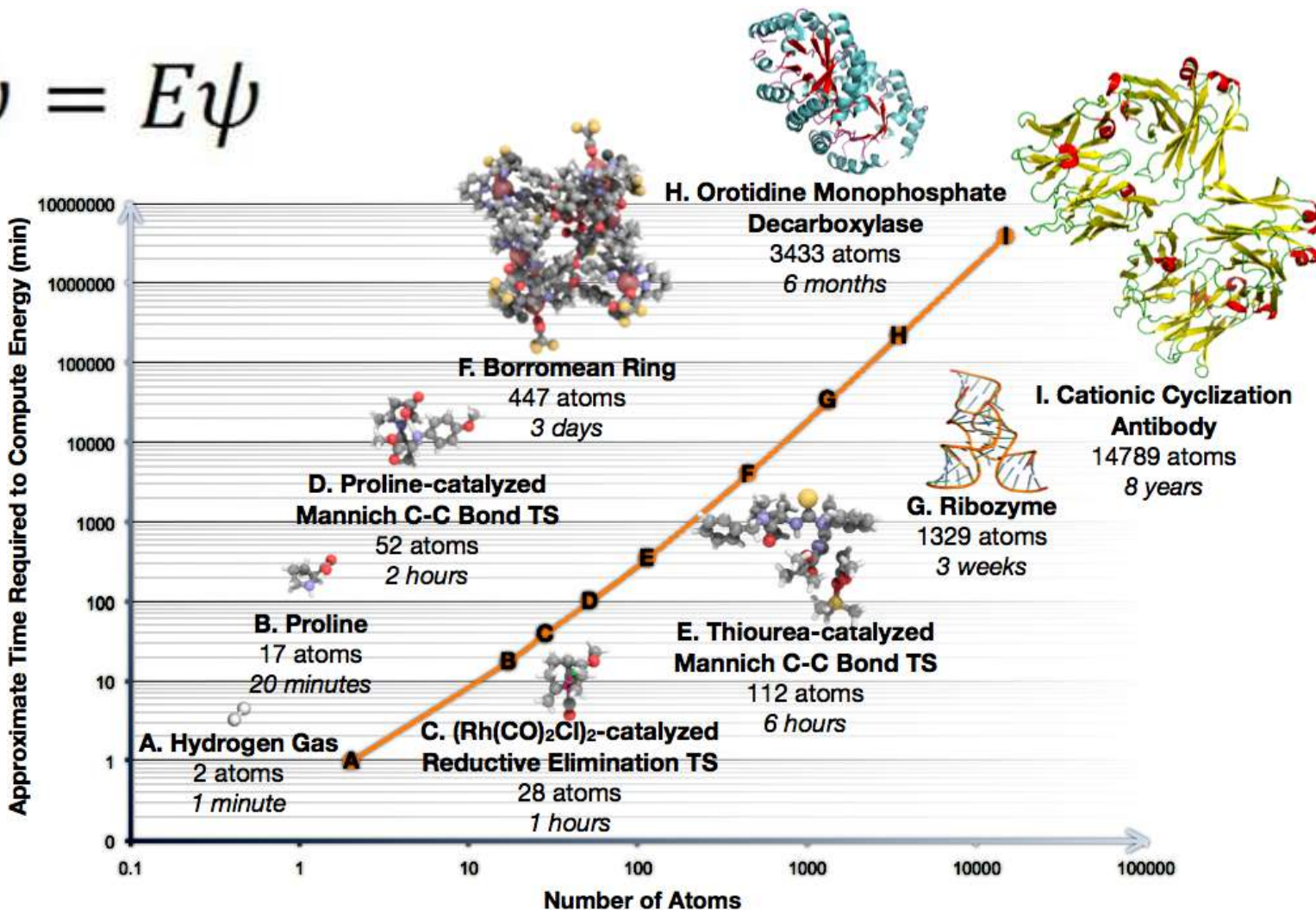
$$3N = 225000$$

Dimensions ($3N - 6$): 224994

Concept 9. Molecular Modeling

9.2. Energy Functional

$$\hat{H}\psi = E\psi$$



Concept 9. Molecular Modeling

9.2. Energy Functional

Molecular Mechanics

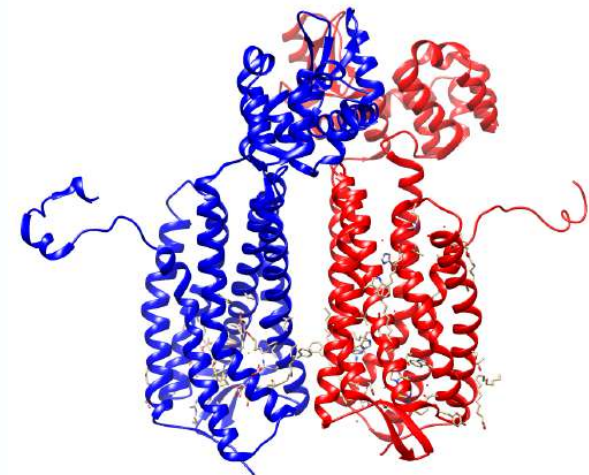
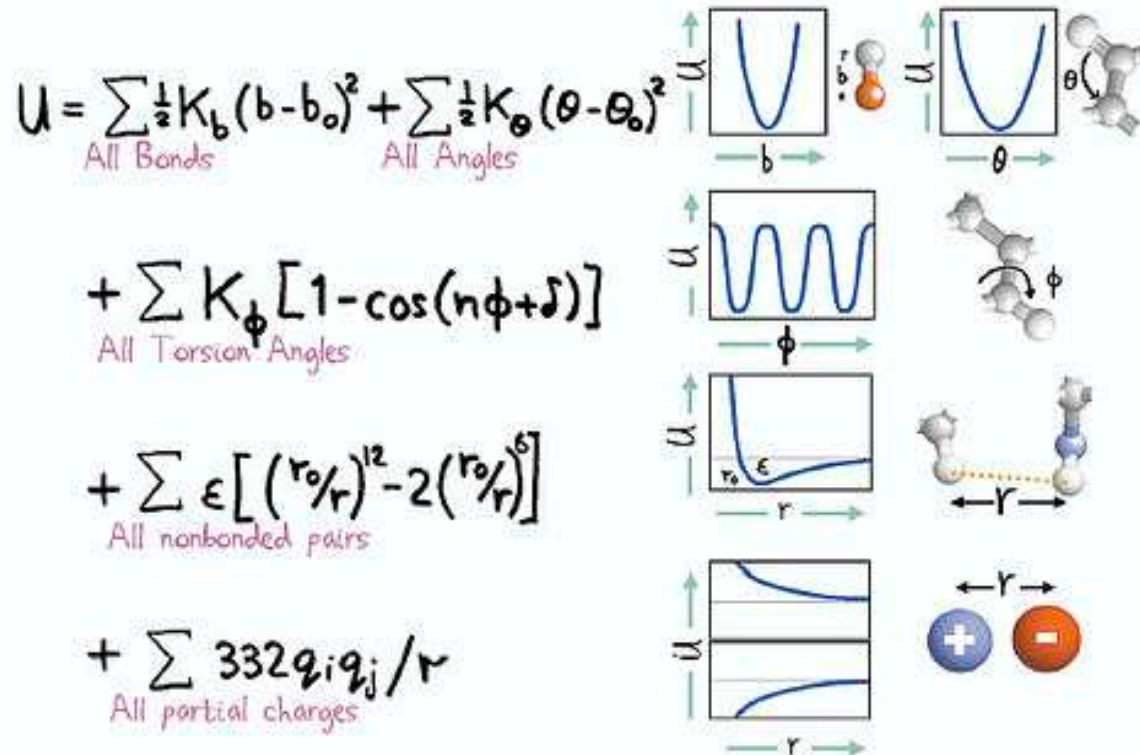


Fig. 1 The total potential energy of any molecule is the sum of terms allowing for bond stretching, bond angle bending, bond twisting, van der Waals interactions and electrostatics. Many properties of a biomolecules can be simulated with such an empirical energy function.

Concept 9. Molecular Modeling

9.3. Molecular Dynamics

Energy Functional Molecular Dynamics

How can we describe protein dynamics?

Quantum Mechanics

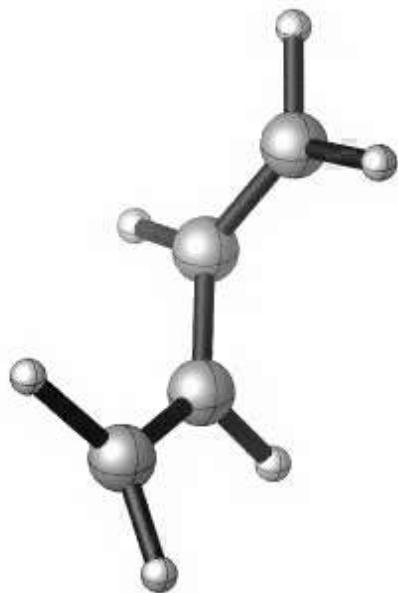
$$\frac{-\hbar^2}{2m} \frac{\partial^2 \Psi(x,t)}{\partial x^2} + U(x)\Psi(x,t) = i\hbar \frac{\partial \Psi(x,t)}{\partial t}$$

Classical Mechanics

$$m \frac{d^2 x(t)}{dt^2} = F = -\frac{dV}{dr}$$

Concept 9.3 Molecules are Dynamic

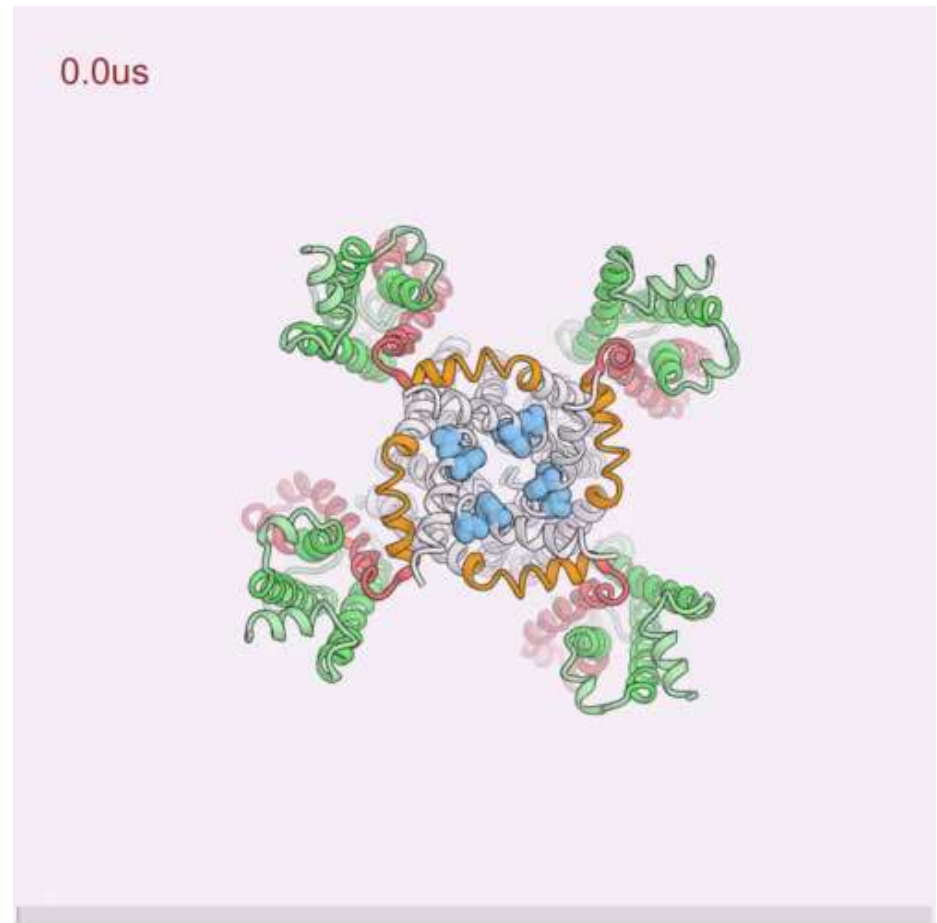
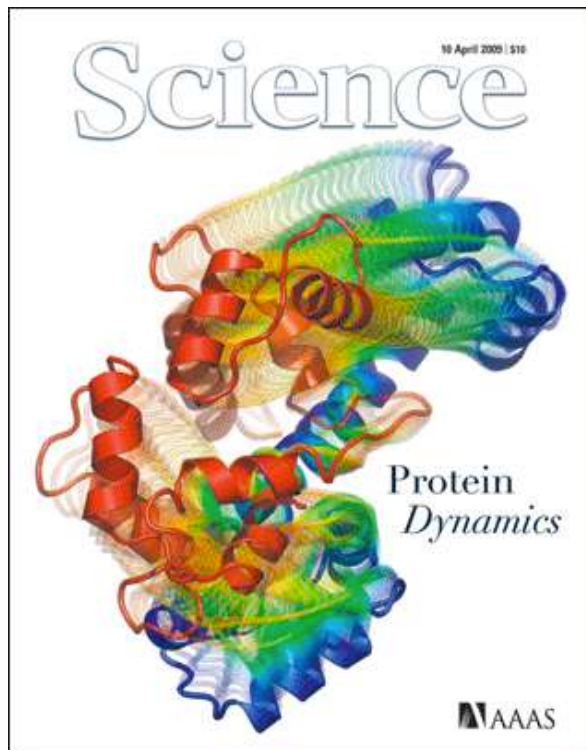
Molecules are not static objects, molecules are **dynamic**.



Diels-Alder Reaction

Concept 9.3 Molecules are Dynamic

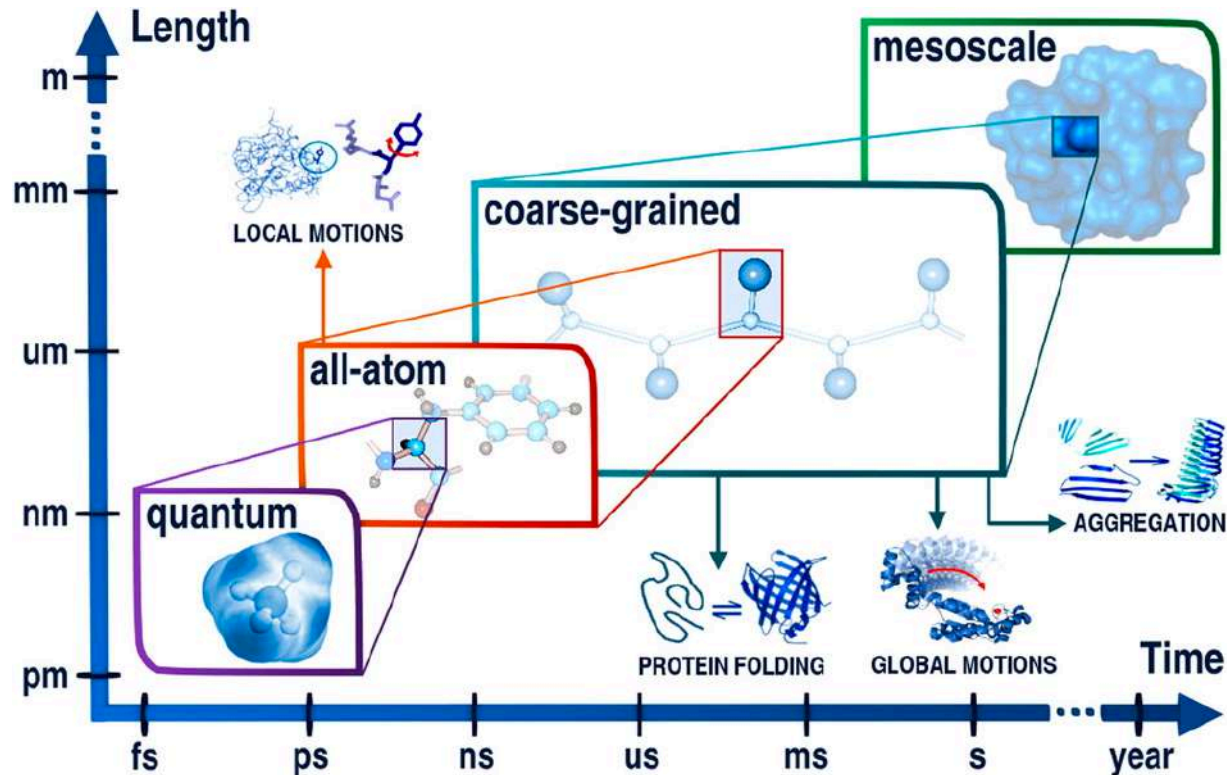
Proteins are dynamic



Potassium ion channel

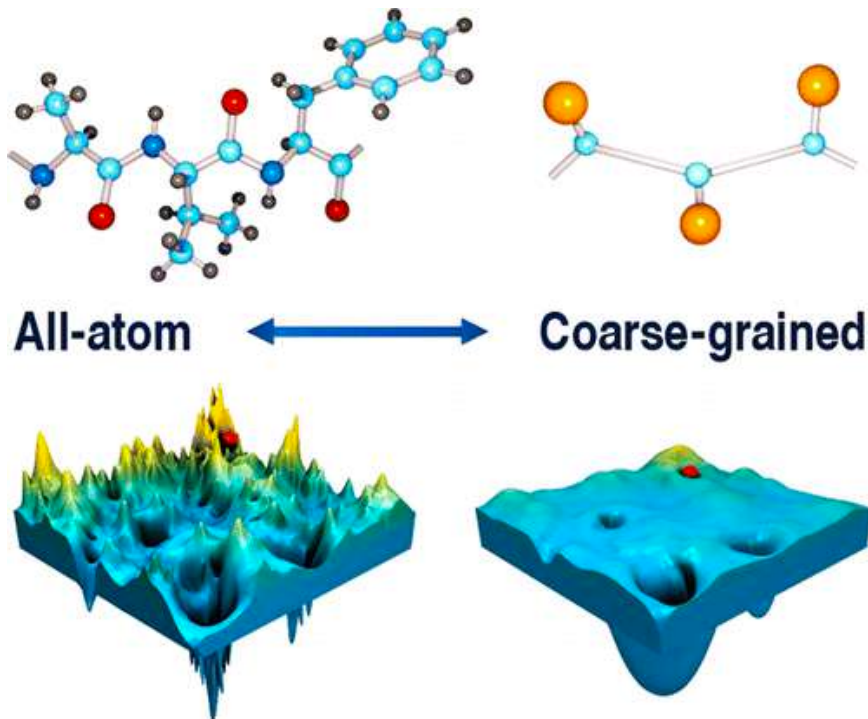
Concept 10. Level of Resolution

- There are different strategies to simplify the many-body problem: electronic level, atomic level, coarse-grain .
- The shape of the potential energy surface will depend on how we describe our system:

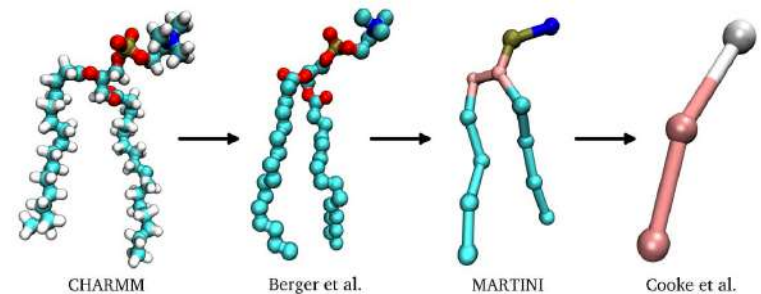


Concept 10. Level of Resolution

- There are different strategies to simplify the many-body problem: electronic level, atomic level, coarse-grain .
- The shape of the potential energy surface will depend on how we describe our system:



- Different levels of coarse-graining. Usually applied to lipid membrane simulations.



– **Dynamic Personalities of Proteins.**

Katherine Henzler–Wildman and Dorothee Kern

<http://www.nature.com/nature/journal/v450/n7172/full/nature06522.html>

– **A Theoretical View of Protein Dynamics.**

Modesto Orozco

<http://pubs.rsc.org/en/content/articlelanding/2014/cs/c3cs60474h#!divAbstract>

– **Quantum mechanical methods in biochemistry and biophysics**

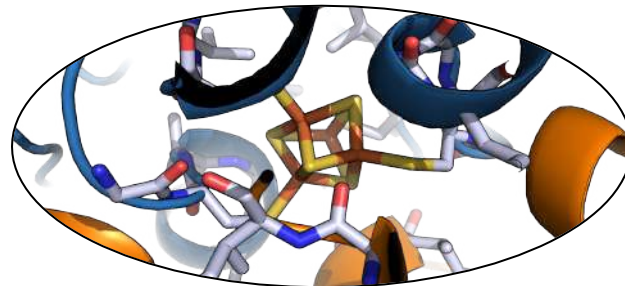
Qiang Cui

<http://aip.scitation.org/doi/full/10.1063/1.4964410>

QUANTUM MECHANICS: For Biomolecules



$$\hat{H}\psi = E\psi$$



- **QM Methods – Post Hartree-Fock**

Quantum Chemistry methods: Interactions between **electrons** determine the **structure and properties** of matter from **molecules** to **proteins** or **solids**.

Hartree-Fock	QM Methods
Configuration Interaction CIS and CISD	
Coupled Cluster CCSD and CCSD(T)	HF
Moller-Plesset Perturbation Theory MP ₂	Post-HF
Multi-Configurational Self-Consistent Field CASSCF and CASP ₂	

$$\hat{H}\psi = E\psi$$

Method + Basis Set

Approximate Schrödinger Equation
Proper description of electron correlation?

Dynamic Correlation: CC high accuracy for energies and molecular properties of small molecular systems dominated by a single configuration. **High computational scaling with system size!**

Largest CCSD(T) calculation: 20 water molecules (1020 basis functions).

Non-Dynamic Correlation: CASSCF and CASPT₂ requires the definition of an active space and a FCI is generated within this space. **High computational scaling with active space size!**

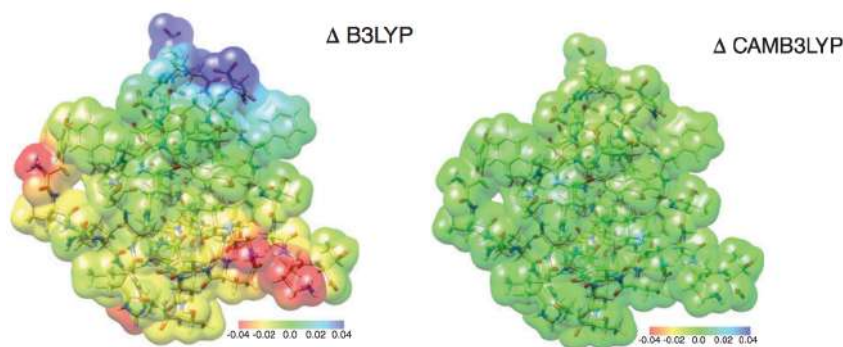
Largest active space for CASSCF/CASPT₂ calculations: 18 electrons and 18 orbitals

• QM Bio – Recent Advances

P. Jorgensen Group (Aarhus)

Divide-Expand-Consolidate (DEC) MP2 and Coupled Cluster

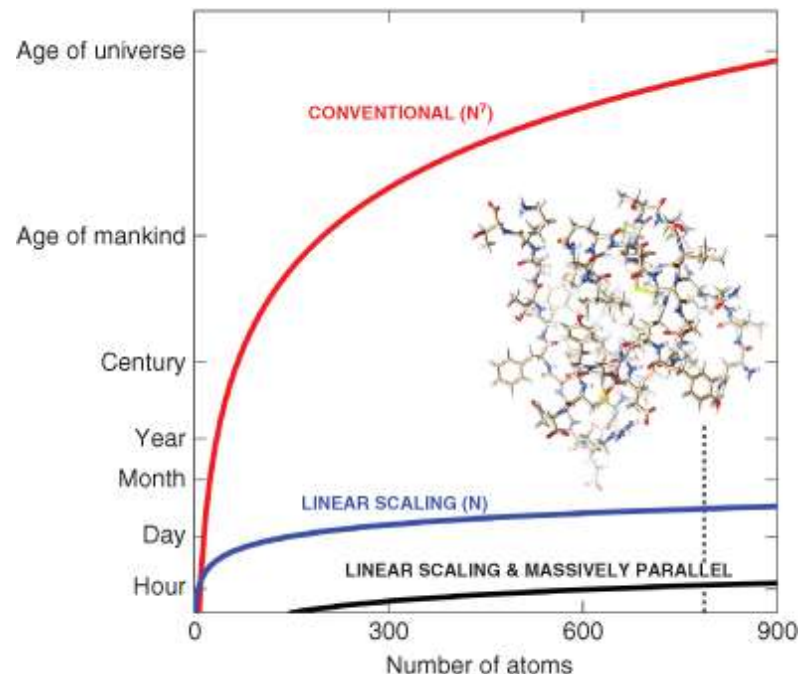
- The DEC Scheme is a linear-scaling and massively parallel framework for high accuracy MP2 and CC calculations on large molecular systems.
- Benchmarking DFT functionals against MP2 electrostatic potential. B3LYP incorrectly predicts partial electron transfer from anionic to cationic sites (self-interaction error)



T. Kjærgaard et al. *WIREs Comp. Mol. Sci.* **2017**, 115, 792-794 ([article](#))

- **Fragmentation-based local correlation approximations:**

$$E \rightarrow \sum_f E_f$$



More about Energy-based Molecular Fragmentation Methods: *Chem. Rev.* **2015**, 7, 5607 ([article](#))

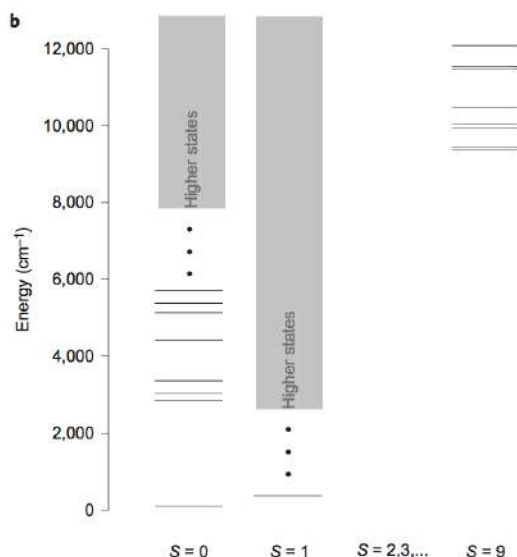
• QM Bio– Recent Advances

G. K. Chan Group (Caltech)

Density-Matrix Renormalization Group (DMRG)

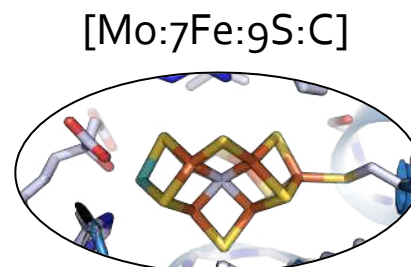
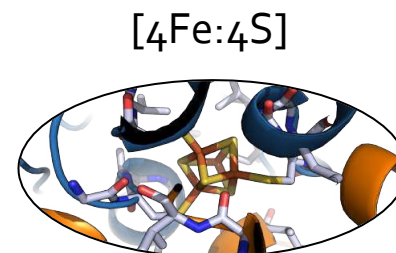
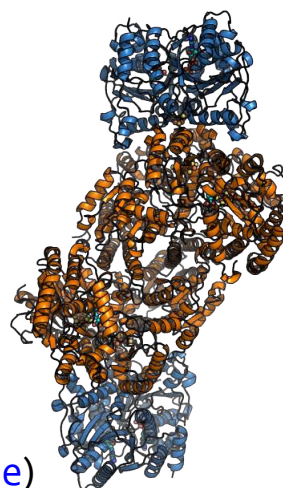
- The group of Garnet K. Chan used DMRG to calculate the electronic levels of $[2\text{Fe-2S}]$ and $[4\text{Fe-4S}]$ clusters and other important biologic cofactors like Mn_4CaO_5 cluster in photosystem II.
- Iron-Sulfur clusters play key roles in biological systems including electron-transfer, redox chemistry and oxygen sensing in processes such as nitrogen fixation, respiration, and photosynthesis.

Ex: Nitrogenase



S. Sharma et al. *Nature Chemistry* **2014**, 6, 927 ([article](#))

Y. Kurashige et al. *Nature Chemistry* **2013**, 5, 660([article](#))



• QM Bio – Recent Advances

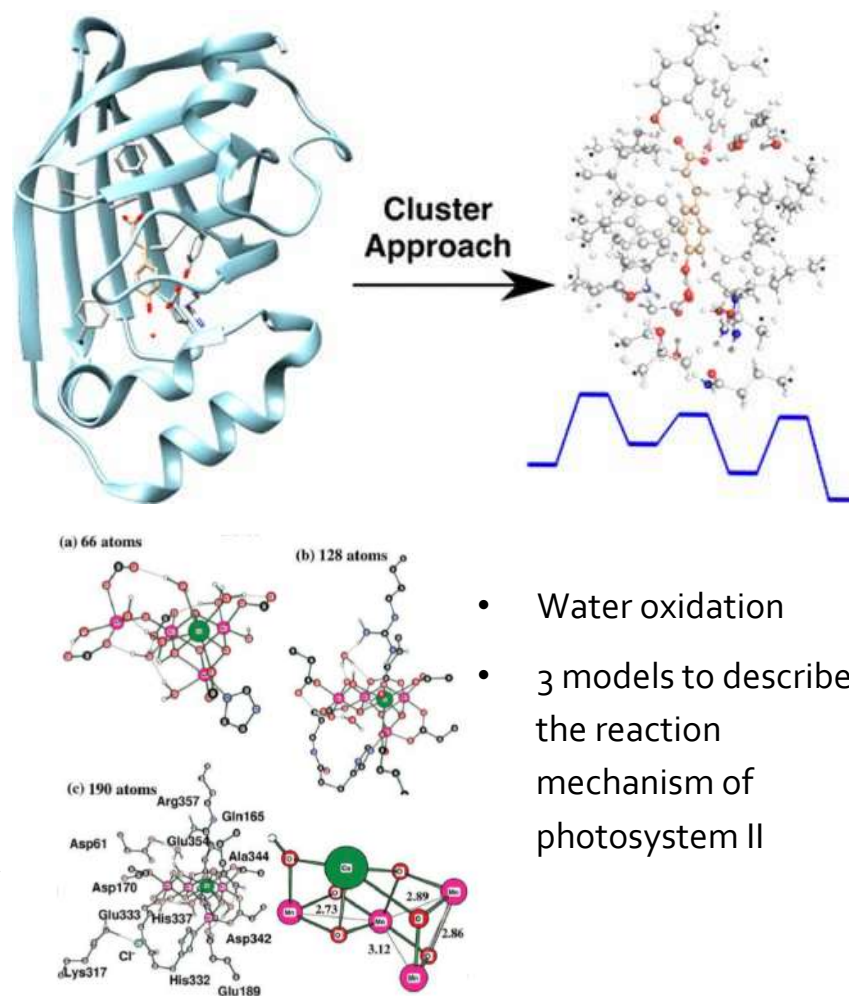
P. Siegbahn and F. Himo (Stockholm)

Enzymatic Reaction Mechanisms – Cluster Model

- The basic idea of the **cluster approach** for modeling **enzyme active sites** and **reaction mechanisms** is to cut out a **relatively small** but well-chosen **part of the enzyme** and treat it with as **accurate quantum chemical methods as possible**.
- To account for the excluded rest of the enzyme, two simple approximations are used, namely the **polarizable continuum model** and a coordinate-locking scheme (the position of some atoms is restrained)
- As the **model size grows** and more **flexibility** is granted to the groups of the model, this approximation works **better** and better.

P. Siegbahn and F. Himo *WIREs Comp. Mol. Sci.* **2011**, 1, 323-336 ([article](#))

F. Himo *J. Am. Chem. Soc.* **2017**, 139, 6780 ([article](#))



QM Bio – Recent Advances

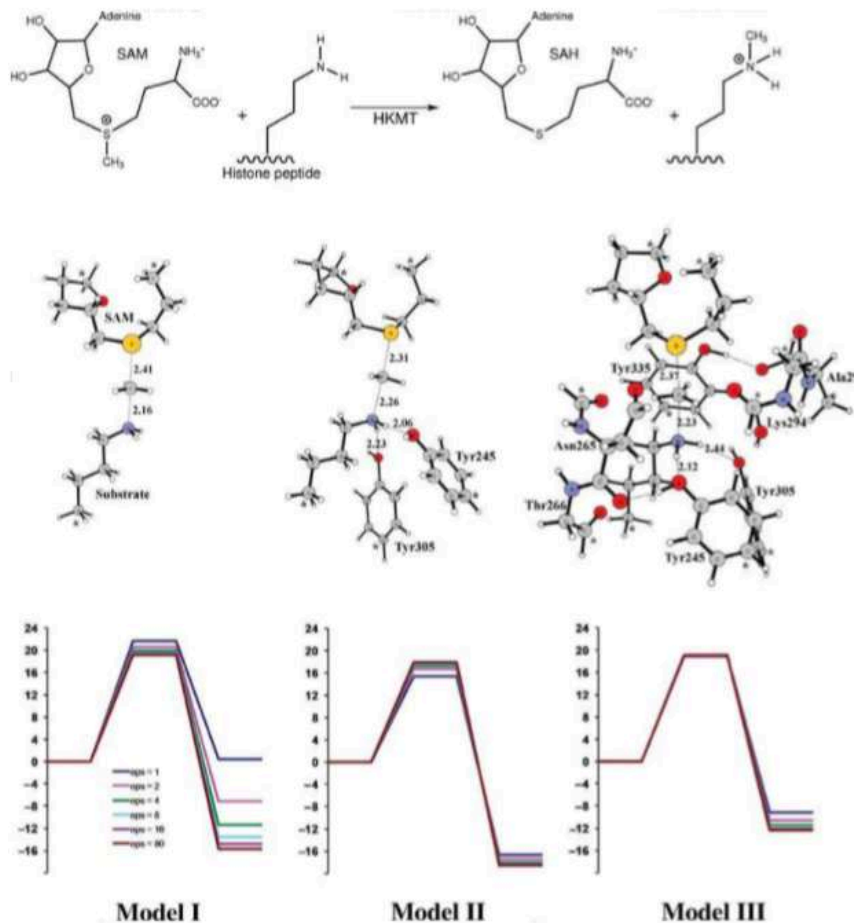
P. Siegbahn and F. Himo (Stockholm)

Enzymatic Reaction Mechanisms – Cluster Model

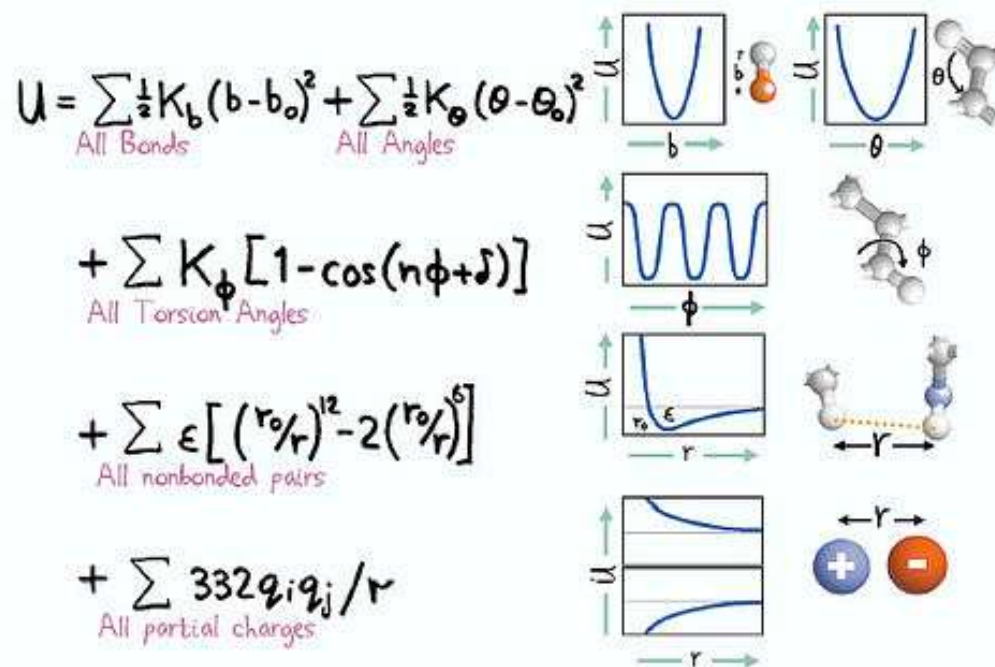
- The basic idea of the **cluster approach** for modeling **enzyme active sites** and **reaction mechanisms** is to cut out a **relatively small** but well-chosen **part of the enzyme** and treat it with as **accurate quantum chemical methods** as possible.
- To account for the excluded rest of the enzyme, two simple approximations are used, namely the **polarizable continuum model** and a coordinate-locking scheme (the position of some atoms is restrained)
- As the **model size grows** and more **flexibility** is granted to the groups of the model, this approximation works **better** and better.

P. Siegbahn and F. Himo *WIREs Comp. Mol. Sci.* **2011**, 1, 323-336 ([article](#))

F. Himo *J. Am. Chem. Soc.* **2017**, 139, 6780 ([article](#))

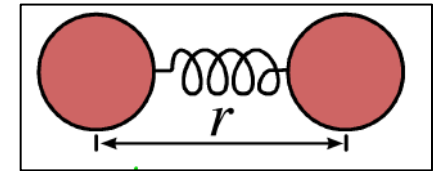


INTRODUCTION TO MOLECULAR MECHANICS: Force Fields and Recent Advances



• Basics of Molecular Mechanics

- **Molecular Mechanics** or **force-field** methods use **classical physical models** to predict the **energy** of a **molecule** as a **function** of its **conformation** (coordinates, r^N).
- **Quantum effects** are explicitly **neglected** giving large **computational efficiency**.
- It consists of using 'balls-on-springs' scheme to represent molecules. Each solid ball corresponds to an atom and each spring a covalent bond.
- Atoms oscillate around an **equilibrium distance**.
- **Atomic Resolution**



Quantum Mechanics (QM)

The positions of **electrons** and **nuclei** are taken into account to calculate the electronic **energy**

Molecular Mechanics (MM)

Only the position of **nuclei (atom)** are taken into account to calculate the **energy**

We assume **Born-Oppenheimer** approximation

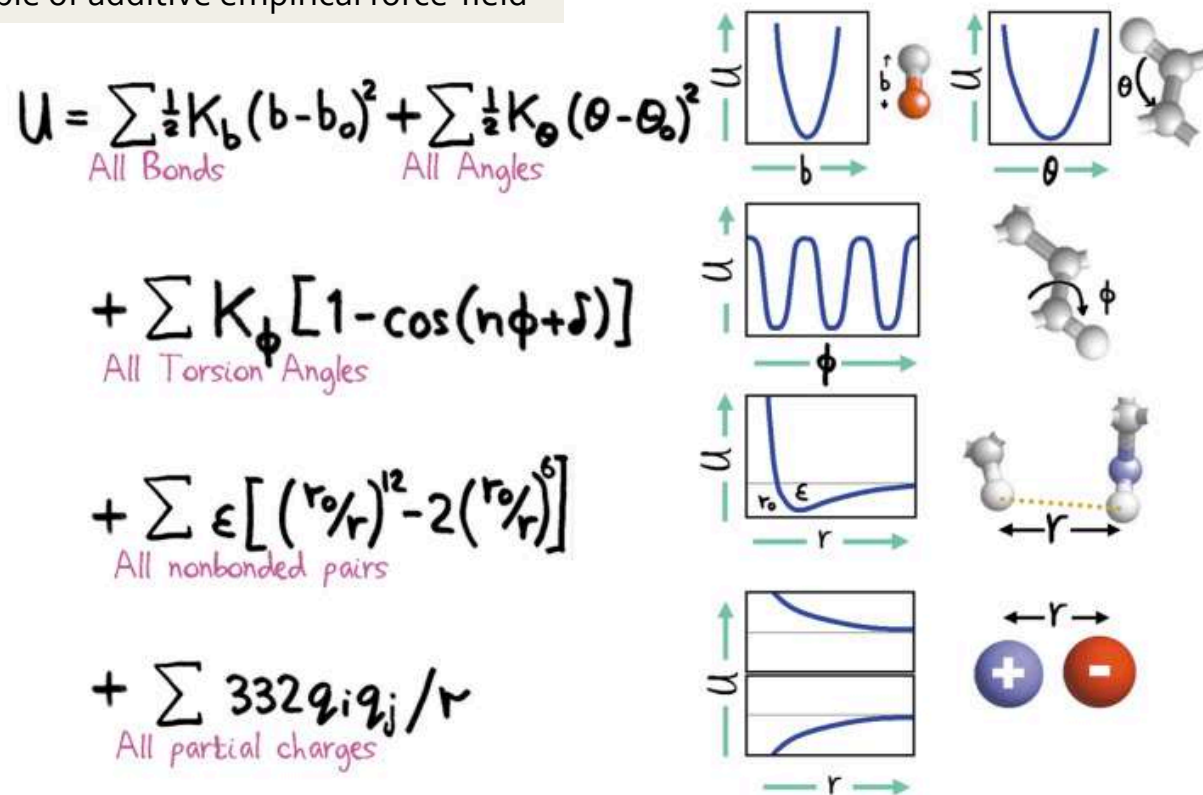
- **Basics of Molecular Mechanics**

- What **can** we study with a **MM Force Field** methods?
 - **Equilibrium geometries** (minima) and **transition states**
 - **Relative energies** between **conformers** or between **different molecules**
- What **cannot** we study with a **MM Force Field** methods?
 - FFs are not appropriate (nor are most quantum mechanics methods!) for **bond-breaking reactions**.
 - MM cannot provide **properties** that depend upon the **electronic distribution** of a molecule (e.g. **electron density**).

- Basics of Molecular Mechanics

- General form of a force-field potential energy (U, V, or E) function:

Example of additive empirical force-field



The birth of computational structural biology

M. Levitt *Nat. Struct. Biol.* 2001, 8, 392 ([article](#))

Basics of Molecular Mechanics

- A potential energy function alone does not make a force-field! It is the **combination** of the **potential energy function** with a **set of parameters** used in that function.

Example of Class I Additive Force Field:

$$U(\vec{r}^N) = \sum_{\text{bonds}} \frac{1}{2} k_{AB} (r_{AB} - r_{AB,eq})^2$$
$$+ \sum_{\text{angles}} \frac{1}{2} k_{ABC} (\theta_{ABC} - \theta_{ABC,eq})^2$$
$$+ \sum_{\text{dihedrals}} \frac{1}{2} V_{ABCD} (1 + \cos(n\omega_{ABCD} + \gamma_{ABCD}))$$
$$+ \sum_A \sum_{B=A+1}^N \left(4\epsilon_{AB} \left[\left(\frac{\sigma_{AB}}{r_{AB}} \right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}} \right)^6 \right] + \frac{q_A q_B}{4\pi\epsilon_0 r_{AB}} \right)$$

Energy bonded interactions (1-4 interactions)

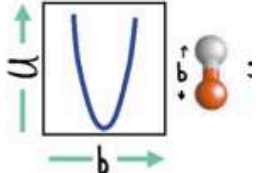
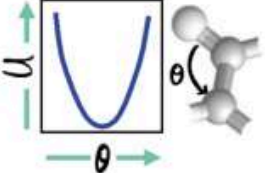
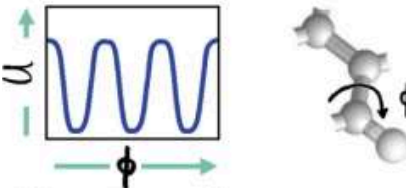
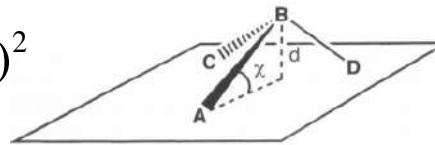
Energy non-bonded interactions

Collection of simple functions:

- Bond and angle** energies are described with harmonic potentials.
- Torsions (dihedrals)** by a sinusoidal term.
- Interactions between atoms (atomic repulsion and dispersion)** by a Lennard-Jones 6-12 term.
- Electrostatic** interactions between atoms are treated via a Coulombic term.

Molecular Mechanics and Molecular Dynamics Simulations

- Bonded Terms of the Force-Field**

FF Term	Variables and Parameters	Expression	Representation
Bond Stretching	r_{AB} is the bond length, $r_{AB,eq}$ is the equilibrium distance and k_{AB} is the bond force constant.	$\sum_{bonds} \frac{1}{2} k_{AB} (r_{AB} - r_{AB,eq})^2$	
Angle Bending	θ_{ABC} is the angle, $\theta_{ABC,eq}$ is the equilibrium angle and k_{ABC} is the angle force constant.	$\sum_{angles} \frac{1}{2} k_{ABC} (\theta_{ABC} - \theta_{ABC,eq})^2$	
Torsional Energy	V_{ABCD} is the dihedral force constant, n is the multiplicity of the cos function, and γ is the phase angle.	$\sum_{dihedrals} \frac{1}{2} V_{ABCD} (1 + \cos(n\omega_{ABCD} + \gamma_{ABCD}))$	
Out-of-plane distortions	Φ is the improper angle, φ_{ABC} the equilibrium improper angle and K_{imp} is the improper angle force constant (not in all FFs)	$\sum_{improper} \frac{1}{2} k_{imp} (\varphi_{ABC} - \varphi_{ABC,eq})^2$	

Force planarity in sp^2 central atoms

- Non-bonded potential energy terms of a Force-Field**

FF Term

Variables and Parameters

Expression

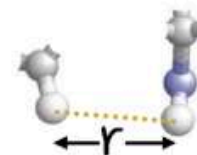
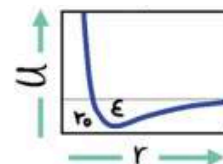
Representation

Repulsion-dispersion Term (van der Waals term)

r_{AB} is the distance between atoms A and B, ϵ_{AB} is the well-depth, σ_{AB} is the interatomic distance at which repulsive and attractive forces exactly balance.

$$\sum_A^N \sum_{B=A+1}^N \left(4\epsilon_{AB} \left[\left(\frac{\sigma_{AB}}{r_{AB}} \right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}} \right)^6 \right] \right)$$

Lennard-Jones 6-12 potential

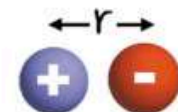
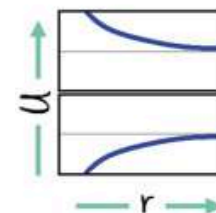


Electrostatic Term

r_{AB} is the distance between atoms A and B, q_A and q_B are the charges of the atoms A and B respectively, and ϵ_0 is the dielectric constant (usually set to 1).

$$\sum_A^N \sum_{B=A+1}^N \left(\frac{q_A q_B}{4\pi\epsilon_0 r_{AB}} \right)$$

Coulomb's law



• Basics of Molecular Mechanics

- In general Class I force fields are adequate for biomolecular simulations at room temperature.
- Class I force fields provide accurate heats of vaporization, molecular volumes, and free energies of solvation.
- The functional forms used are a compromise between **accuracy** and **computational efficiency**.
- In energy **minimization** and **molecular dynamics** it is needed to compute **first** and **second derivatives** of the energy with respect to the atomic coordinates. Therefore, it is important to select force-field expressions that are easily differentiable by a computer.
 - Exercise: estimate the first and second derivatives of the potential energy function that we have discussed before. Is the force-field expression easily differentiable?

The importance of Atom Type

- In a force field it is necessary to assign an **atom type** to each atom of the system.
- The atom type contains information about the **hybridization** and **local environment**. For example, sp^3 carbon atoms adopt a tetrahedral geometry and sp^2 carbons are trigonal, while sp -carbons are linear.

Examples of atom types (MM2)

Symbol	Description
C_{sp^3}	sp^3 -carbon
C_{sp^2}	sp^2 -carbon, alkene
$C_{sp^2,C=O}$	sp^2 -carbon, carbonyl
$C_{sp^2,arom}$	sp^2 -carbon, aromatic
$O_{sp^2,C=O}$	sp^2 -oxygen, carbonyl
H_{OH}	Alcohol, OH
H_{COOH}	Carboxyl (COOH)

Force constants and equilibrium distances MM2

Bond	k_{AB} (kcal/Å ²)	$r_{AB,eq}$ (Å)
$C_{sp^3}-C_{sp^3}$	317	1.523
$C_{sp^3}-C_{sp^2}$	360	1.497
$C_{sp^3}-C_{sp^2,C=O}$	317	1.509
$C_{sp^2,C=O}-O_{sp^2,C=O}$	777	1.208

MM2 is a force field designed to reproduce equilibrium geometries of small organic molecules (Allinger *J. Am. Chem. Soc.* 99 8127-8134 ([article](#)))

• Coulombic Term

- Electrostatic term describe the **Coulomb interaction** between **atoms A and B** with **fixed partial charges** (e.g., in carbonyls), according to:

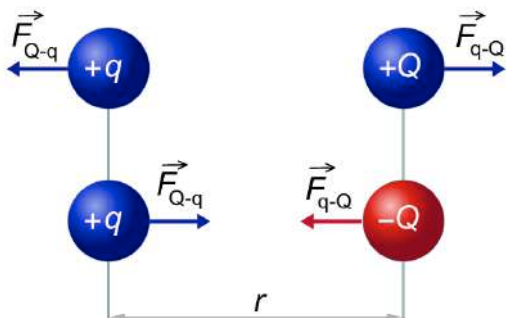
$$\sum_A^N \sum_{B=A+1}^N \left(\frac{q_A q_B}{4\pi\epsilon_0 r_{AB}} \right)$$

q_A partial charge of atom A

q_B partial charge of atom B

r_{AB} interatomic distance

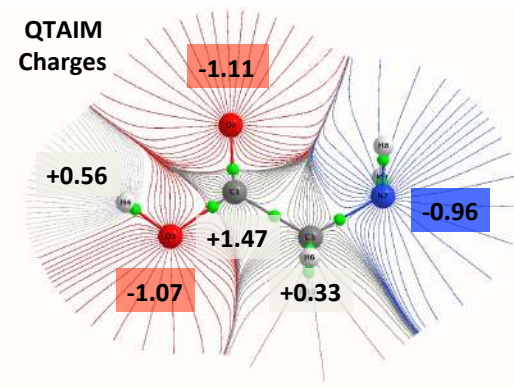
ϵ_0 dielectric constant



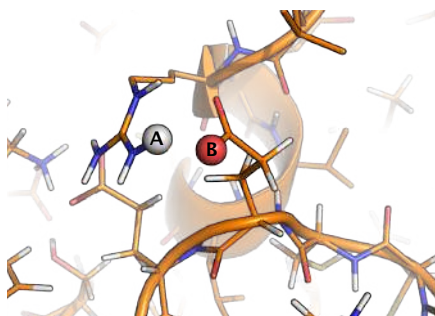
- **Atoms don't have charges (we can not measure it)!!** Quantum mechanics tells us **electrons are delocalized**, and their charge is shared among nearby atoms.

- **No unique way to assign electrons to atoms.**

- Several methods can be used to obtain atomic charges (Chemical Bonding course).



Glycine



Molecular Mechanics and Molecular Dynamics Simulations

- Some famous Force-Fields

Small Molecule Force-Fields

Protein Force-Fields

Allinger Group Force-Fields
(University of Georgia, USA)

Force Fields **MM₂, MM₃,
MM₄**

Systems *Small organic
molecules
(hydrocarbons)*

Lifson Group Force-Fields
(Weizmann Institute, Israel)

Force Fields **CFF**

Systems *Small organic
molecules and
metals*

Jorgensen Group Force-Fields
(Yale, USA)

Force Fields **OPLS**

Systems *Water and
organic liquids,
simulations on
condensed
phase*

Kollmann Group Force-Fields
(UCSF, USA)

Force Fields **AMBER**

Systems *Proteins, DNA,
Lipids*

Karplus Group Force-Fields
(Harvard, USA and
Strasbourg, France)

Force Fields **CHARMM**

Systems *Proteins,
DNA, Lipids*

van Gunsteren and Berendsen
Force-Fields
(ETH Zurich, Switzerland and
Groningen, The Netherlands)

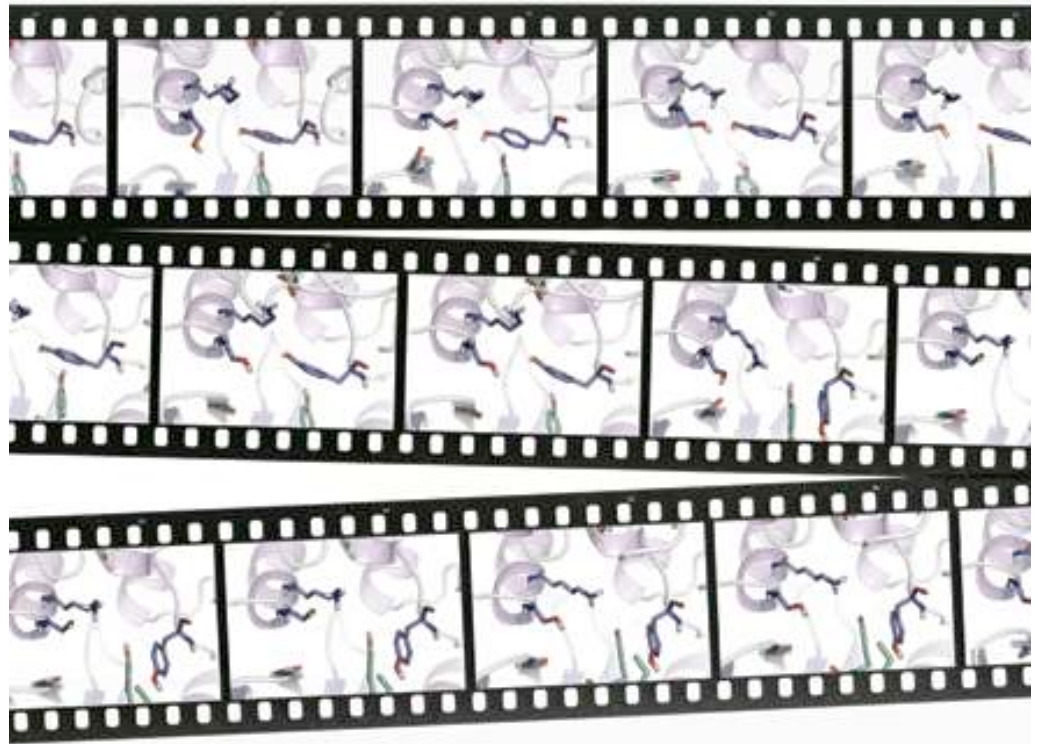
Force Fields **GROMOS**

Systems *Proteins, DNA,
Lipids*

• **General Aspects on Biomolecular Force Fields**

- A highly accurate potential energy function is crucial to an atomic description of a biological function.
- Most popular biomolecular force-fields employ simple functional forms and explicitly describe a limited range of physical phenomena associated with intra- and intermolecular interactions. For example: charge transfer is not considered.
- The accuracy of the force-field relies to a significant degree on error compensation among different components. This limits the degree of transferability between different molecules.
- Force-fields are designed to match QM calculations or experimental observables.
- Molecular Mechanics can be used in Monte Carlo and Molecular Dynamics simulations to simulate the ensemble of conformations that a molecule can adopt.

INTRODUCTION TO MOLECULAR DYNAMICS: For Biomolecules



REPORT

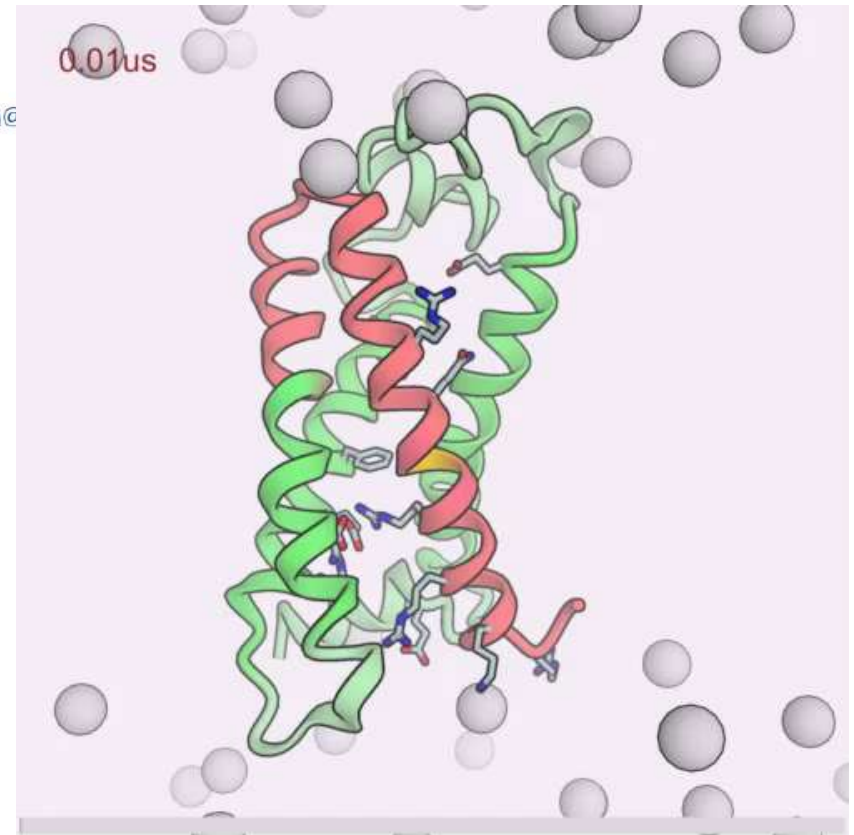
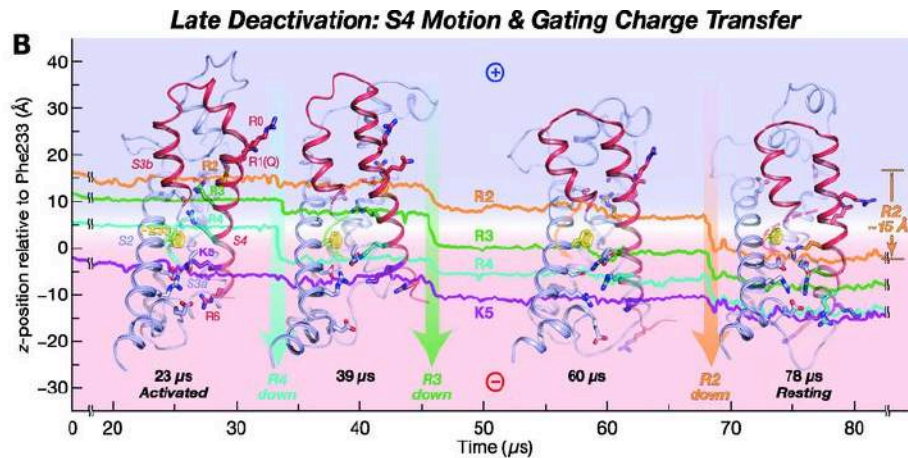
Mechanism of Voltage Gating in Potassium Channels

Morten Ø. Jensen^{1,*}, Vishwanath Jogini¹, David W. Borhani¹, Abba E. Leffler¹, Ron O. Dror¹, David E. Shaw^{1,2,*}

+ Author Affiliations

←*To whom correspondence should be addressed. E-mail: morten.jensen@david.shaw@DEShawResearch.com (D.E.S.)

Science 13 Apr 2012:
Vol. 336, Issue 6078, pp. 229-233
DOI: 10.1126/science.1216533



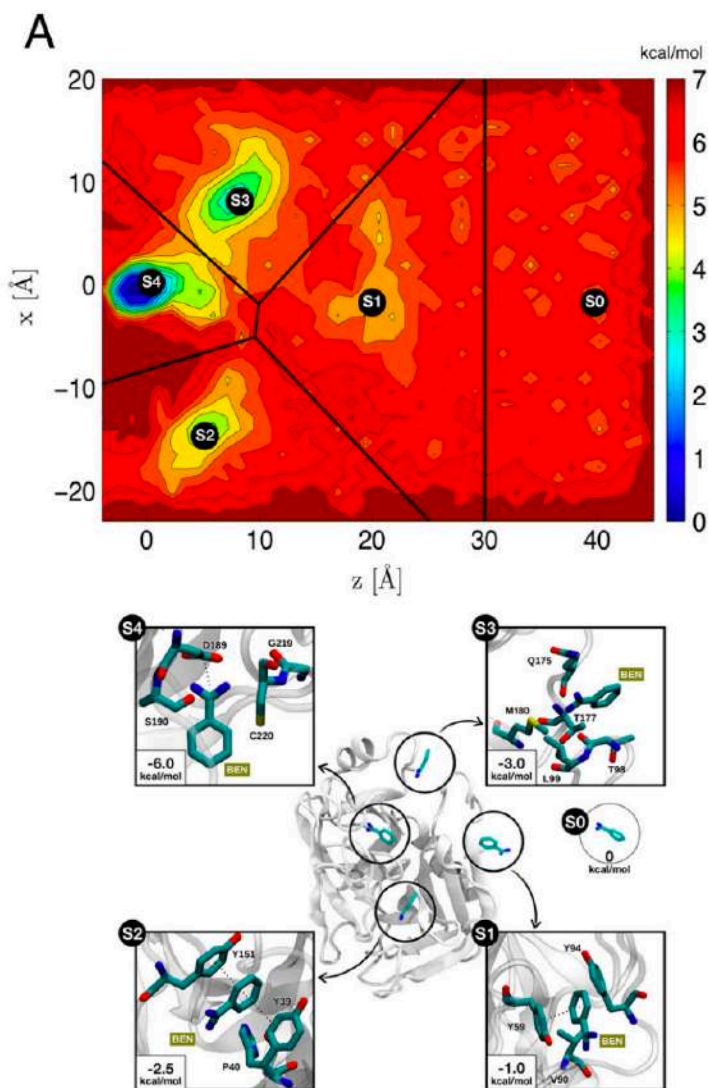
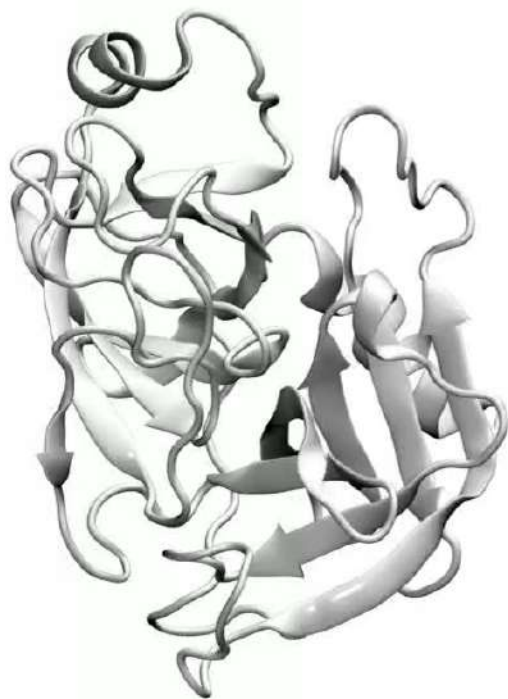
Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations

Ignasi Buch, Toni Giorgino, and Gianni De Fabritiis¹

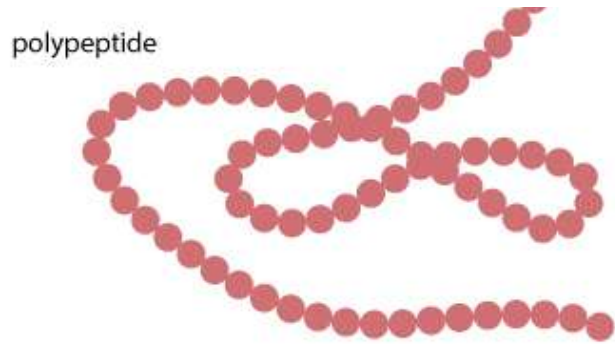
Computational Biochemistry and Biophysics Laboratory, Universitat Pompeu Fabra, Barcelona Biomedical Research Park, C/Doctor Aiguader 88, 08003 Barcelona, Spain

Edited by Arieh Warshel, University of Southern California, Los Angeles, CA, and approved May 11, 2011 (received for review March 4, 2011)

trypsin-benzamidine

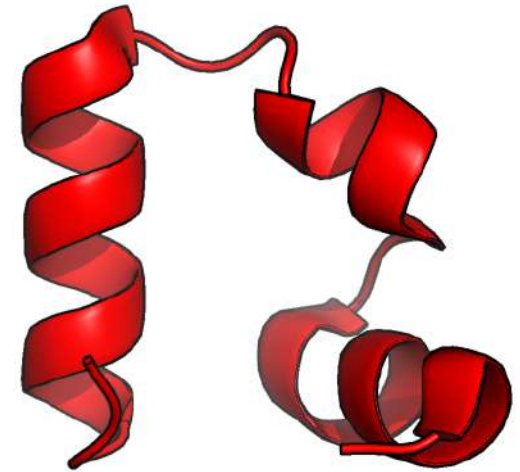


Molecular Mechanics and Molecular Dynamics Simulations

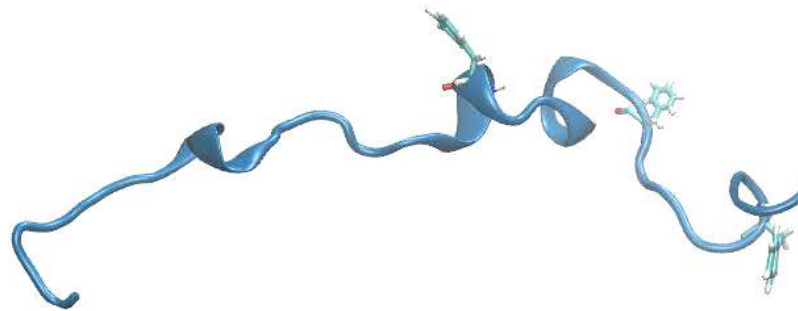


Unfolded state

Protein
Folding



Native state



Molecular dynamics simulations permit the study of complex, dynamic processes that occur in biological systems. These include, for example:

- NMR or X-ray structure refinements
- Protein structure prediction
- Protein folding kinetics and mechanics
- Conformational dynamics
- Global optimization
- DNA/RNA simulations
- Membrane proteins/lipid layers simulations

Molecular Mechanics and Molecular Dynamics Simulations

Molecular dynamics simulations calculate the motion of the atoms in a molecular assembly using **Newtonian dynamics** to determine the net force and acceleration experienced by each atom. Each atom i at position \mathbf{r}_i is treated as a point with a mass m_i and a fixed charge q_i .

$$\mathbf{F}_i(\mathbf{t}) = m_i \mathbf{a}_i(\mathbf{t}) \quad (i = 1, 2, \dots, N)$$

$$\mathbf{F}_i(\mathbf{t}) = m_i \frac{d^2 \mathbf{r}_i(\mathbf{t})}{d\mathbf{t}^2}$$

$$\mathbf{F}_i(\mathbf{t}) = -\frac{dU}{d\mathbf{r}_i}$$

$$-\frac{dU}{d\mathbf{r}_i} = m_i \frac{d^2 \mathbf{r}_i(\mathbf{t})}{d\mathbf{t}^2} = m_i \mathbf{a}_i(\mathbf{t})$$

Knowledge of the atomic forces and masses can be used to solve the position of each atom along a series of extremely small time steps (on the order of femtoseconds = 10^{-15} seconds).

Molecular Mechanics and Molecular Dynamics Simulations

In MD, successive configurations of the system are generated by integrating Newton's laws of motion. The result is a **trajectory** that specifies how the positions and velocities of the particles in the system vary with time.

TRAJECTORY: The resulting series of **snapshots** of structural changes over time. MD is a time-dependent technique, from which time-average of properties can be calculated.



$$-\frac{dU}{dr_i} = m_i \frac{d^2 r_i(t)}{dt^2} = m_i a_i(t)$$

This equation describes the motion of a particle of mass m_i along the coordinate r with F being the force on the particle in that direction.

It is helpful to distinguish three different types of problem to which Newton's laws of motion can be applied:

- Simplest case: no force acts on each particle between collisions. From one collision to the next, the position of the particle thus changes by $v_i \Delta t$.
- The particle experiences a constant force between collisions.
- The force on each particle depends on its position relative to other particles. More realistic situation, but as the motions of all particles are coupled together \rightarrow many-body problem that cannot be solved analytically. \rightarrow equations of motion integrated using a finite difference method.

Finite Difference Methods

Finite difference methods are used to generate molecular dynamics trajectories with continuous potential models (force fields), which we will assume to be pairwise additive.

Essential idea: the integration is broken down into many small stages, each separated in time by a **fixed time δt** .

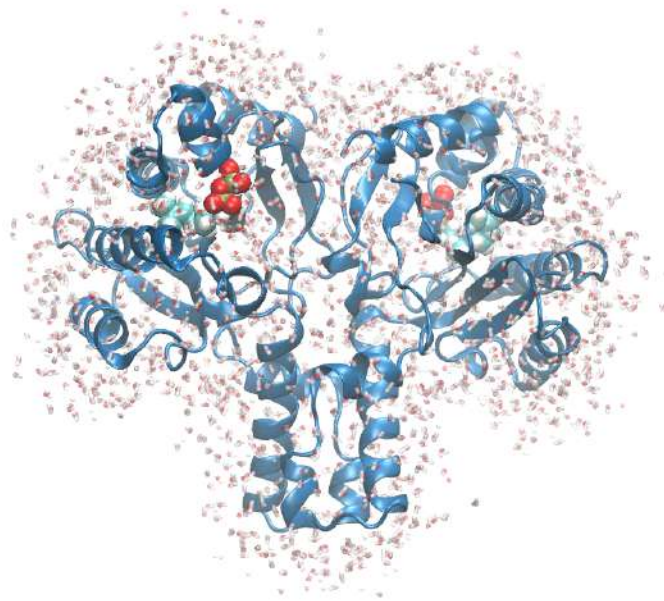
Total force on each particle in the configuration at a time t : sum of its interactions with other particles (force field).

F \rightarrow **accelerations** are determined $\rightarrow a + r + v$ at time $t \rightarrow$ calculate the positions and velocities at a time $t + \delta t$

Then: **F** on the particles in their new positions are determined, leading to new positions and velocities at a time $t + 2\delta t$, and so on.

Steps in Molecular Dynamics Simulations

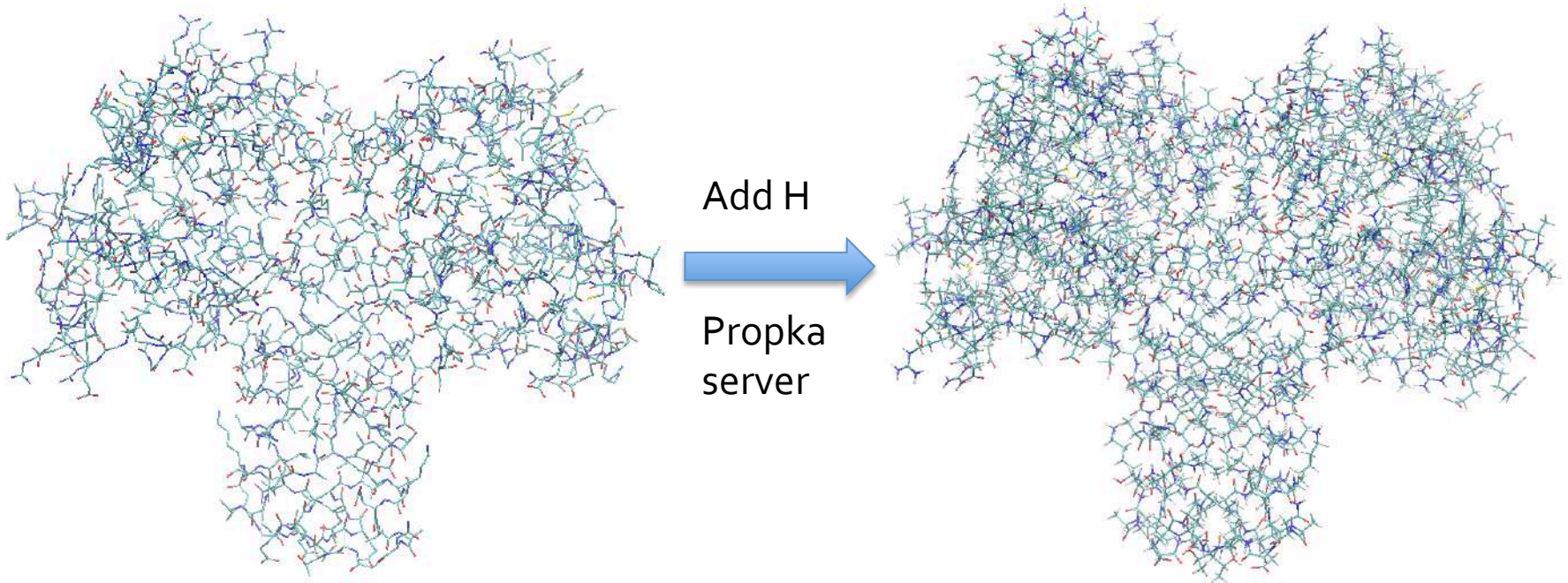
1. Build realistic atomistic model of the system
2. Simulate the behavior of your system over time using specific conditions (temperature, pressure, volume, etc):
3. Analyze the results obtained from MD



Steps in Molecular Dynamics Simulations

1. Build realistic atomistic model of the system

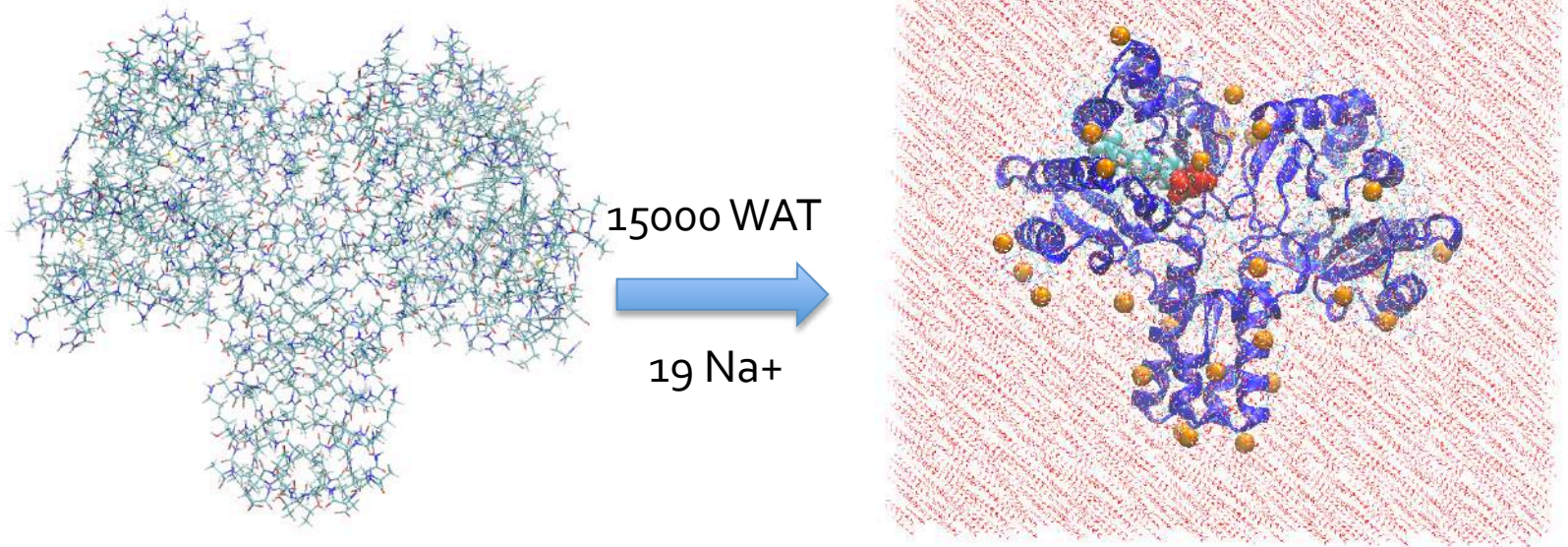
Step 1. PDB Protonation



Steps in Molecular Dynamics Simulations

1. Build realistic atomistic model of the system

Step 2. Solvate the system and add ions to neutralize systems



The solvation box is composed by **rigid water models** which rely on non-bonded interactions. The most popular models include **TIP₃P**, **TIP₄P**, **TIP₅P**, **SPC**, **SPC/E**. All these models are consistent with the **SHAKE algorithm** that constrains the hydrogen bonds to fixed lengths.

Steps in Molecular Dynamics Simulations

1. Build realistic atomistic model of the system

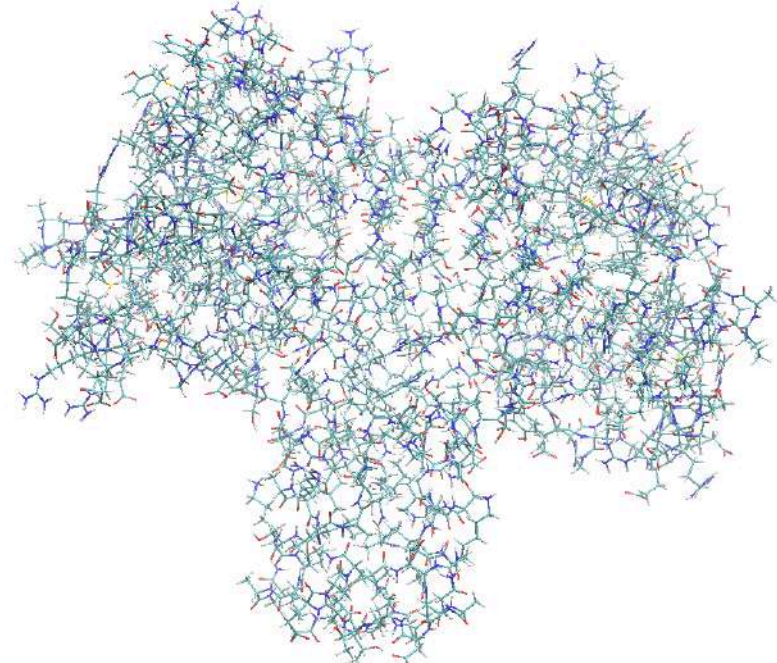
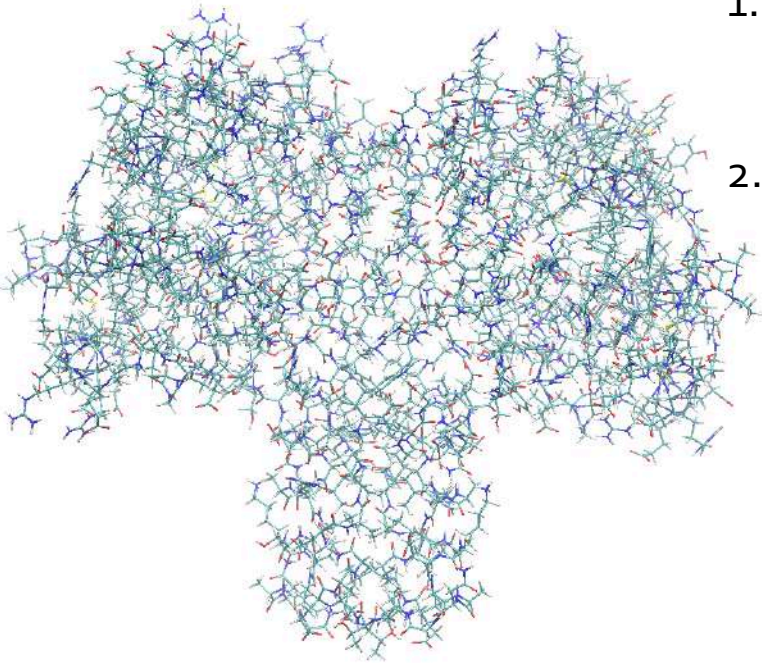
Step 3. Protein and Solvent Minimization

1. Solvent and counterion minimization
2. Whole system

2000 steps



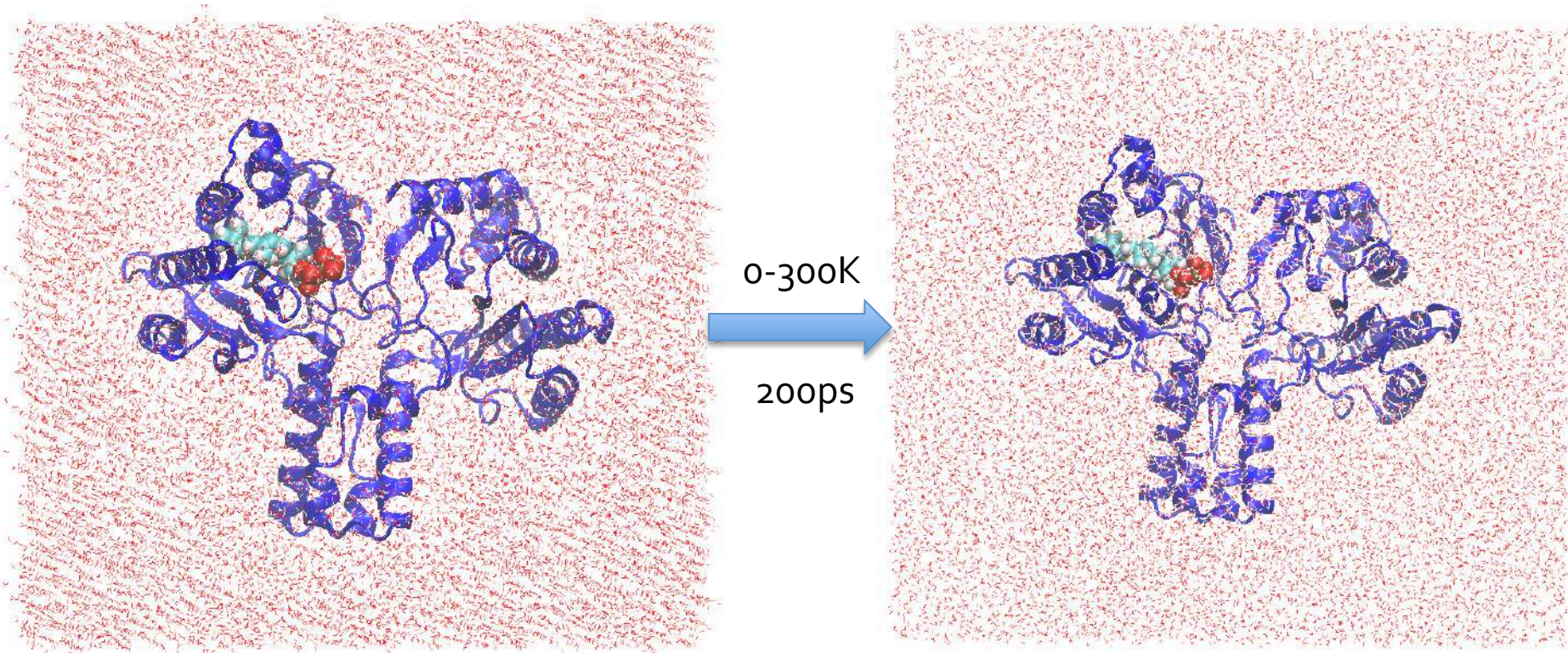
Minimization



Steps in Molecular Dynamics Simulations

1. Build realistic atomistic model of the system

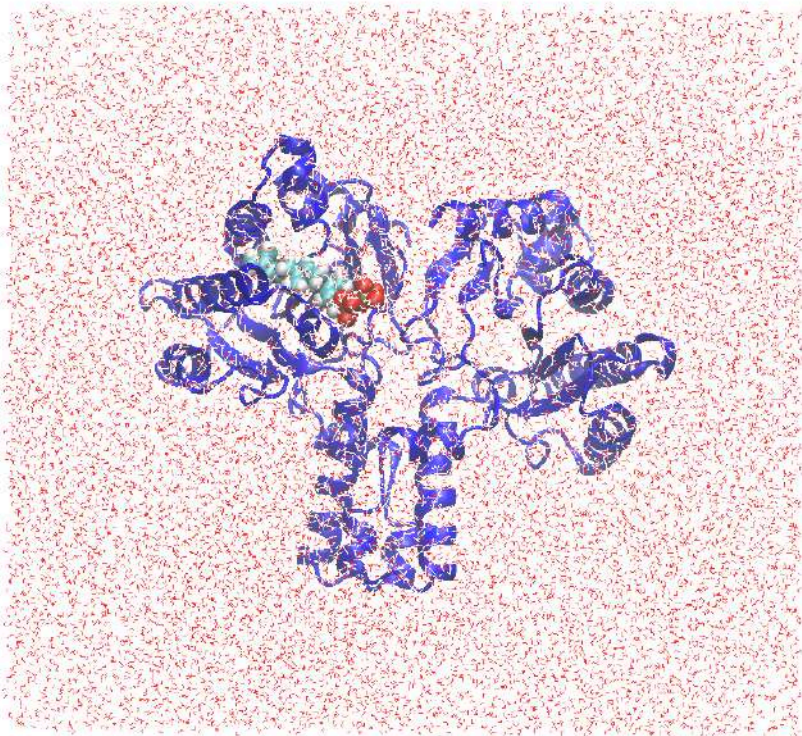
Step 4. Heating from 0K to 300K



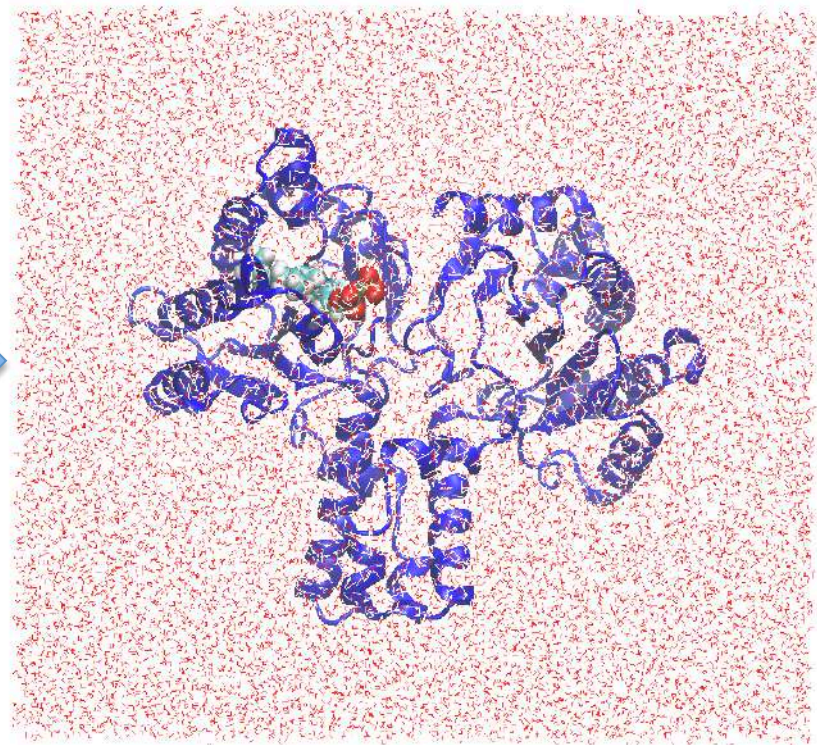
Steps in Molecular Dynamics Simulations

1. Build realistic atomistic model of the system

Step 5. Equilibration at Constant Pressure and Temperature



NPT
1ns



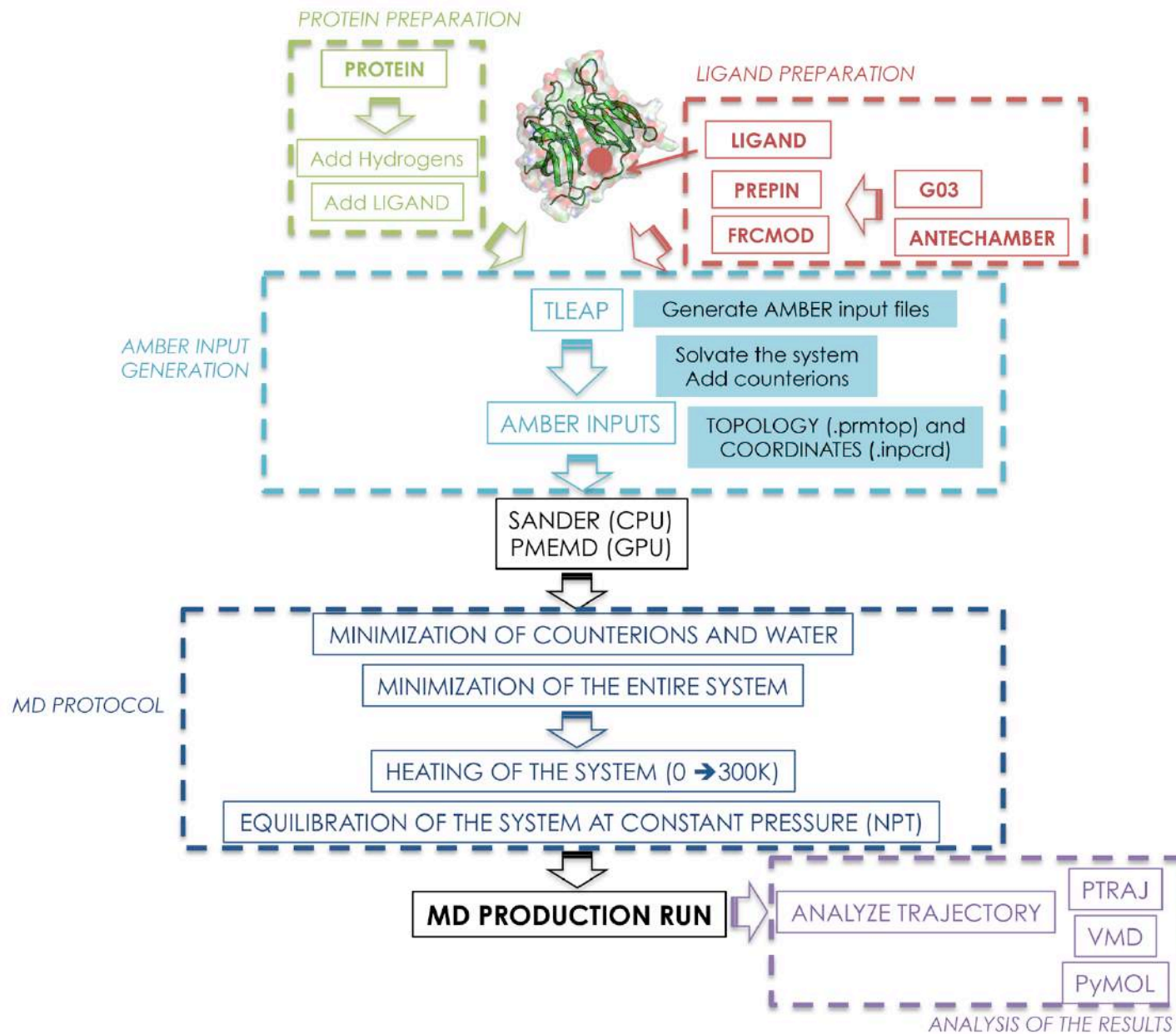
Steps in Molecular Dynamics Simulations

2. Simulate the behavior of your system over time using specific conditions (temperature, pressure, volume, etc): **MD PRODUCTION RUN**

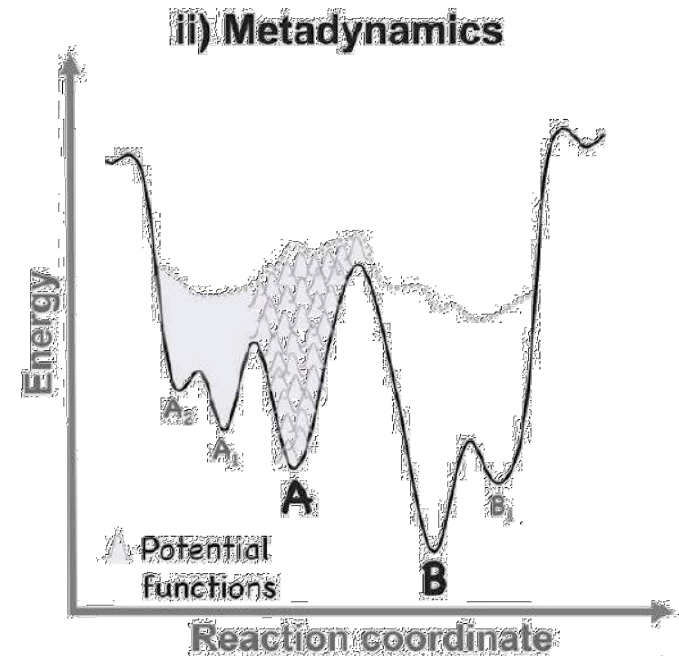
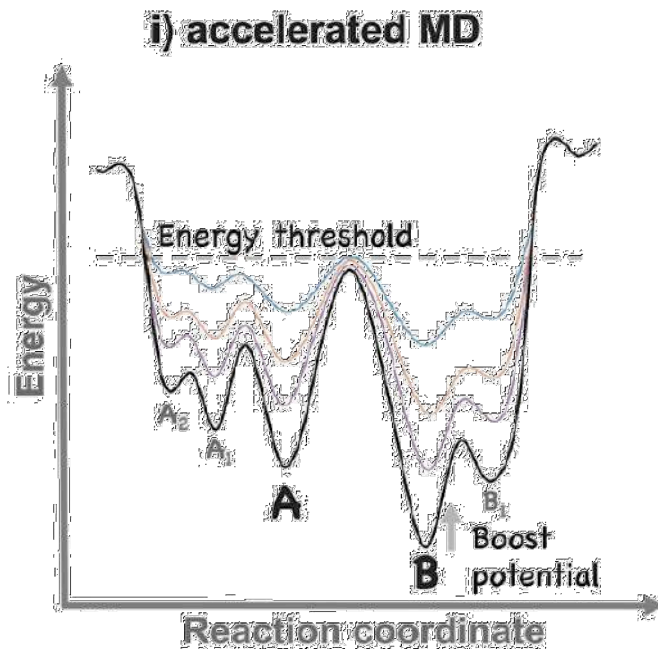
3. Analyze MD data

- Ramachandran plots
- Distance contacts maps
- RMS deviations
- Autocorrelation Functions
- Free energy contour maps
- Radius of gyration

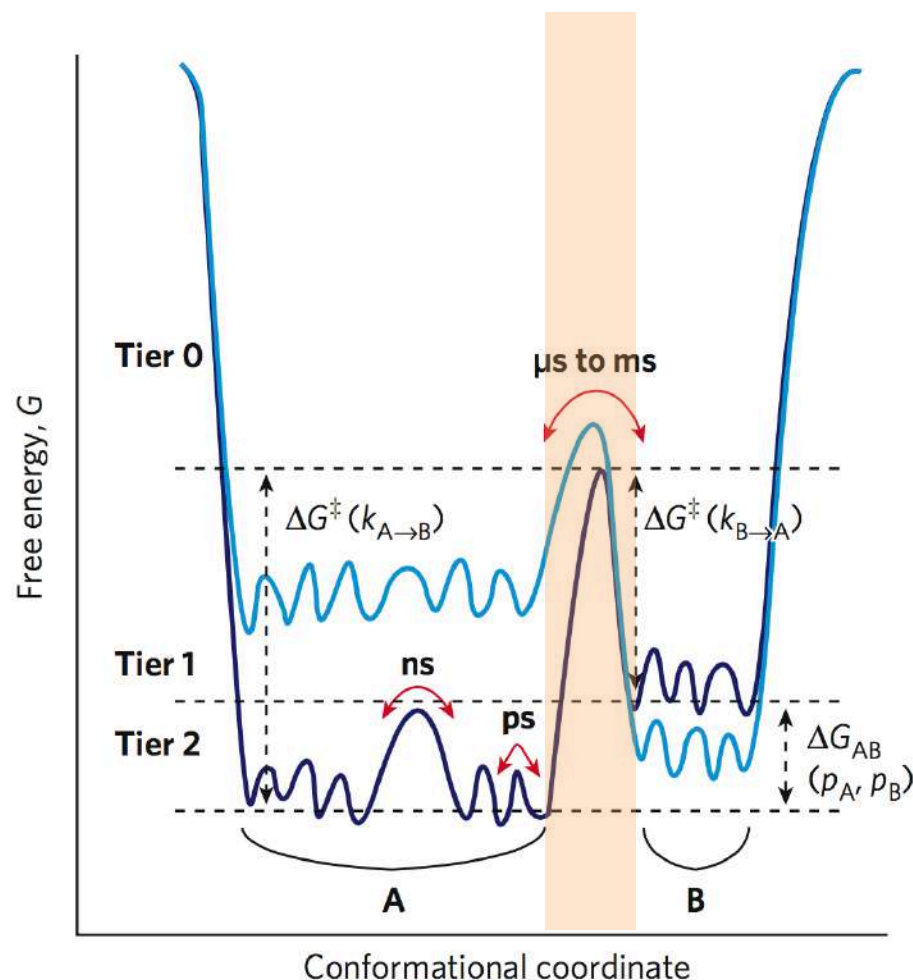
Molecular Mechanics and Molecular Dynamics Simulations



INTRODUCTION TO ENHANCED SAMPLING METHODS: For Biomolecules



Many biologically important processes such as enzyme catalysis, signal transduction, protein folding, ligand binding and allosteric regulation occur on the micro- to millisecond time-scale.



Limitations of Conventional Molecular Dynamics Simulations

MD simulations are computationally demanding for two reasons:

- 1) The force calculation at each time step requires substantial computation (one billion arithmetic operations for a system with one hundred thousand atoms).
- 2) The force calculation must be repeated many times

Individual steps are limited to a few femtoseconds by fast atomic vibrations, so simulating a millisecond of physical time requires nearly one trillion time steps.

On a single high-end processor core, such a simulation would take thousands of years to complete.

Limitations of Conventional Molecular Dynamics Simulations

Multiple CPUs

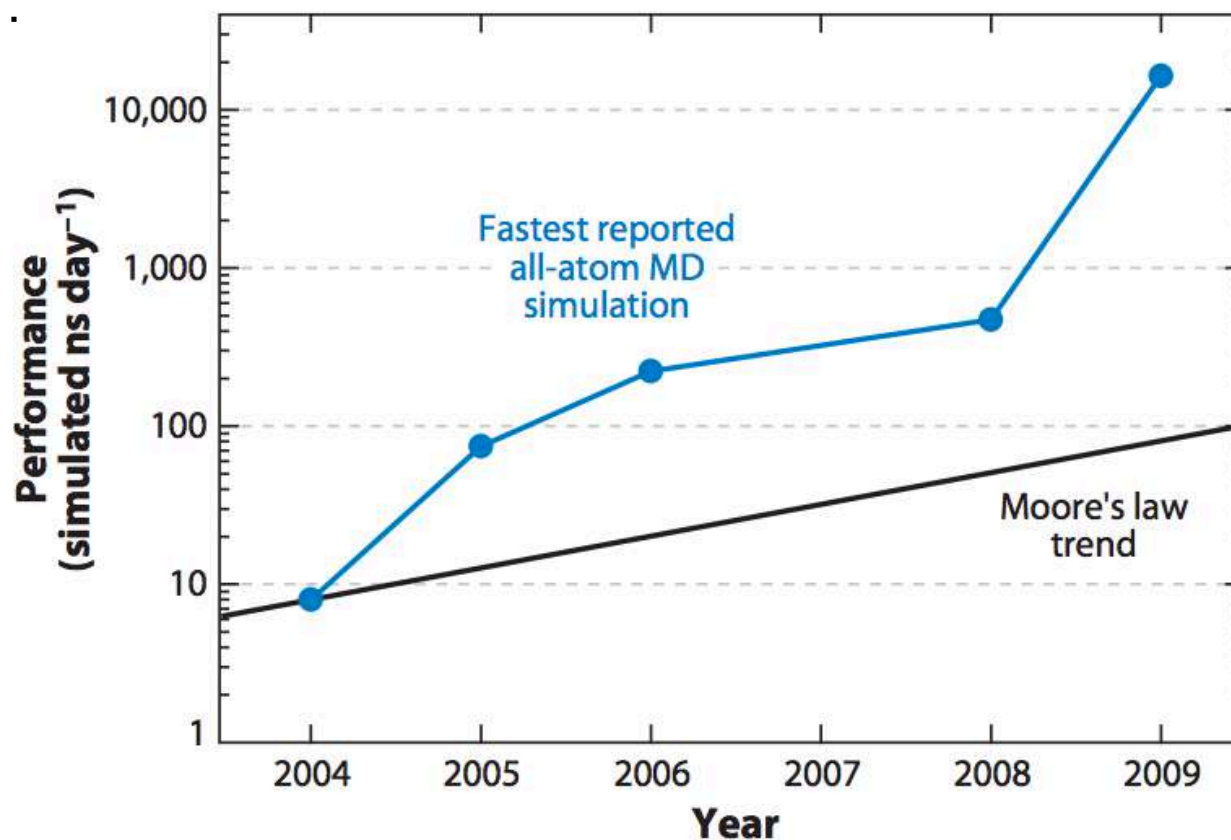
Parallelizing a force calculation across multiple processors requires that those processors communicate with one another, and the amount of communication required increases with the number of processors.

Beyond some point, adding more processors actually slows down the calculation.

The difficulty of parallelizing an MD simulation implies that the transistor density improvements predicted by Moore's law do not automatically lead to improved MD performance

Limitations of Conventional Molecular Dynamics Simulations

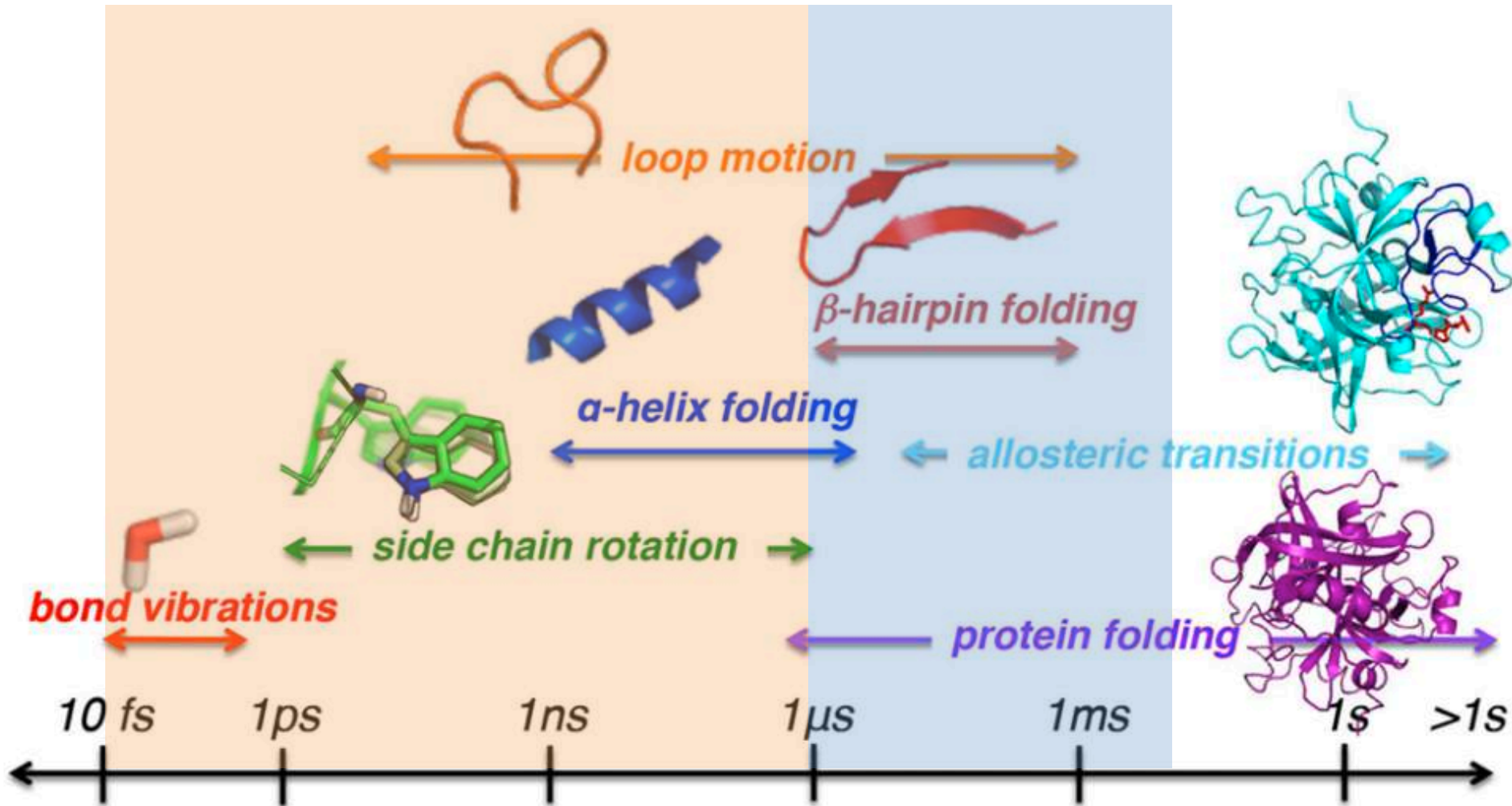
However...



These improvements are attributable to a variety of hardware, software, and algorithm innovations

CONVENTIONAL MD

**SPECIALIZED SUPERCOMPUTERS/
ENHANCED SAMPLING TECHNIQUES**



Recent Advances in Simulation Methodology

Recent advances in hardware, software, and algorithms have increased the timescales accessible to simulation by several orders of magnitude.

1. Parallelization across general-purpose computer chips

The widely used MD codes NAMD, GROMACS, and AMBER have all substantially improved their parallel performance in recent years. These packages can now deliver performance of over 100 ns/day on commodity computer clusters, with the number of processors required to achieve this performance **scaling roughly linearly with the number of atoms in the system.**

Enhancing conformational sampling with multiple-CPU

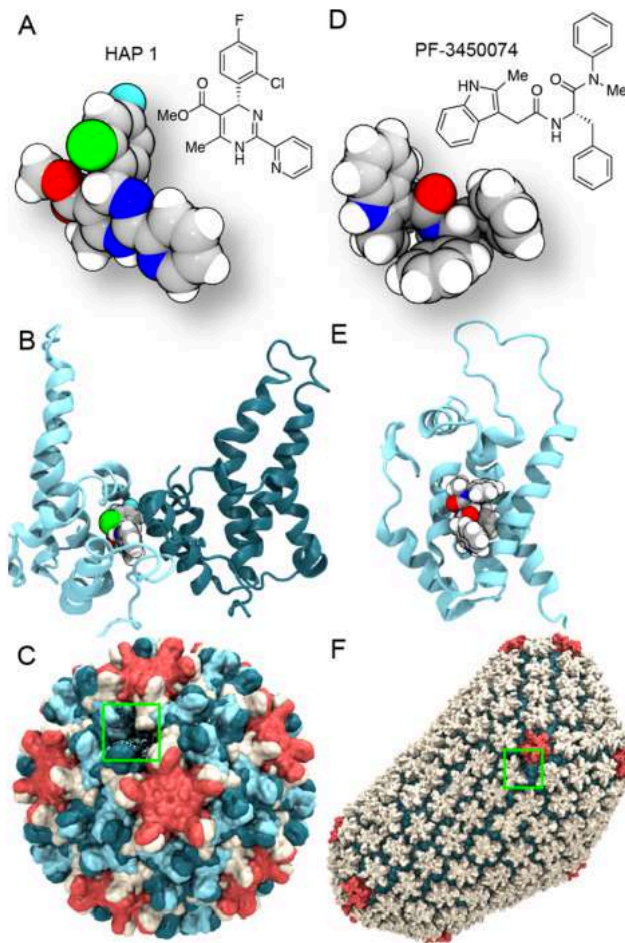
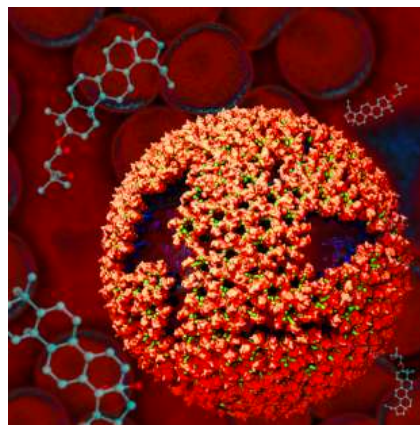
All-Atom Molecular Dynamics of Virus Capsids as Drug Targets

CHARMM36

NAMD

NSF Blue Waters Supercomputer

Simulations totaling **100 ns** were performed for both the apo-capsid and HAP1-bound capsids (~6 million atoms).



Recent Advances in Simulation Methodology

2. Graphical processing units

Originally designed specifically to accelerate the rendering of three dimensional graphics, graphics processing units (GPUs) have become increasingly popular for general-purpose scientific computation thanks to their ability to perform large numbers of identical computations in parallel on a single chip.

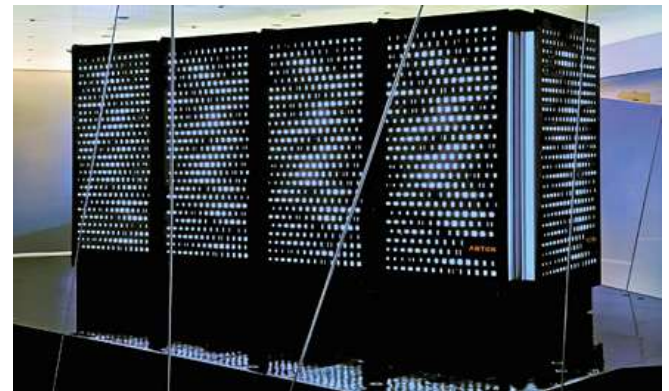


GPUs offer an excellent price-to-performance ratio, however, enabling reasonably fast simulations at a cost substantially lower than that for a cluster of general-purpose processors.

Recent Advances in Simulation Methodology

3. Special Purpose Parallel Architecture

By far the greatest recent speedups have been achieved through the use of special-purpose chips, in combination with new parallelization algorithms and software

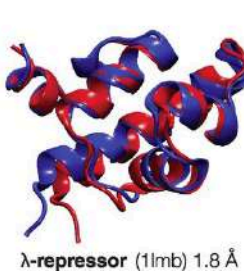
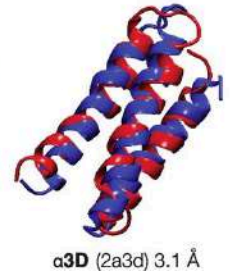
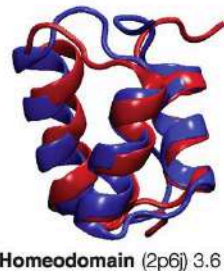
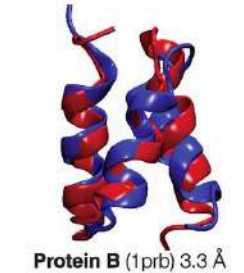
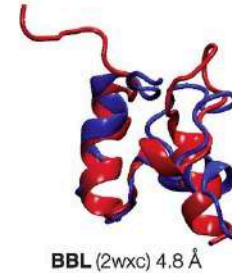
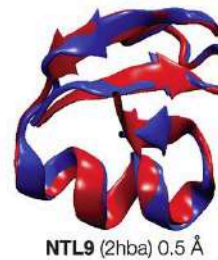
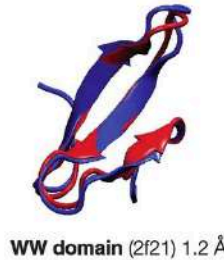
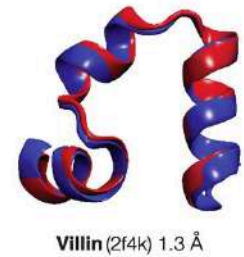
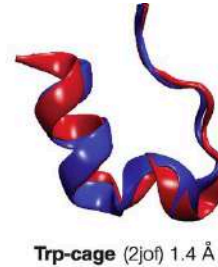
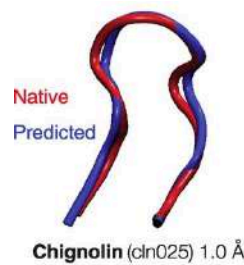
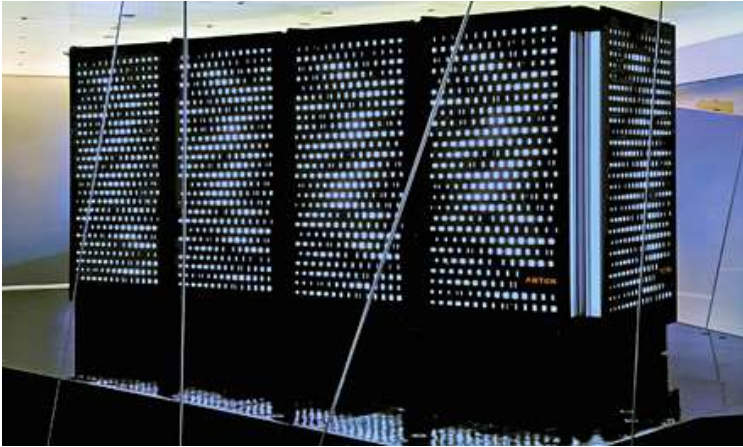


Anton is able to achieve this speed because it is designed specifically for MD simulations. The entire computation is performed on special-purpose chips developed at D. E. Shaw Research, which are directly connected to one another in a custom network

cMD specialized supercomputer. Protein Folding

How Fast-Folding Proteins Fold

Kresten Lindorff-Larsen,^{1,†} Stefano Piana,^{1,†} Ron O. Dror,¹ David E. Shaw^{1,2,†}



~8 ms of simulation, containing more than 400 folding or unfolding events.

Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. Science , 334, 517-520 (2011)

Recent Advances in Simulation Methodology

4. Distributed Computing

In 2000, the group of Pande at Stanford launched the first Distributed Computing (DC) project designed to perform computationally intensive simulations of bio-molecular systems. The so-called ‘folding@home’ network allows people throughout the world to download and run molecular dynamics software on their individual machines.

Using recently developed algorithms based on the theory of exponential kinetics and **Markov-state** models, **short MD** simulations performed on each individual machine are banded together to reconstruct the extended free energy surface of the system of interest.

Markov State Models. Basics.

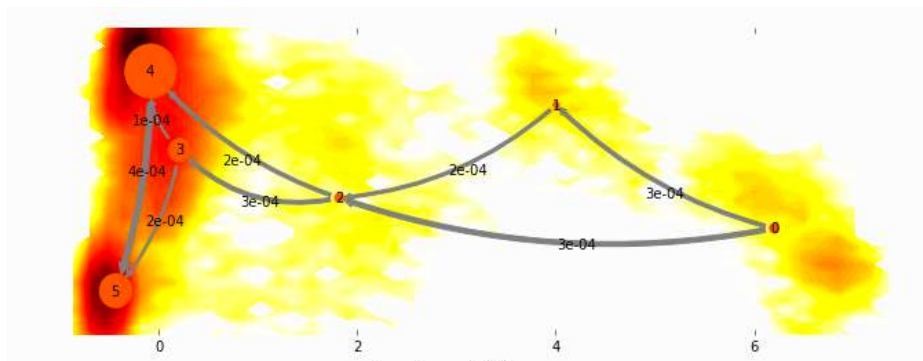
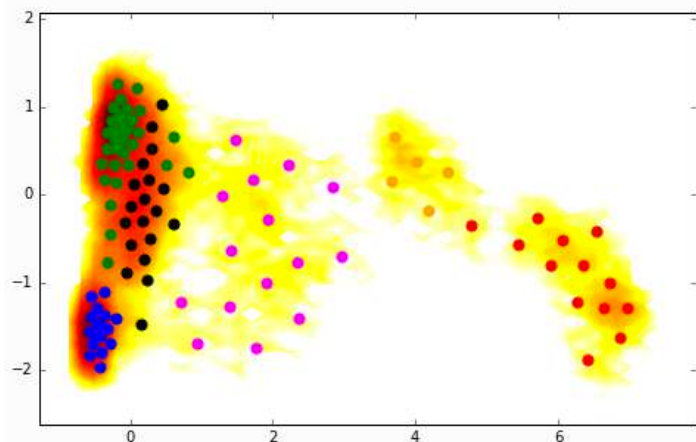
Even with expensive special-purpose hardware such as Anton, simulation trajectories can barely reach biomolecular timescales of typical interest, let alone exceed to permit their statistical characterization beyond simple anecdotal observation.

MSMs can predict both stationary and kinetic quantities on long timescales (e.g. milliseconds) using a set of atomistic molecular dynamics simulations that are individually much shorter, thus addressing the well-known sampling problem in molecular dynamics simulation.

Markov State Models. Basics.

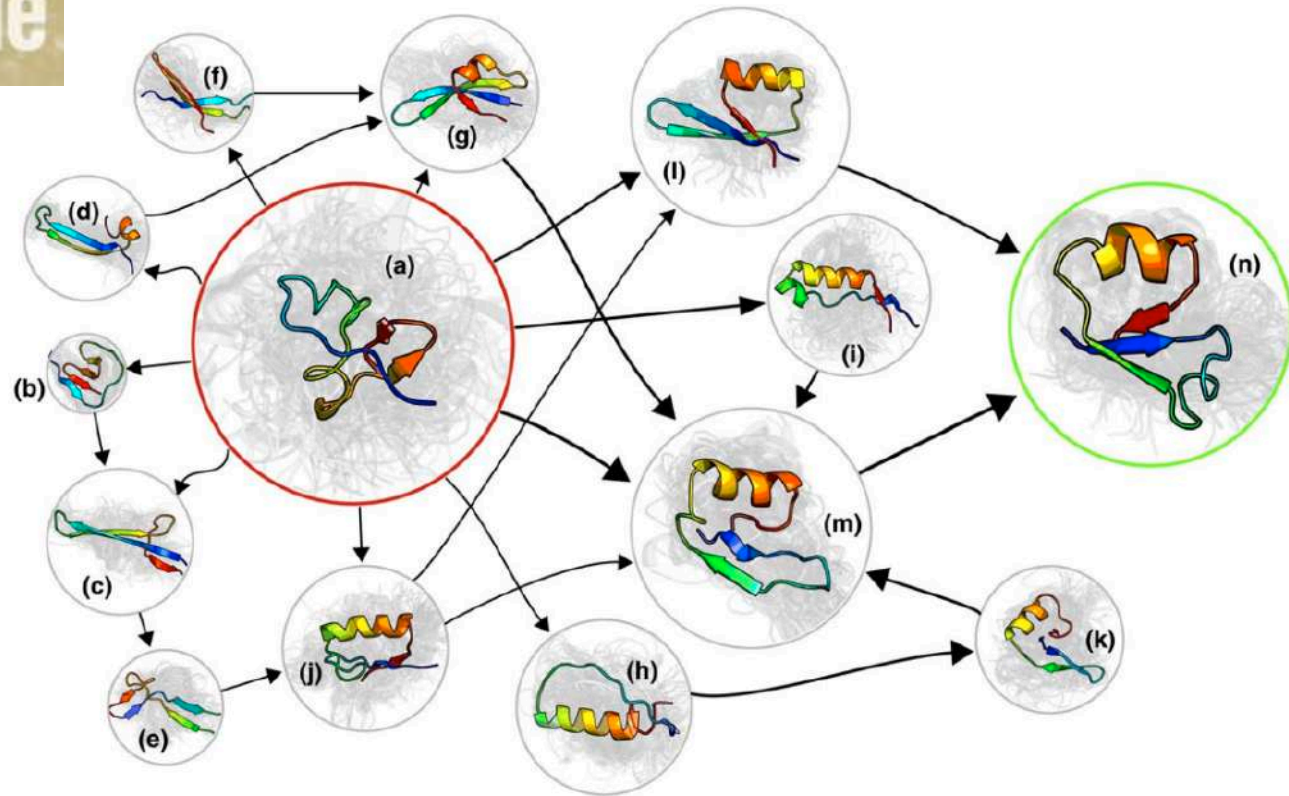
These Markov state models (MSMs) describe the stochastic dynamics of a biomolecular system using two objects:

1. A **discretization** of the high-dimensional molecular state space, usually into n disjoint conformational sets S_1, \dots, S_n
2. A model of the stochastic **transitions** between these discrete states, usually described by a matrix of conditional **transition probabilities** estimated from the simulation trajectories–



Markov State Models. Protein Folding.

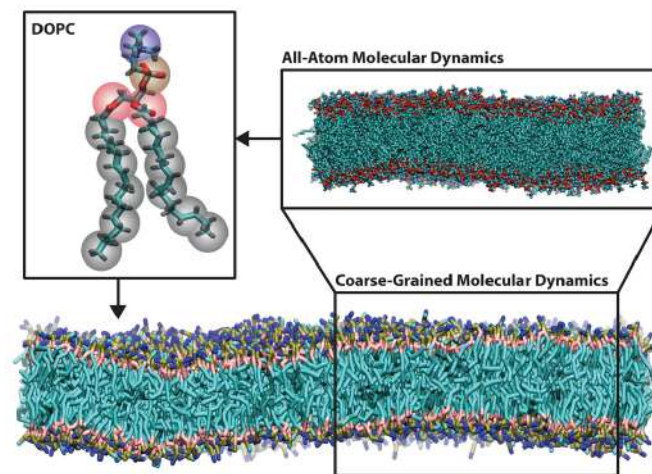
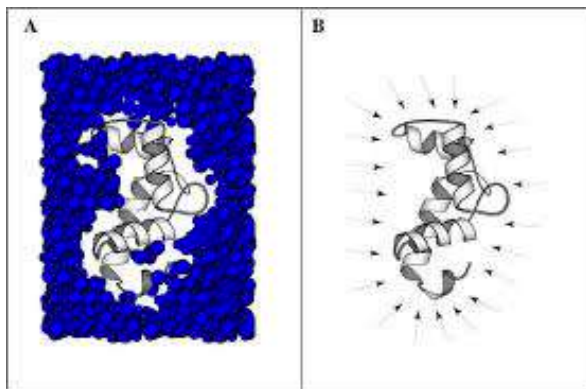
Folding@home



Recent Advances in Simulation Methodology

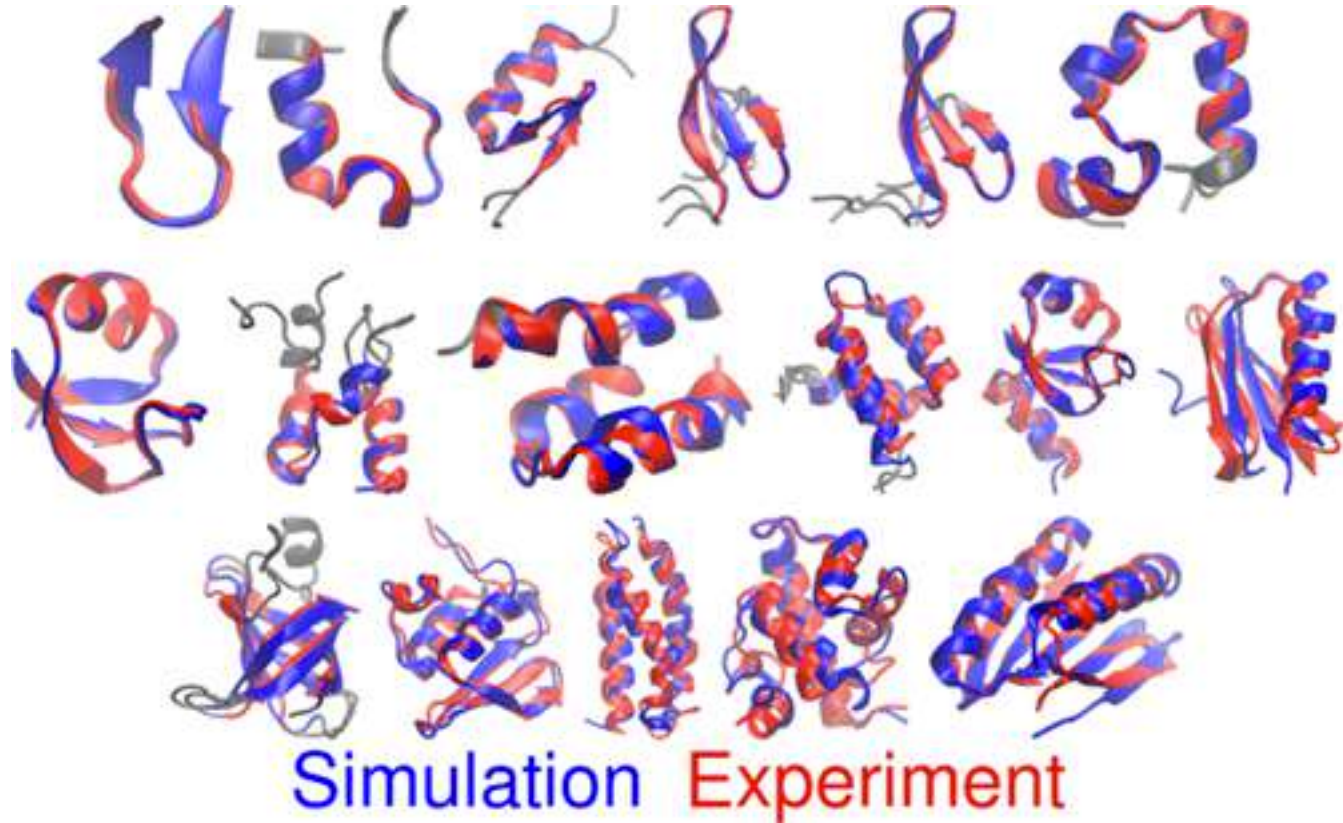
5. Coarse Graining and Implicit Solvent

One can often substantially accelerate simulations at the cost of reduced accuracy by employing simplified system representations, such as **coarse-grained models**, in which each simulated particle represents several atoms, or **implicit solvent models** in which water atoms are replaced by a continuum representation. Models of both types have seen substantial development in recent years.



Speeding up MD simulations with Implicit Solvent

Folding Simulations in Implicit Solvent



Implicit Solvent 1 μ s/day

Recent Advances in Simulation Methodology

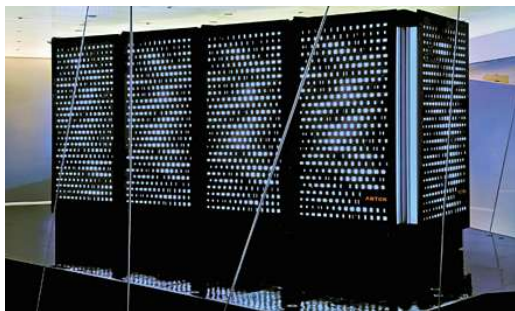
6. Enhanced Sampling

In general, these methods can be divided into two groups:

1. The first class of these methods is based on **biasing collective variables (CVs)** of biomolecules. They include:
 - **Umbrella sampling**, blue moon sampling, adaptive biasing force, **metadynamics**, conformational flooding, string method, λ -dynamics, adiabatic free energy dynamics, temperature accelerated MD and **targeted MD**.

Molecular Mechanics and Molecular Dynamics Simulations

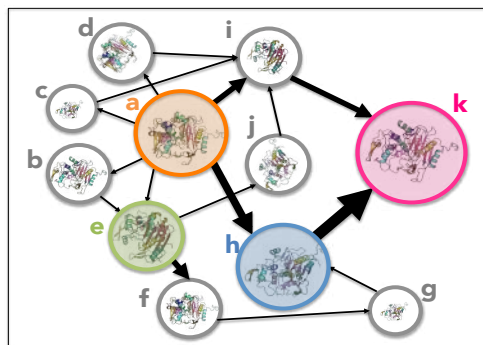
Specialized-SuperComputers



Long Simulations Identification of infrequent events

- Rare events are rarely sampled more than one time

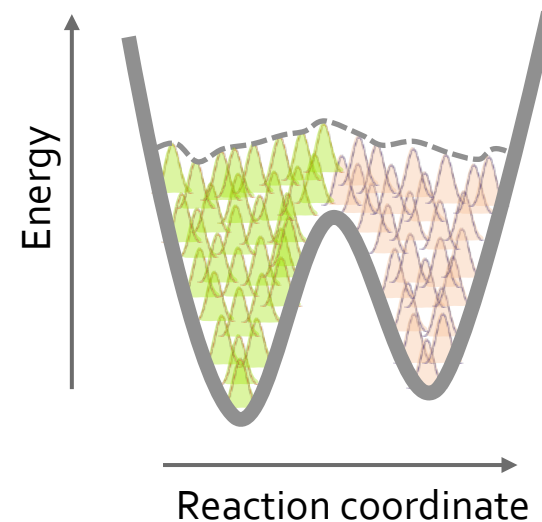
Markov-State Models



Multiple (short) Simulations

- All events are equally sampled
- Thermodynamic and kinetic properties
- High computational cost required

Biasing CVs: Metadynamics



  Gaussian potential functions

- Depends on the a priori definition of a reaction coordinate. Sometimes it is difficult to define.

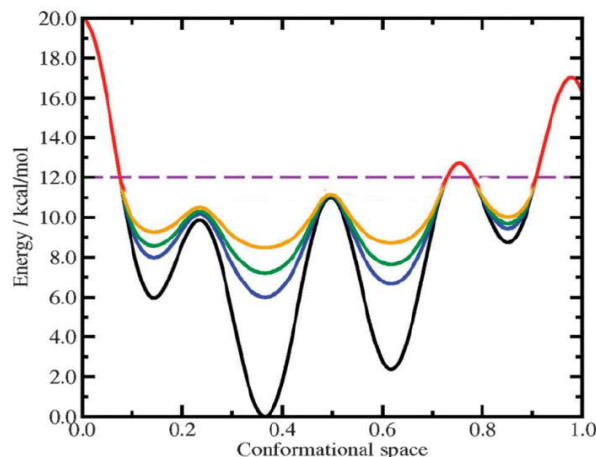
Recent Advances in Simulation Methodology

6. Enhanced Sampling

In general, these methods can be divided into two groups:

2. The second class of enhanced sampling methods that do not require predefined system-dependent CVs such as:
 - **Replica exchange (RE), or parallel tempering (PT), self-guided molecular/ Langevin dynamics (SGMD and SGLD), essential energy space random walk (EESRW) and accelerated molecular dynamics (aMD).**

Unconstrained Methods: Accelerated Molecular Dynamics



- Unconstrained conformational sampling. Identification of relevant conformational states.
- Difficult to reweight aMD simulations for large systems

- Based on Voter hyperdynamics

VOLUME 78, NUMBER 20

PHYSICAL REVIEW LETTERS

19 MAY 1997

Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events

Arthur F. Voter

Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 4 February 1997)

- Extended to biological processes by Hamelberg and McCammon

JOURNAL OF CHEMICAL PHYSICS

VOLUME 120, NUMBER 24

22 JUNE 2004

Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules

Donald Hamelberg^{a)}

NSF Center for Theoretical Biological Physics and Department of Chemistry and Biochemistry, University of California–San Diego, La Jolla, California 92093-0365

John Mongan

NSF Center for Theoretical Biological Physics and Bioinformatics Program, University of California–San Diego, La Jolla, California 92093-0365

J. Andrew McCammon

Howard Hughes Medical Institute, NSF Center for Theoretical Biological Physics, Department of Chemistry and Biochemistry and Department of Pharmacology, University of California–San Diego, La Jolla, California 92093-0365

D. Hamelberg, D. Mongan, J. A. McCammon *J. Chem. Phys.* 2004, 120, 11919

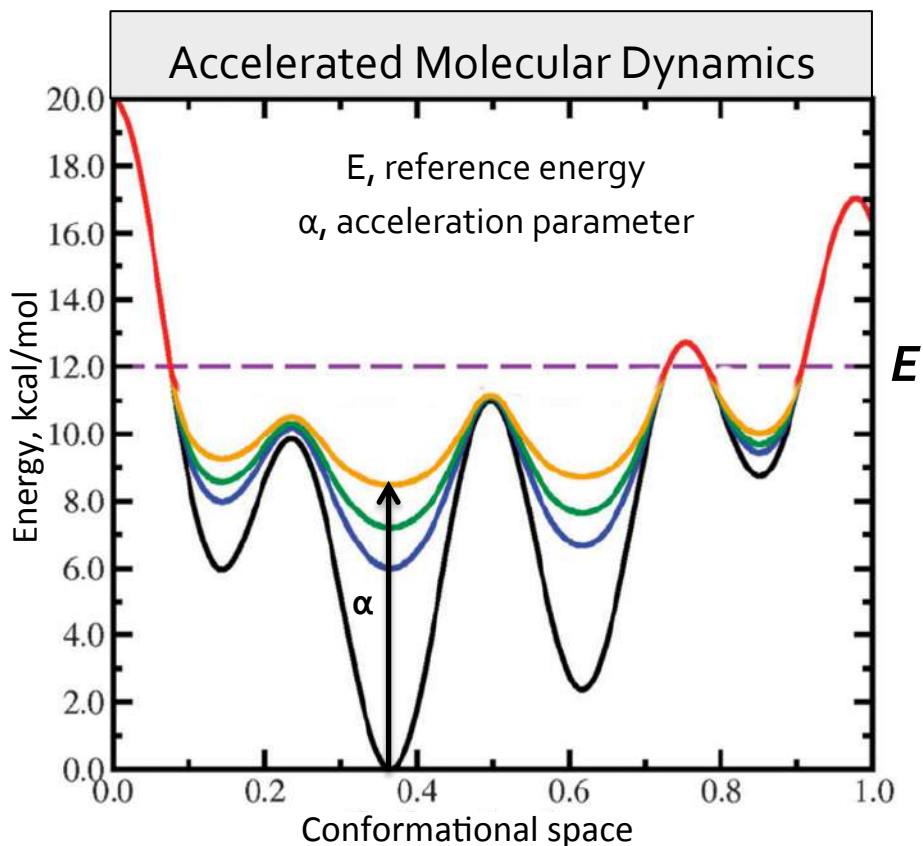
D. Hamelberg, C. A. F. de Oliveira, J. A. McCammon *J. Chem. Phys.* 2007, 127, 155102

Molecular Mechanics and Molecular Dynamics Simulations

ΔV_{tot} 'total-boost'

$$V = \sum_{\text{bonds}} k_b (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \underbrace{\sum_{\text{dihedrals } n=1}^N \sum K_\phi^{(n)} [1 + \cos(n\phi - \delta)]}_{\Delta V_{\text{dih}} \text{ 'dihedral-boost'}} + \sum_{i,j}^{\text{vdW}} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i,j}^{\text{elec}} \left(\frac{q_i q_j}{D r_{ij}} \right)$$

ΔV_{dih} 'dihedral-boost'



$$V^*(\vec{r}) = V(\vec{r}), V(\vec{r}) \geq E,$$

$$V^*(\vec{r}) = V(\vec{r}) + \Delta V(\vec{r}), V(\vec{r}) < E$$

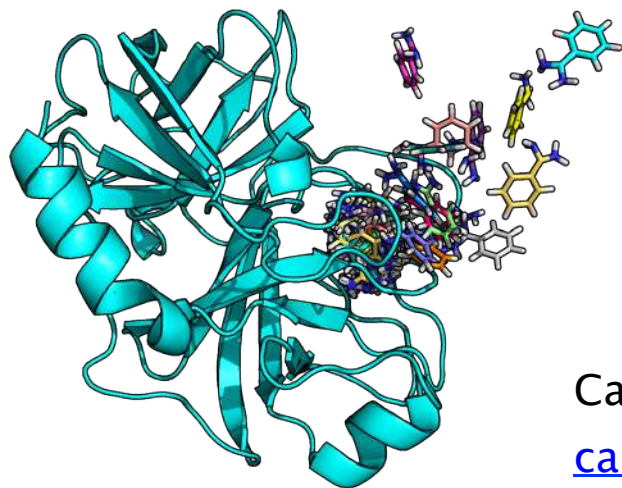
Boost potential

$$\Delta V(\vec{r}) = \frac{(E - V(\vec{r}))^2}{\alpha - E - V(\vec{r})}$$

'Dual-Boost' $\Delta V(\vec{r}) = \Delta V_{\text{dih}}(\vec{r}) + \Delta V_{\text{tot}}(\vec{r})$

Higher Lower

MOLECULAR MECHANICS AND MOLECULAR DYNAMICS SIMULATIONS



Carla Calvó-Tusell
carla.calvoa@udg.edu

Ferran Feixas
ferran.feixas@udg.edu

Institut de Química Computacional i Catàlisi (IQCC)
Universitat de Girona, Catalonia, Spain