Computational Biology
(BIOSC 1540)
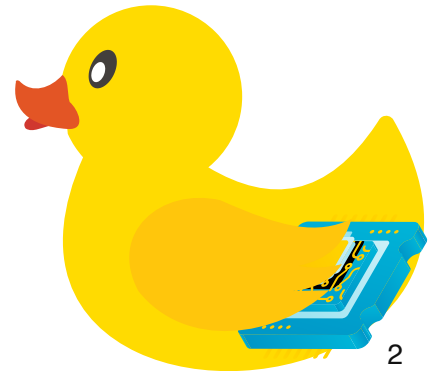
Lecture 13:
Molecular simulation principles

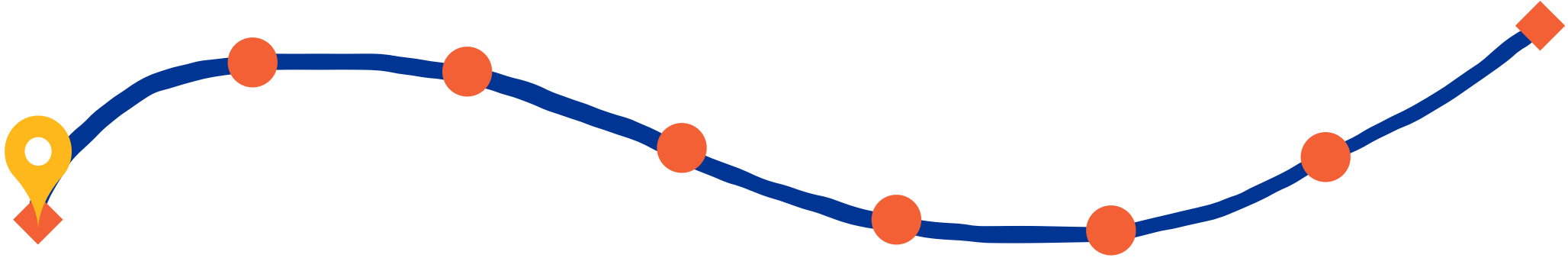Oct 17, 2024

University of Pittsburgh
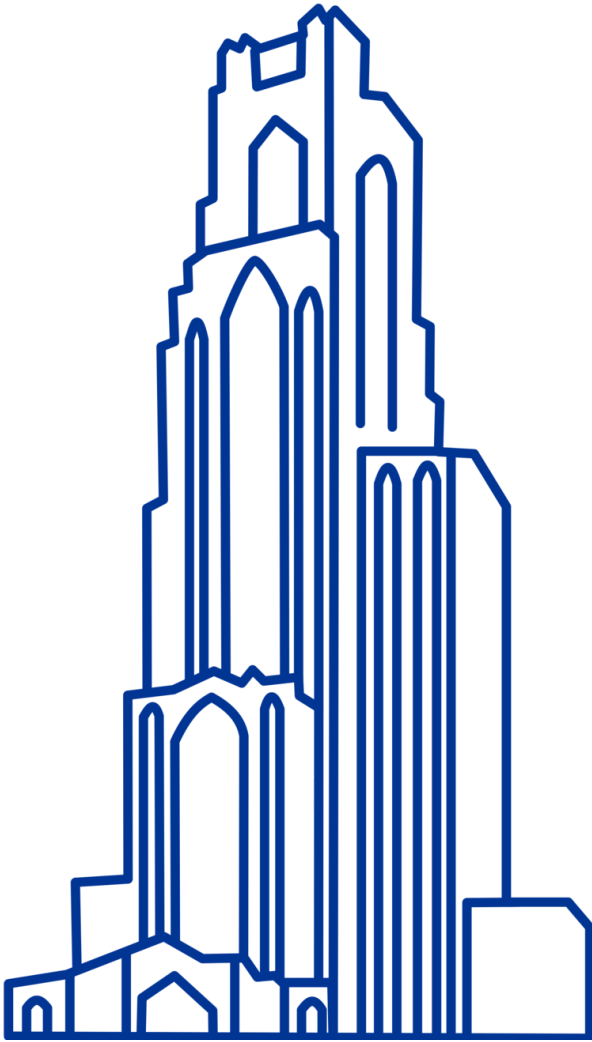
Announcements

- A05 is due Oct 24 at 11:59 pm

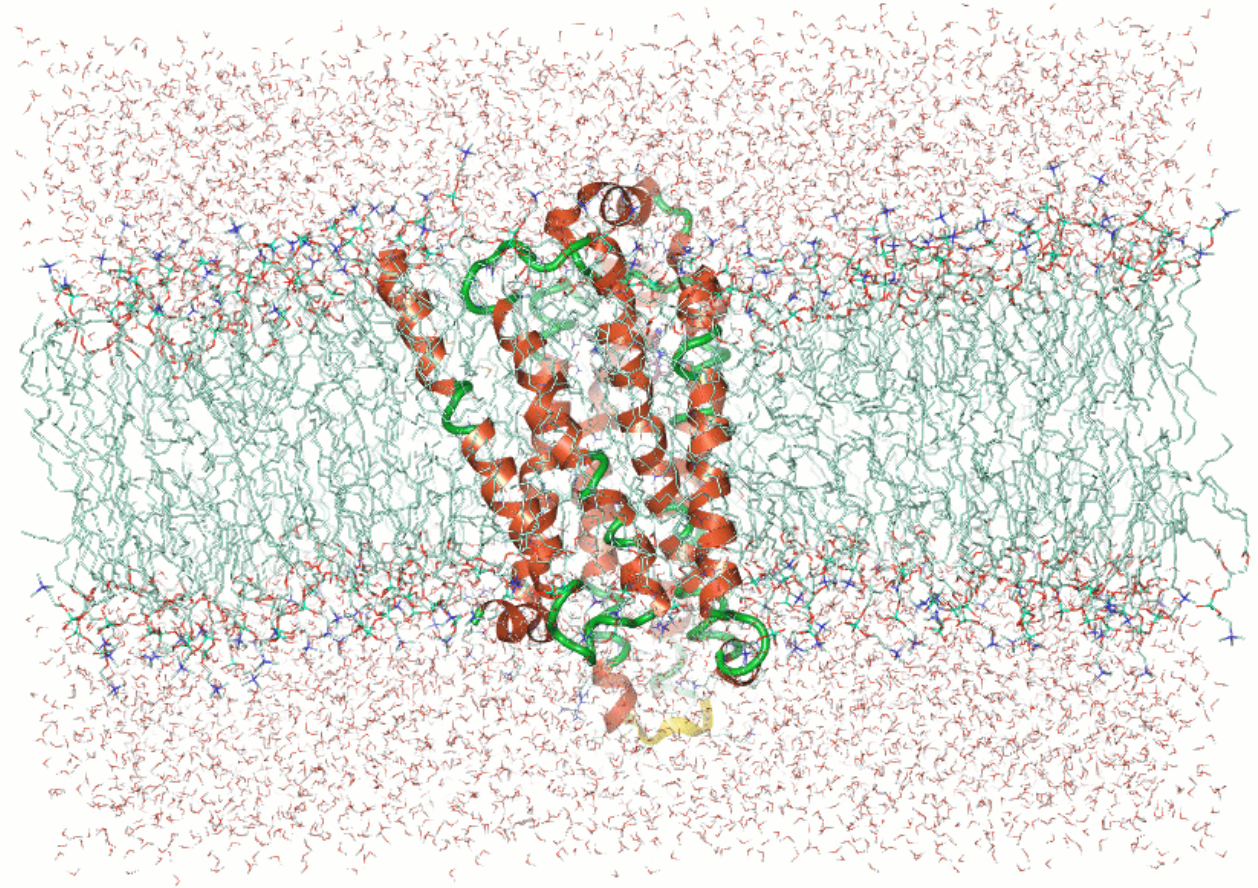# After today, you should be able to

Understand the importance of molecular dynamics (MD) simulations for proteins.

Proteins undergo movements like folding, unfolding, and domain motions

These motions are essential for binding, catalysis, and signal transduction
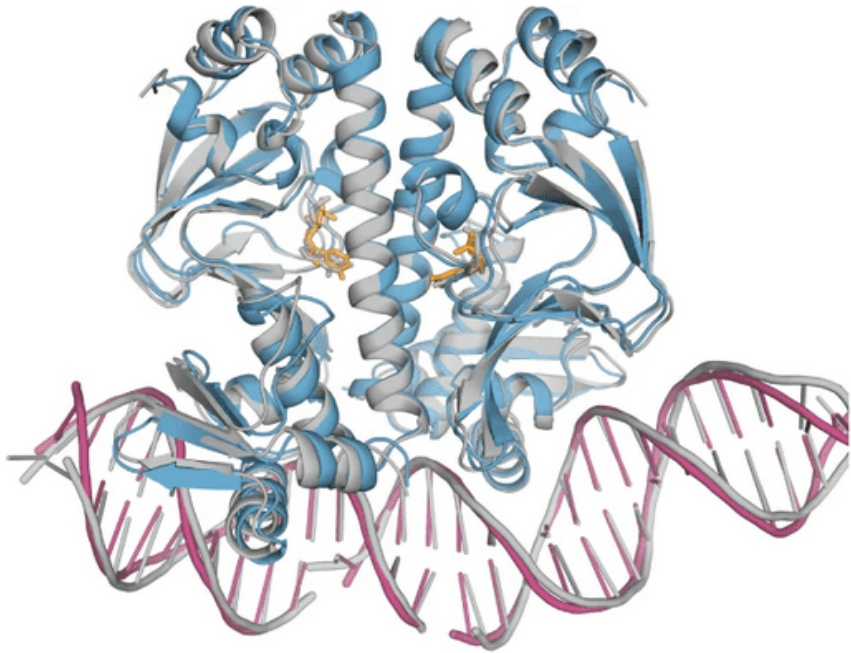
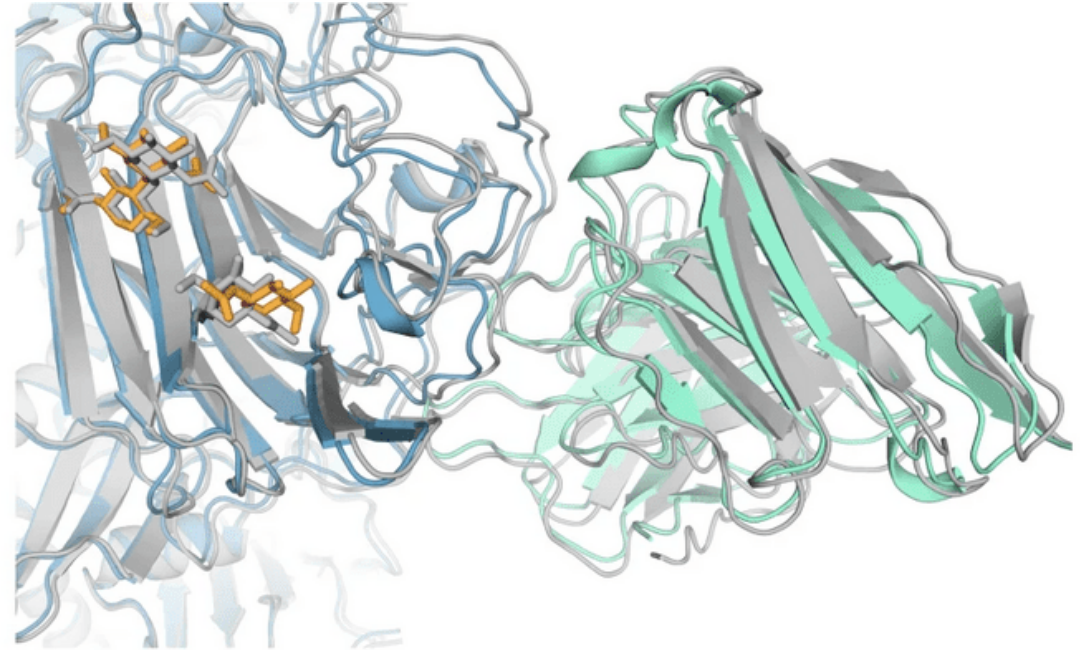Understanding dynamics is crucial for drug design, protein design, biotech, etc.

Protein structure determination and prediction provide fixed snapshots

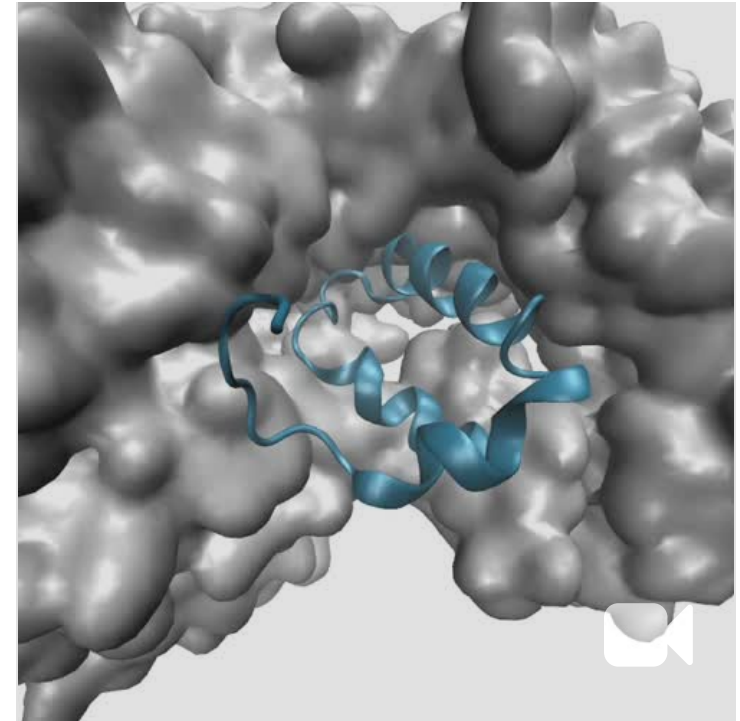Do not capture the full range of functional conformations

Molecular dynamics (MD) provide time-resolved insights into protein behavior

Simulation of Atomic Movements

- MD computes trajectories of atoms over time scales of femtoseconds to microseconds.
- It can capture both small-scale vibrations and large-scale conformational changes.

Visualization and Analysis:

- Provides detailed information on atomic interactions and energy changes.
- Enables the study of mechanisms at an atomic level.

MD simulations provide more realistic analysis of proteins

Refinement of Predicted Structures:

- MD helps minimize energy and relax structures obtained from modeling.
- Improves accuracy by accounting for environmental effects.

Studying Intrinsically Disordered Proteins:

- MD captures the flexible nature of disordered regions.
- Aids in understanding functions that depend on disorder.

Folding and Misfolding Pathways:

- Simulates the folding process to identify intermediates.
- Investigates misfolding mechanisms relevant to diseases.

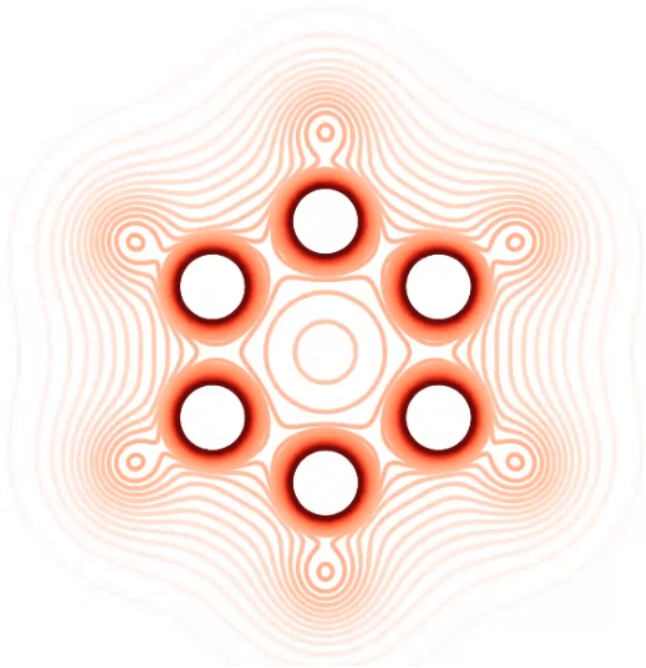# After today, you should be able to
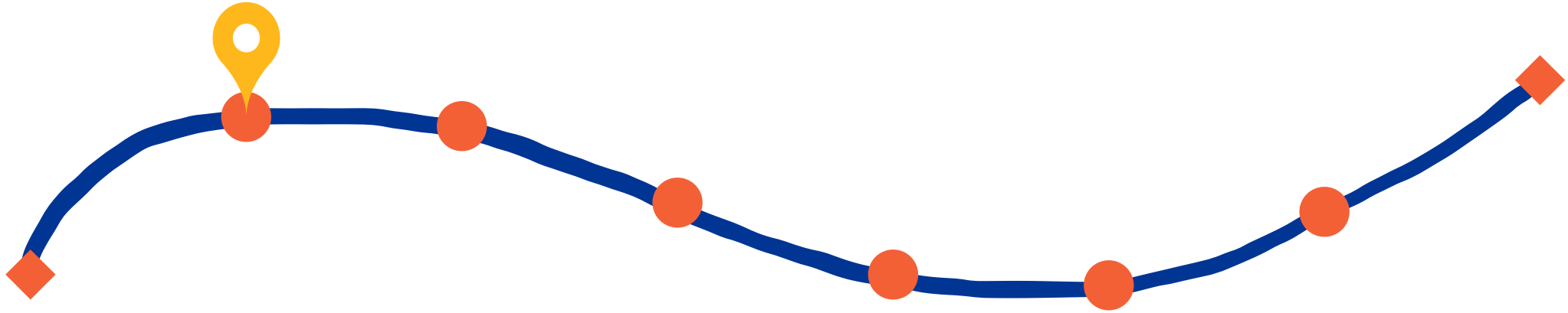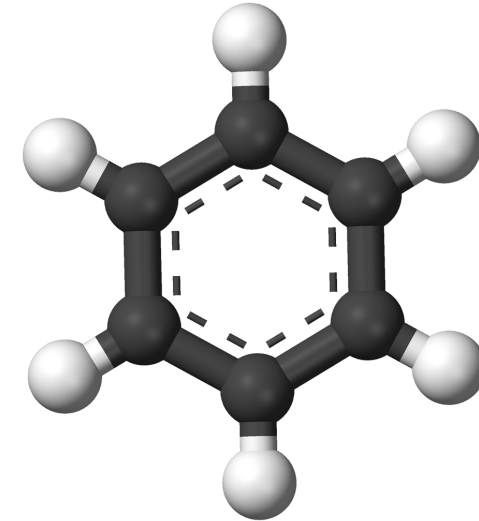
Identify the validity of the classical approximation

In biomolecular MD simulations, **atoms are treated as classical particles**

We treat atoms as hard spheres

Quantum mechanics

Classical mechanics

Classical approximations neglect quantum effects

Classical Mechanics

- Describes the motion of macroscopic objects.
- Assumes particles have well-defined positions and velocities.
- Governed by Newton's Laws of Motion.

Quantum Mechanics

- Necessary for describing behavior at atomic and subatomic scales.
- Accounts for wave-particle duality, uncertainty principle, proton tunneling
- Electrons exhibit quantum behavior that cannot be captured classically.

Classical approximation impacts

Nuclei

- Nuclei (protons and neutrons) are much heavier than electrons.
- Their de Broglie wavelengths are very small, making quantum effects less significant.
- At room temperature and above, thermal energies dominate over quantum zero-point energies.

Electrons

- Electrons are not explicitly simulated in classical MD.
- Their effects are included implicitly through potential energy functions (force fields).
- The electronic structure is assumed to remain in the ground state during simulation.

# Classical approximations are valid in most biomolecular cases

Suitable Systems:

- Biological macromolecules (proteins, nucleic acids, lipids).
- Materials where electronic excitations are not critical.
- Processes where bond breaking/forming does not occur.

Limitations:

- It cannot accurately simulate chemical reactions involving electronic transitions.
- Quantum phenomena like tunneling and zero-point energy are not captured.

After today, you should be able to
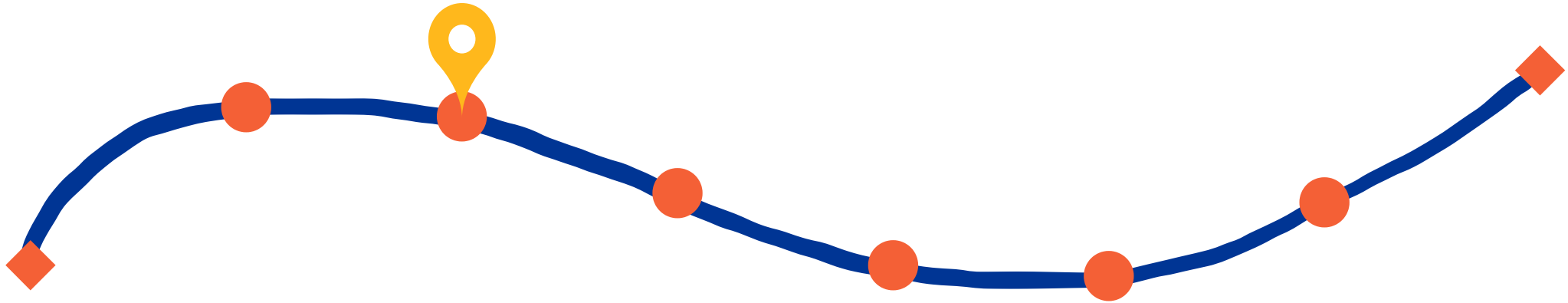
Discuss the concept of equations of

motion in MD simulations.

Classical particles follow Newton's equations of motion

Newton's Second Law

The acceleration of an object is directly proportional to the net force acting on it and inversely proportional to its mass

$$\vec{a} = \frac{\vec{F}}{m}$$

Force vector acting on the particle [kcal/mol/Å]

$\vec{F}$

Mass of the particle [amu]

m

Given atomic forces, we can calculate atomic movements

Acceleration vector of the particle [Å/fs^2]

$\vec{a}$

How can we compute atomic forces?

Forces are the negative gradients of potential energy

The potential energy, U, is dependent on positions of all atoms

Forces are obtained from the negative gradient of potential energy

$$\overrightarrow{F} = -\nabla U(\overrightarrow{r})$$

$$\nabla = \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z}$$

Determines acceleration and thus motion of atoms

Time evolution of the system is computed by integrating equations of motion

Continuous motion approximated using discrete time steps

Δt

- Determine forces
- Move a small amount forward in time
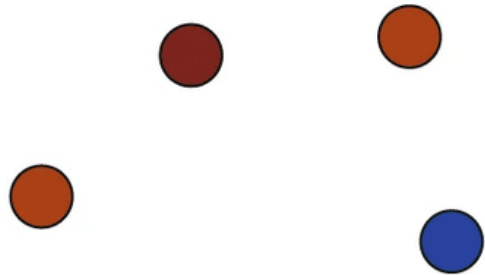- Repeat

Time step length determines how "smooth" the animation/trajectory
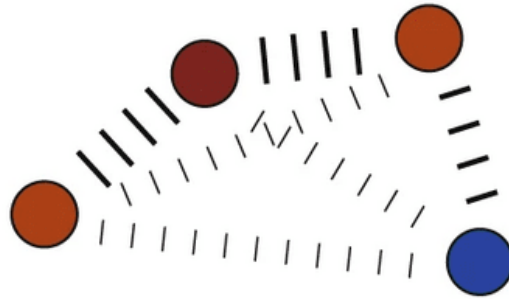

Think claymation

Molecular simulations compute an atomistic trajectory
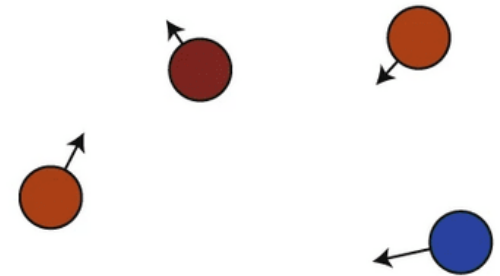
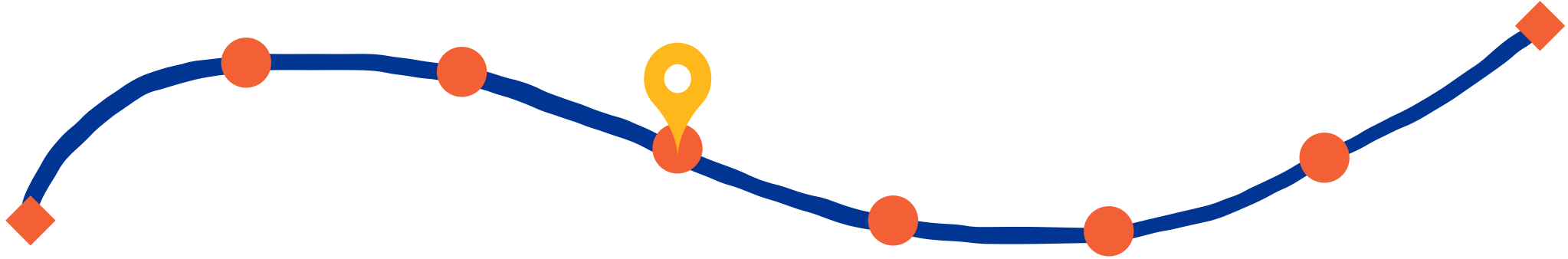3D coordinates of atoms in our system

These atoms exert forces on each other

Using Newton's equation of motion, we can predict their movement

$$\overrightarrow{r}_i$$

$$\overrightarrow{F}$$

$$\overrightarrow{F} = m_i \frac{d^2 \overrightarrow{r}_i}{dt^2}$$

After today, you should be able to

Explain the role of integration

algorithms in MD simulations.

Integration Algorithms Numerically Solve the Equations of Motion in Molecular Dynamics

Purpose of Integration Algorithms:

- **Numerical Solution:** Approximate the continuous equations of motion using discrete time steps.
- **Update Positions and Velocities:** Calculate the new positions and velocities of particles based on current forces.

Challenges Addressed by Integration Algorithms:

- **Stability:** Prevent numerical errors from accumulating over many time steps.
- **Accuracy:** Ensure that the trajectories closely follow the true physical behavior.
- **Efficiency:** Balance computational speed with the precision of the simulation.

Common Integration Algorithms

Verlet: Uses current and previous positions to calculate the next position

$$\mathbf{r}(t + \Delta t) = 2\,\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \frac{\mathbf{F}(t)}{m}\,(\Delta t)^2$$

Velocity Verlet: An extension of the Verlet algorithm that explicitly calculates velocities

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\,\Delta t + \frac{\mathbf{F}(t)}{2m}\,(\Delta t)^2$$

Time step length determines how "smooth" the trajectory

Say we want to simulation system for 100 fs

Smaller time steps lead to more calcualtions to simulate same amount of time

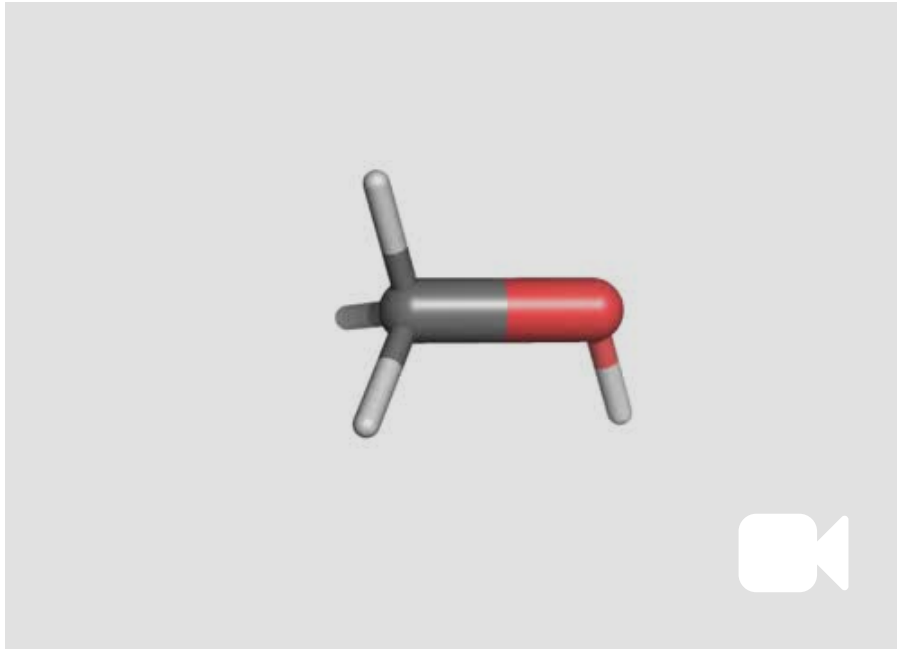| Δt	Number of total atomic force calculations
0.5 fs	200
1.0 fs	100
2.0 fs	50

After today, you should be able to
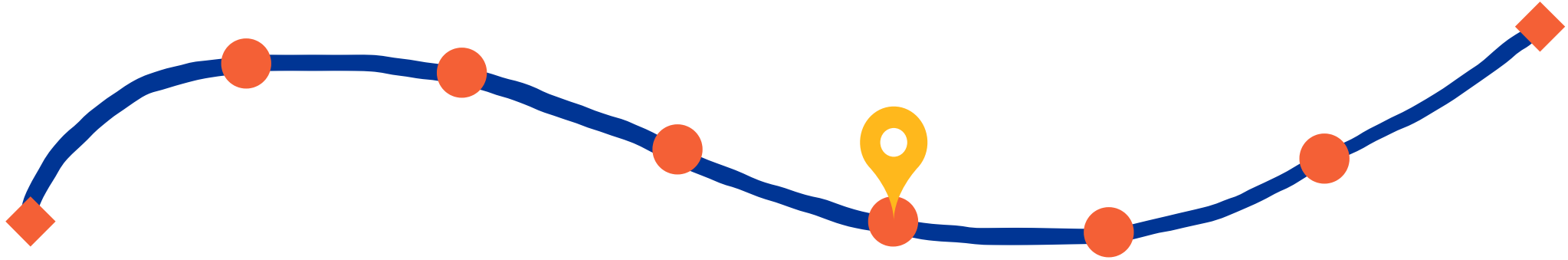
Describe the components of a

molecular mechanics force field.

We use "force fields" to compute energies and atomic forces
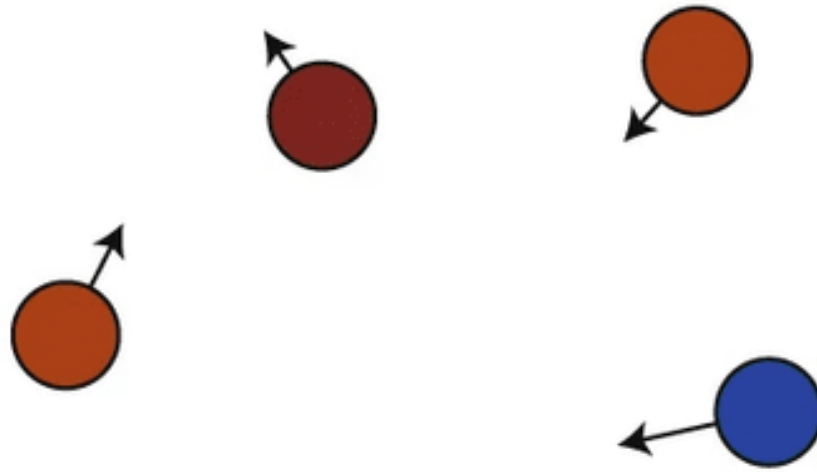
How can we do this quickly and accurately for large systems (e.g., proteins)?

How can we decompose dynamics into fundamental atomic motions?

GFN2-xTB; 1 fs time step; 1000 fs

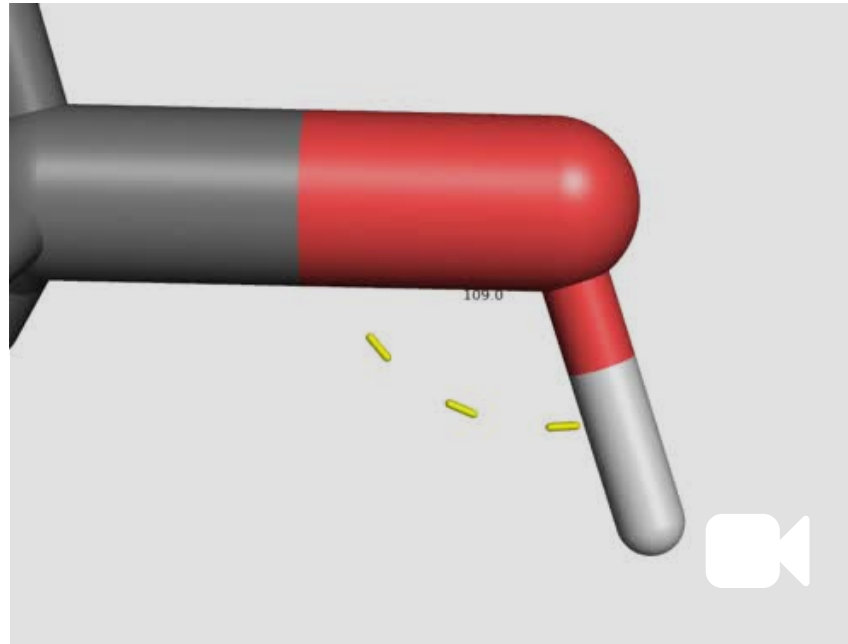Let's consider a QM simulation a single methanol molecule (H3COH)

How would we model these atomistic dyanmics classically?

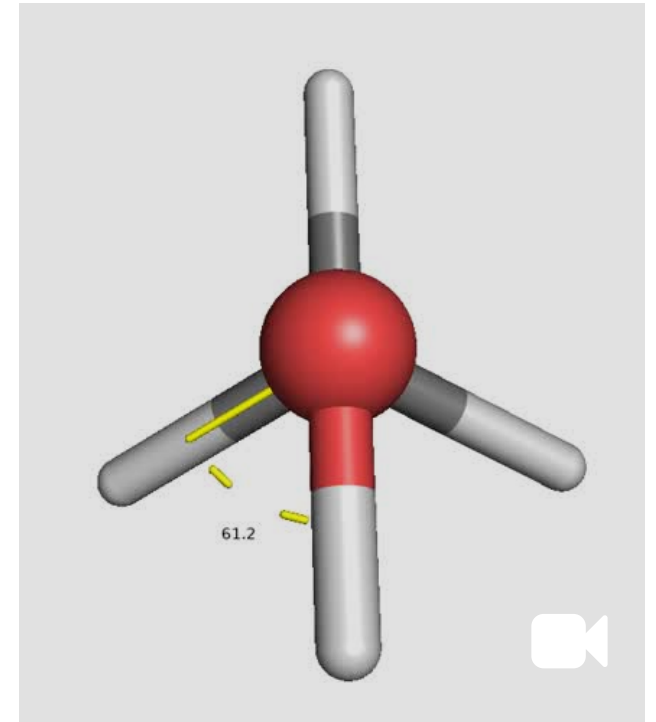# The dynamics of a molecule can be described as combinations of ...
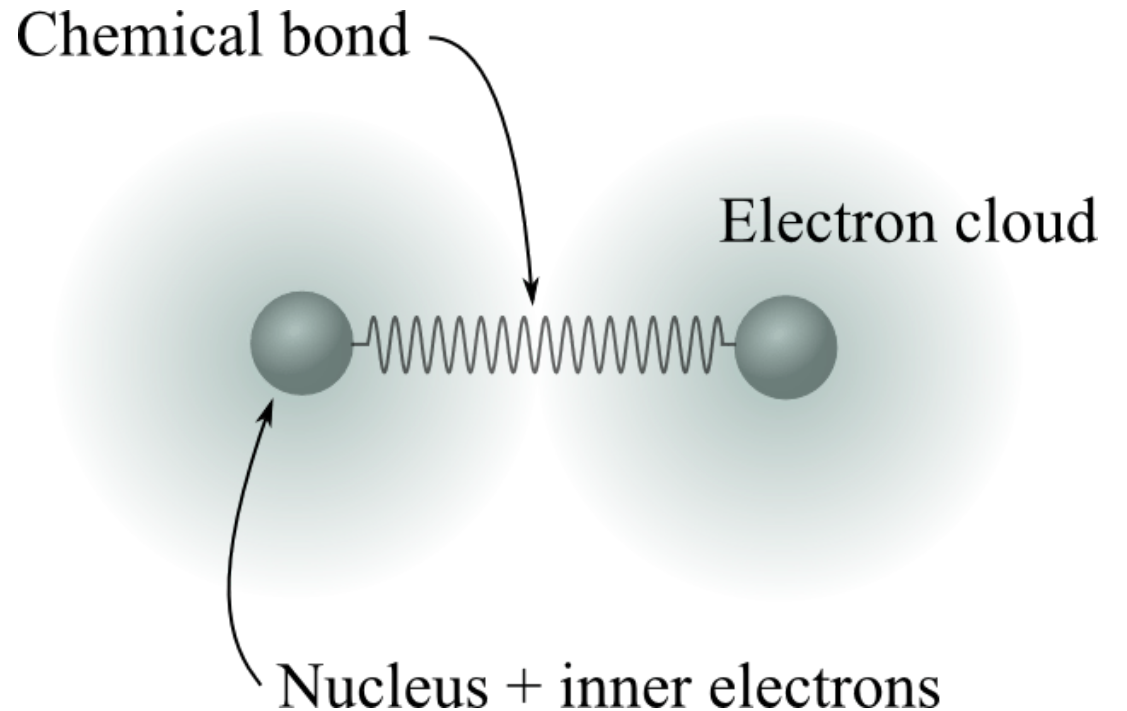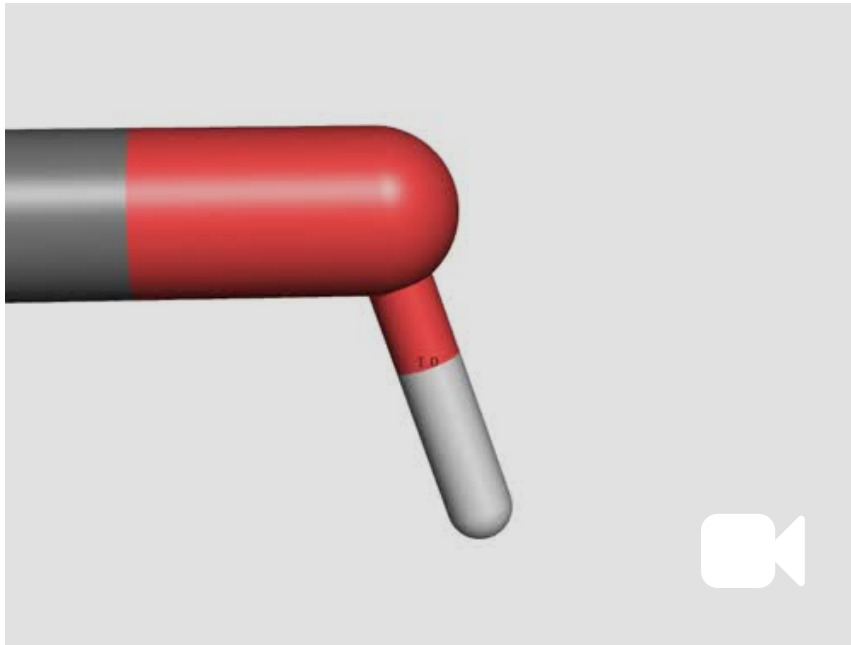
Bond lengths

Bond angles

Dihedral angles

Chemical bonds behave like springs

Two spheres (atoms) connected by a single spring

Chemical bond

Electron cloud

Nucleus + inner electrons

The spring resists changes in the distance
between the two atoms

Approximate bond vibrations as harmonic oscillators
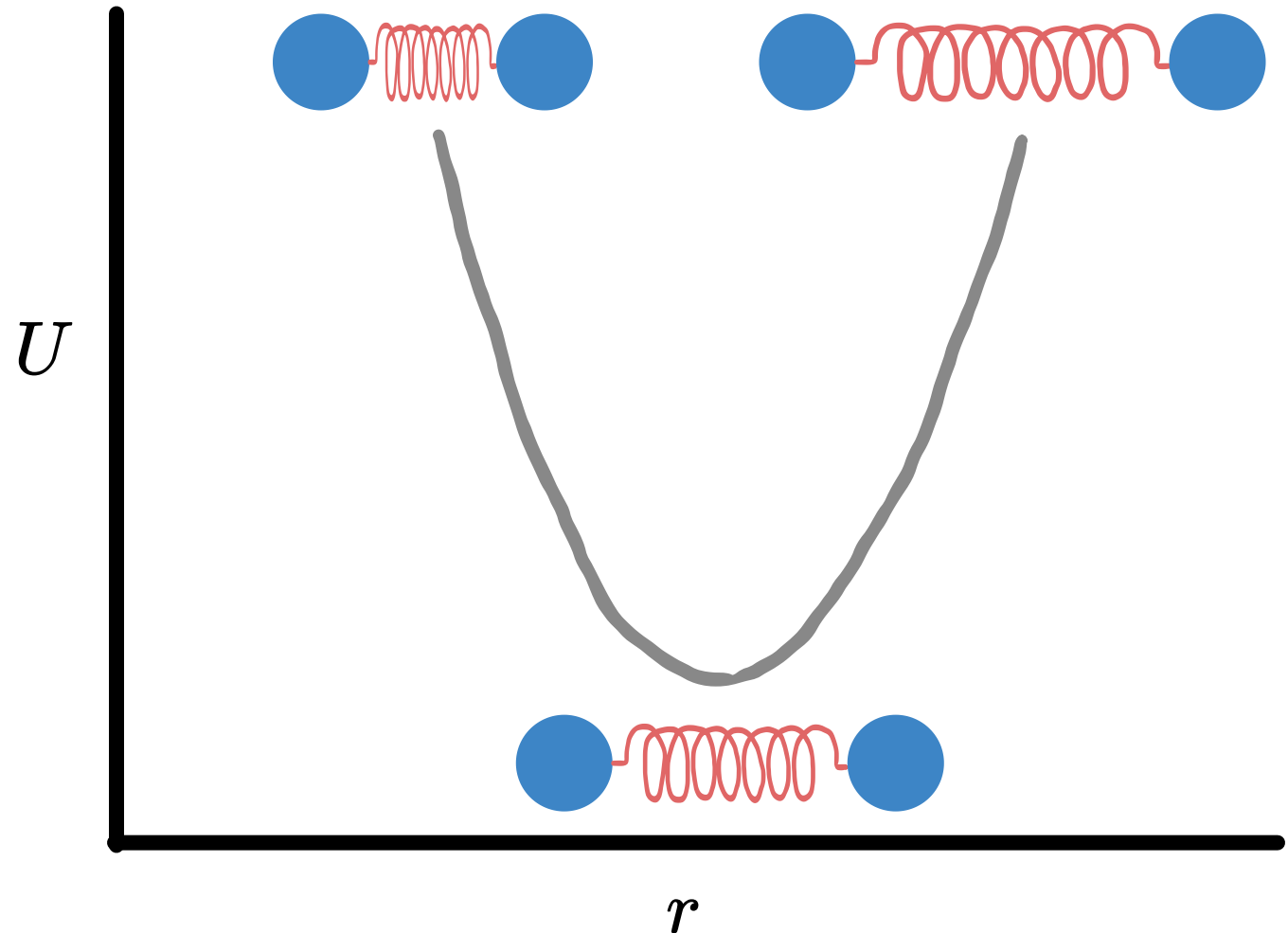
$$U = k\left(r - r_{eq}\right)^2$$

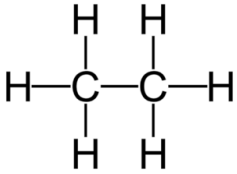r_{eq} Equilibrium bond length
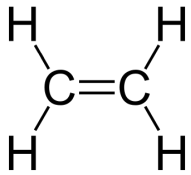
r Current bond length

k Spring constant (i.e., bond stiffness)
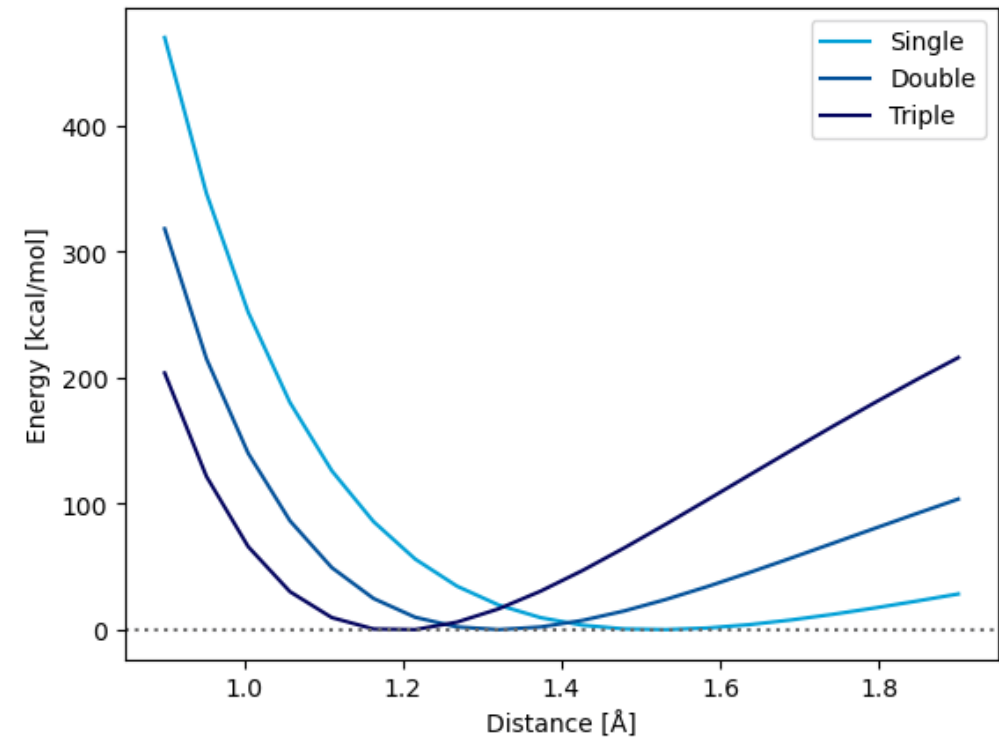
We can also get the force

$$F = -\frac{\partial U}{\partial r} = -2k\left(r - r_{eq}\right)$$

Spring constants are determined by bond order

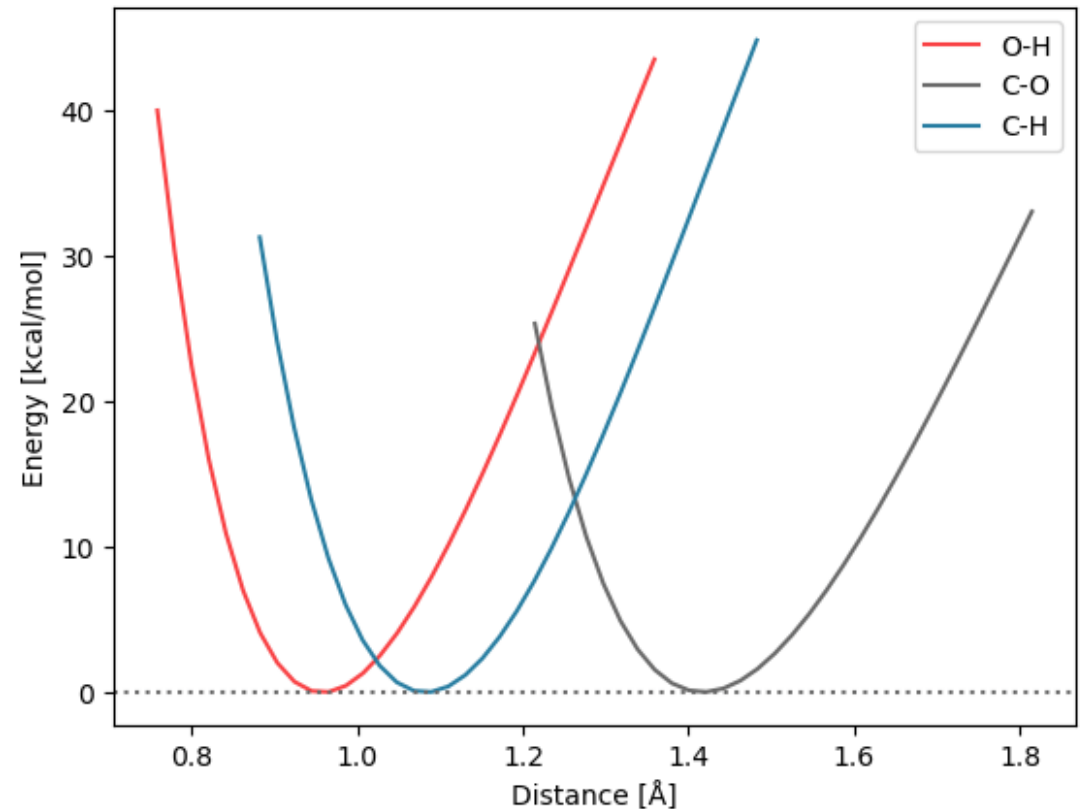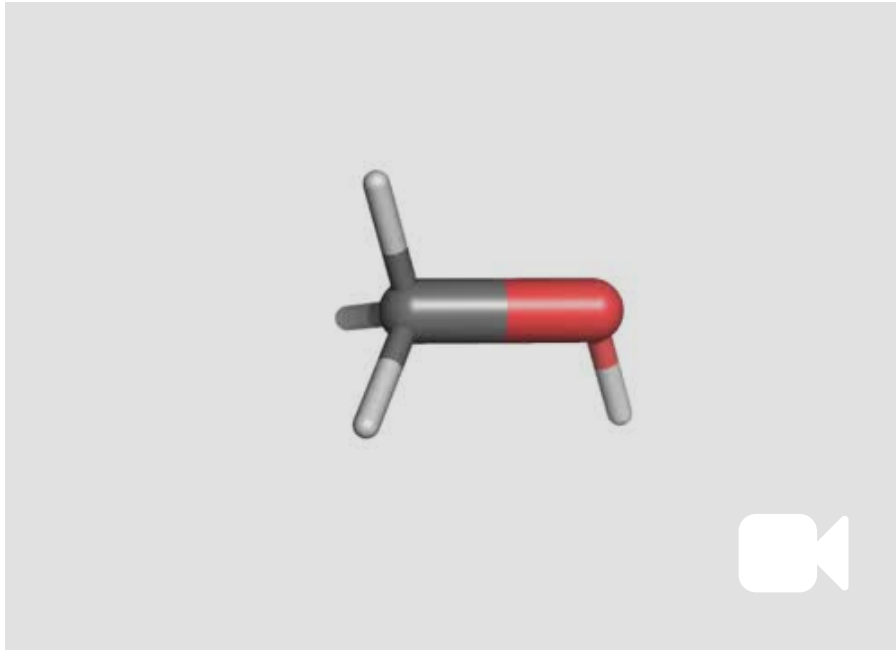|		k	r_{eq}
Single		300 kcal/mol/Å^2	1.54 Å
Double		450 kcal/mol/Å^2	1.34 Å
Triple		600 kcal/mol/Å^2	1.20 Å

(These are approximate values.)
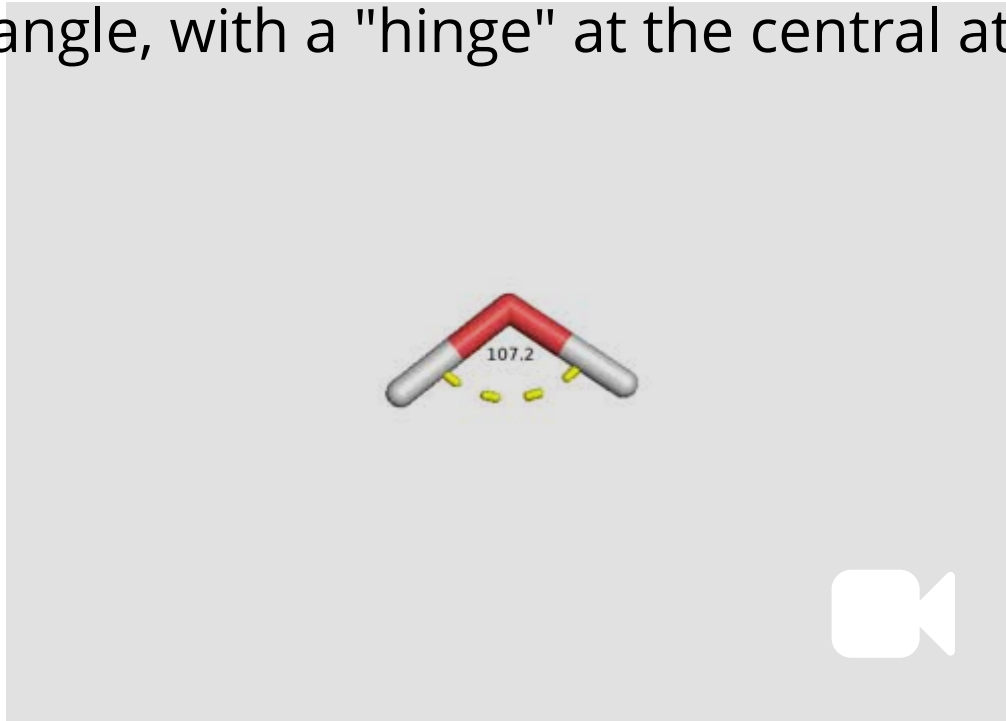
CISD/cc-pVTZ

Spring constants are determined by bond order and atom types

We can model each type of bond with a specific spring constant
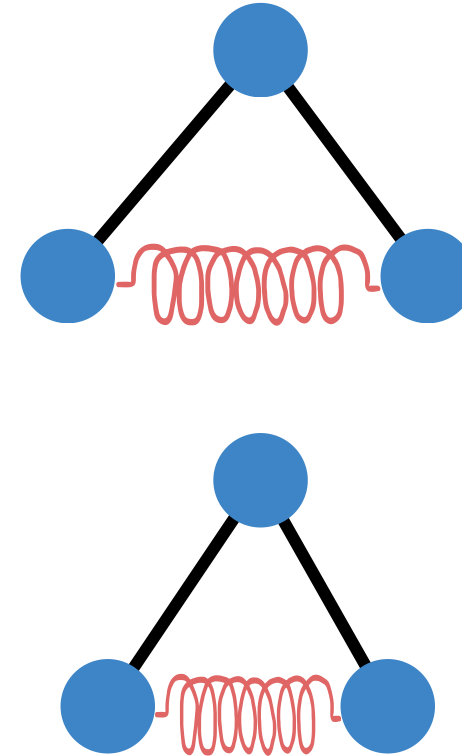
Bond angles behave like ... harmonic oscillators

Three balls connected by two springs forming an angle, with a "hinge" at the central atom.

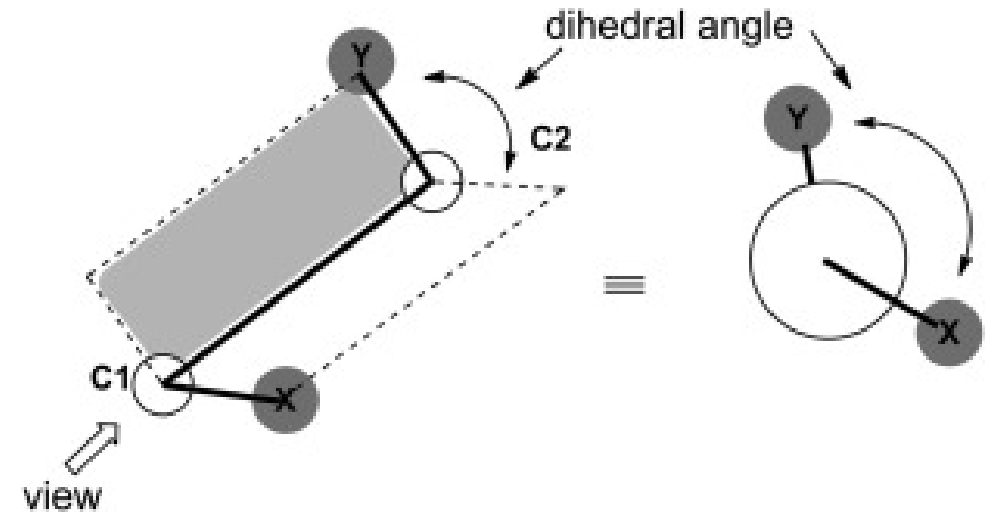We also have separate spring constants for bond angles

$$V = k\left(\theta_i - \theta_{eq}\right)^2$$

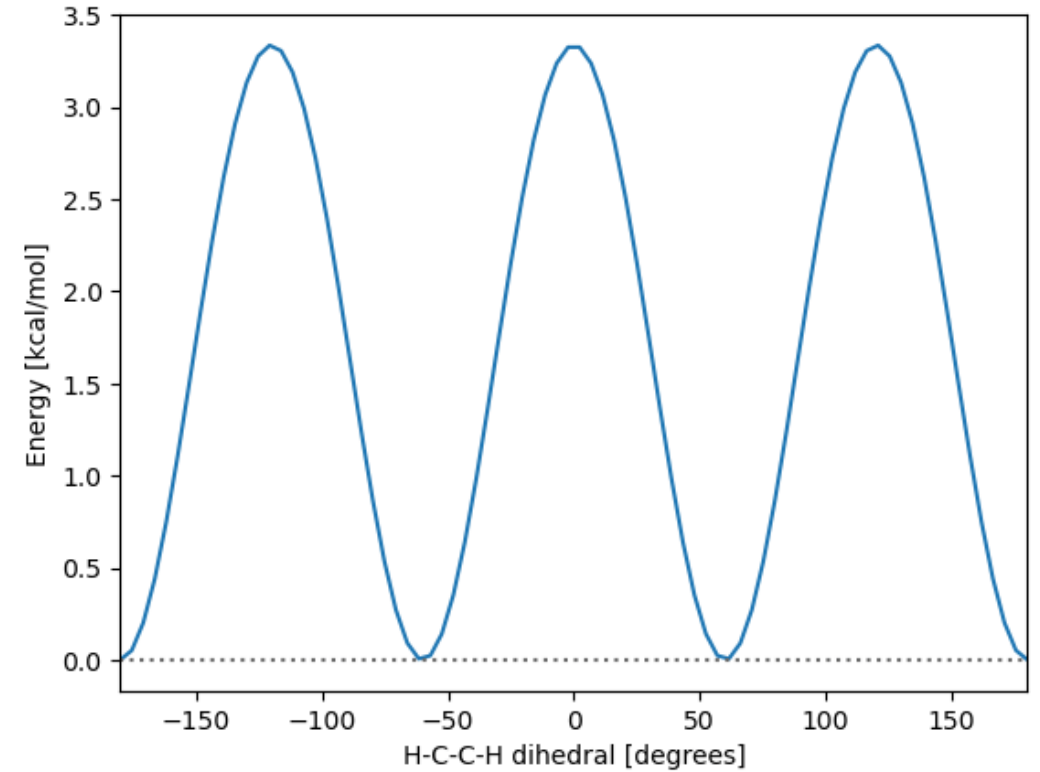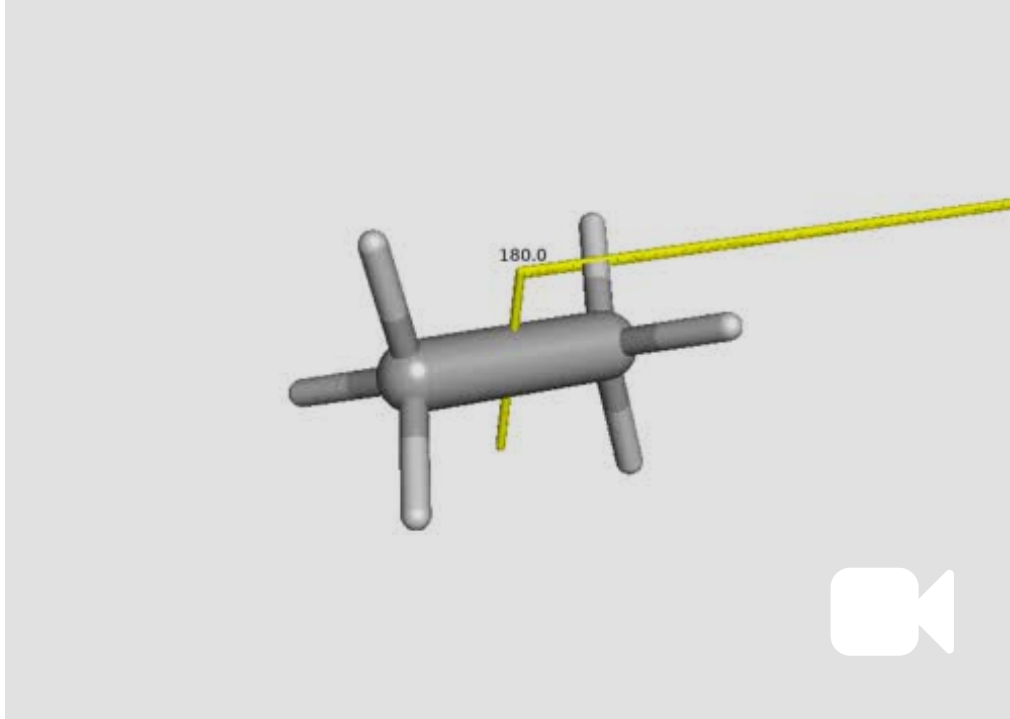# A dihedral angle is the angle between two planes

A dihedral angle is the angle between two planes formed by four sequentially bonded atoms (A–B–C–D)

The dihedral angle ϕ is the angle between these two planes

It describes the rotation around the bond between atoms B and C

Dihedral angles behave … not like springs

Energies are computed
with MP2/cc-pVTZ

Dihedrals vs. Bonds and Angles

- **Bonds and Angles**: Govern local geometry (bond lengths and bond angles) using **quadratic (harmonic)** potentials that favor specific distances and angles.
- **Dihedrals**: Govern torsional or rotational flexibility around bonds, typically using **periodic** or multi-well potentials to allow for multiple stable conformations.
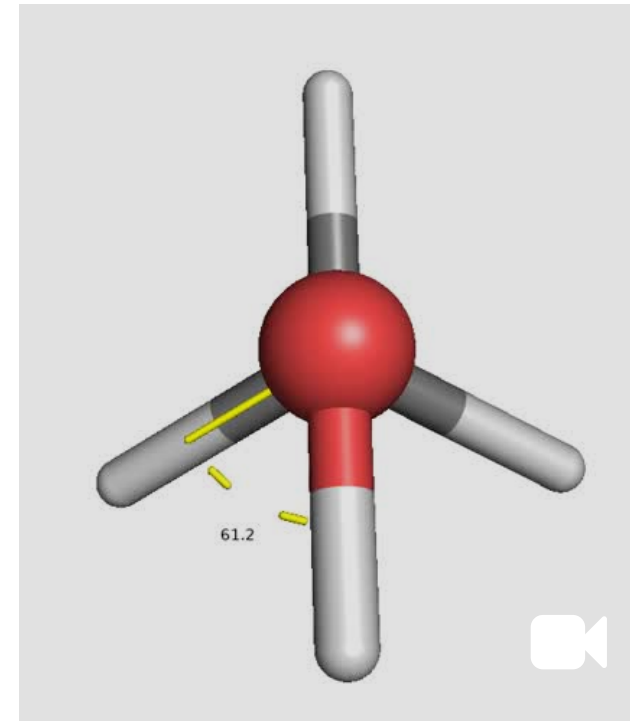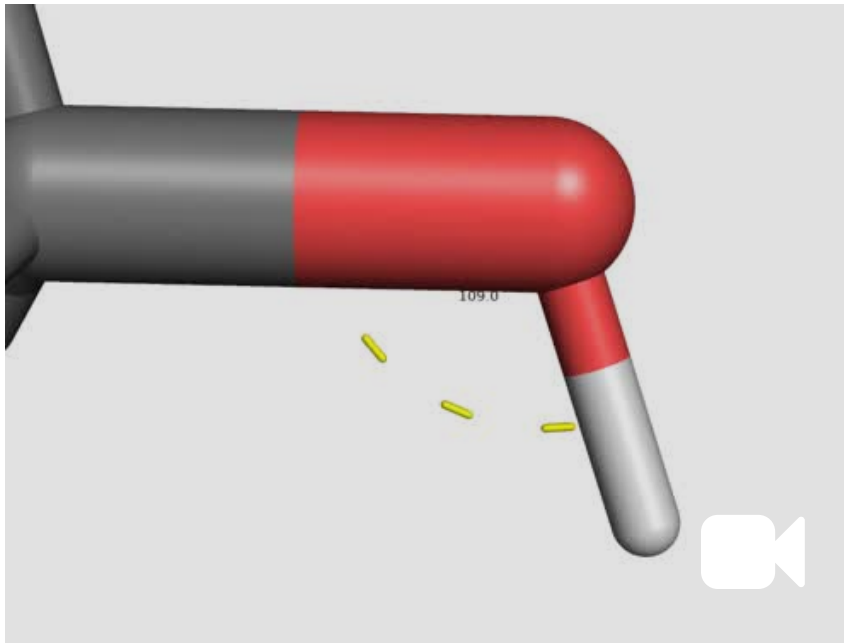
Dihedral potentials must capture arbitrary functions with rotational symmetry

Here, we have a periodic energy function with varying minima

How do we model this?
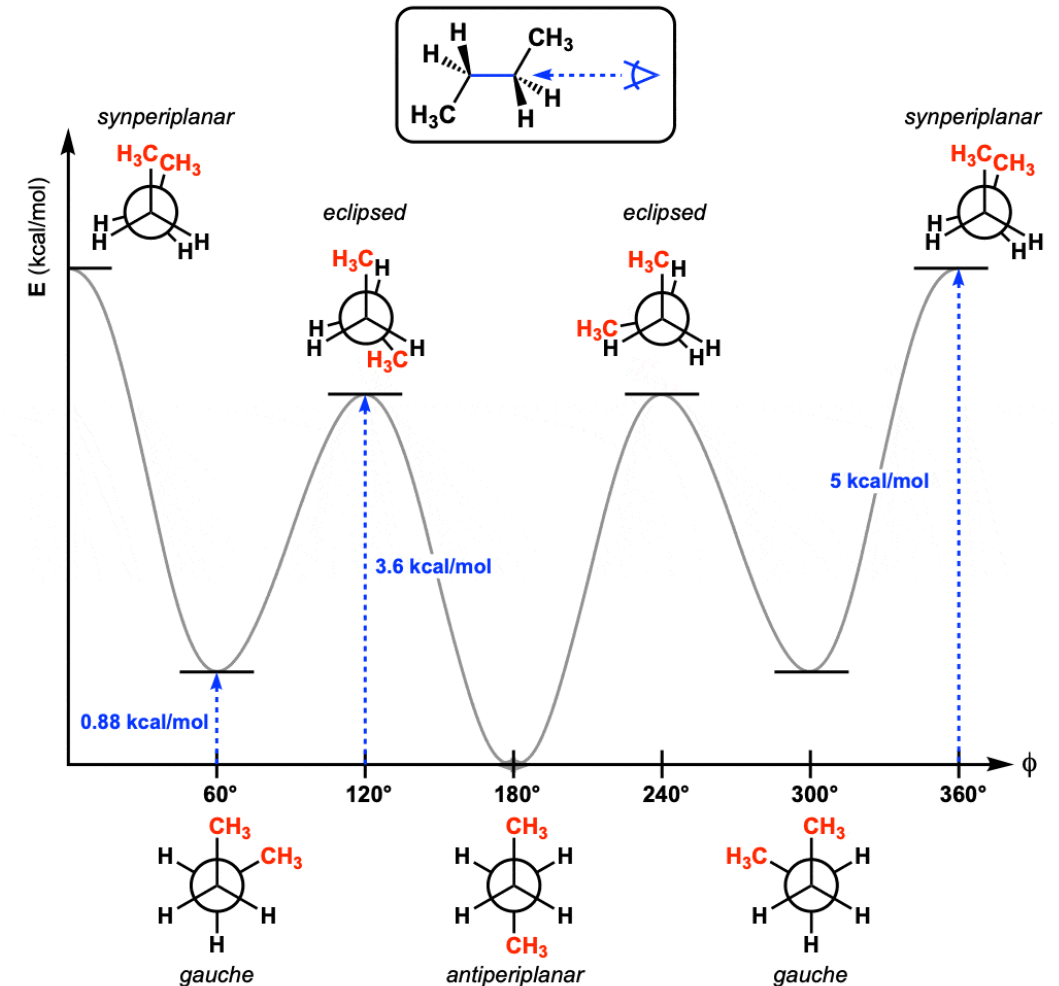
Fourier Series approximate functions as a sum of sine and cosine waves

$$h_n(x) \qquad\qquad\qquad f(x)$$

$$\sum h_n(x) = f(x)$$

$$\longrightarrow$$

$$\sum h_n(x) = f(x)$$

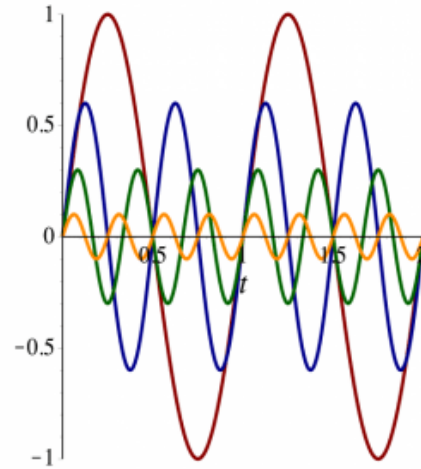$$\longrightarrow$$

37
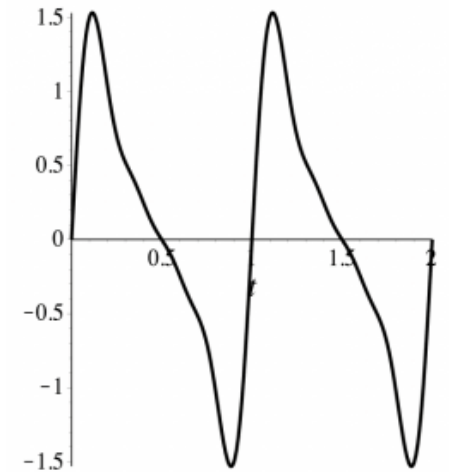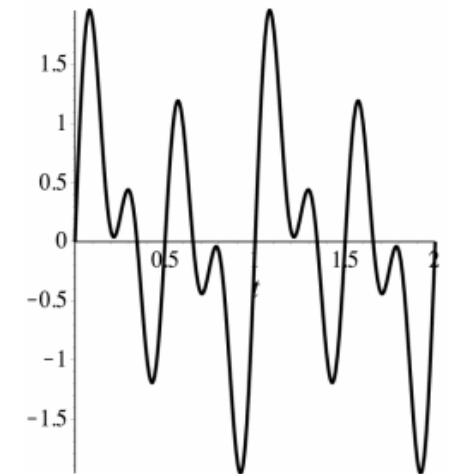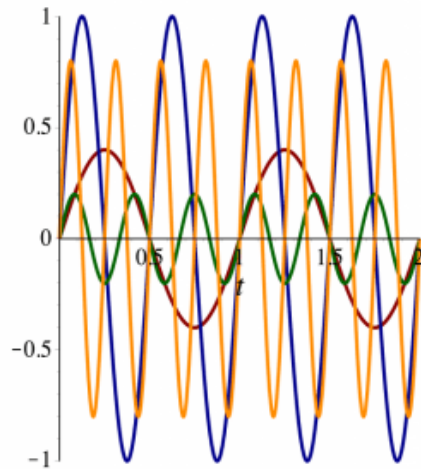
Fourier series can approximate (any) symmetrical rotational energy function

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left[a_n \cos(nx) + b_n \sin(nx) \right]$$

a_0 Constant term (i.e., average of function)

a_n b_n Amplitude coefficients

n Harmonic number (i.e., frequency)

Higher harmonics add finer details to the approximation, enhancing the accuracy of the representation
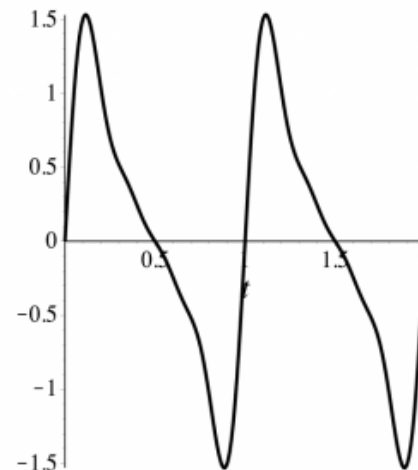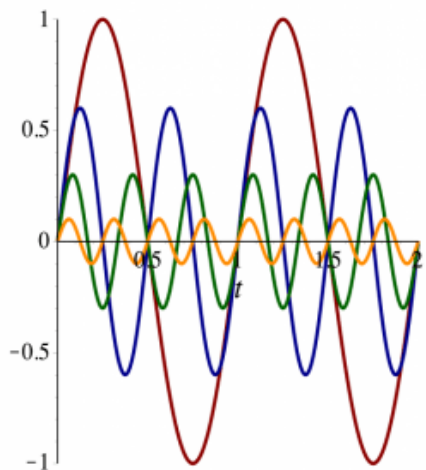
Adding more sine and cosine terms improves the approximation, allowing the Fourier Series to closely match the original complex function

Dihedral potentials use custom Fourier series

$$U(\phi) = \sum_{n=1}^{N} \frac{V_n}{2} \left[1 + \cos(n\phi - \gamma_n)\right]$$

61.2
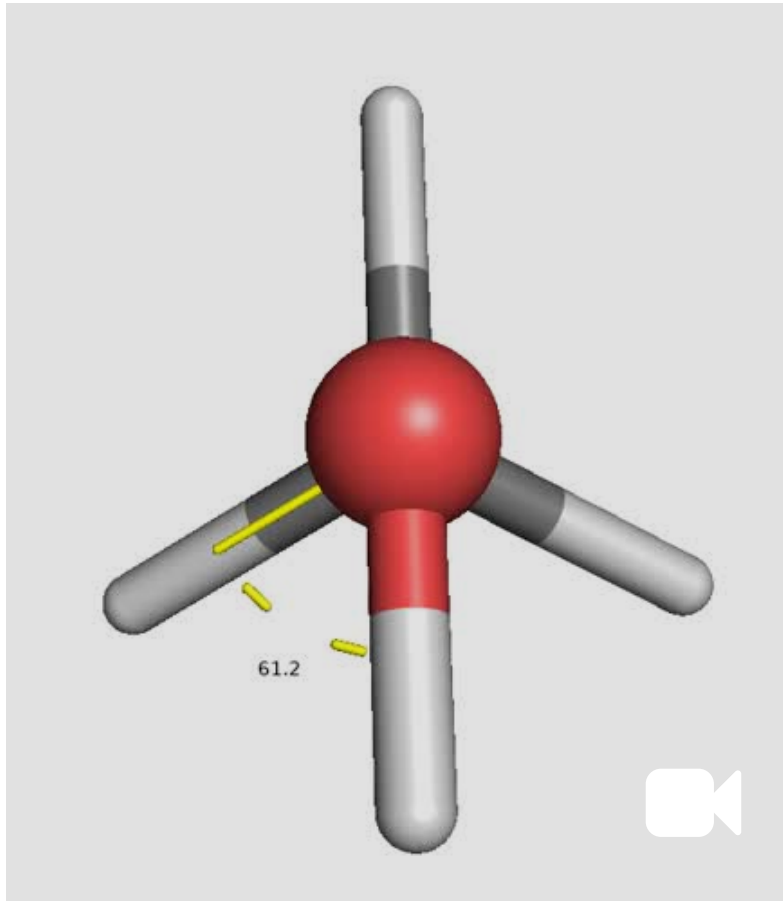
ϕ Dihedral angle

V_n Amplitude for n-th Fourier

γ_n Phase shift
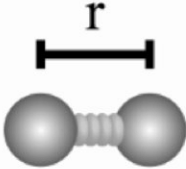
N Number of periodic terms

Sum of all bonded terms provides a single-molecule force field
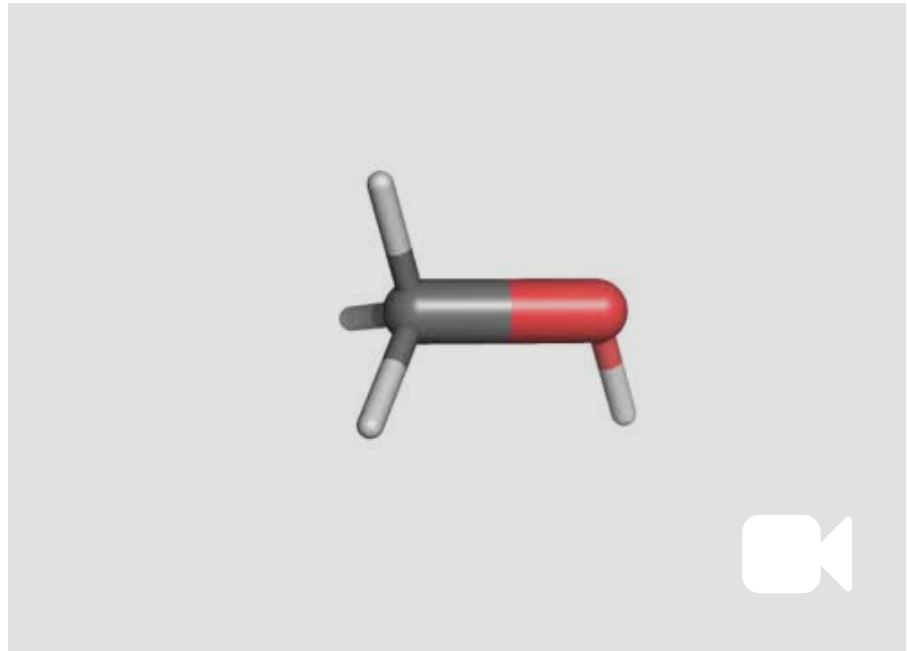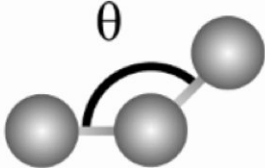
Bonded
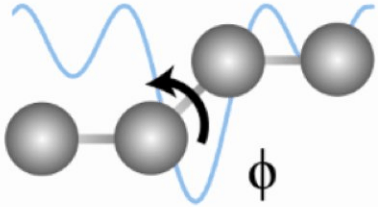
$$E_{total} = \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)]$$

We rarely simulate one molecule; what about multiple?
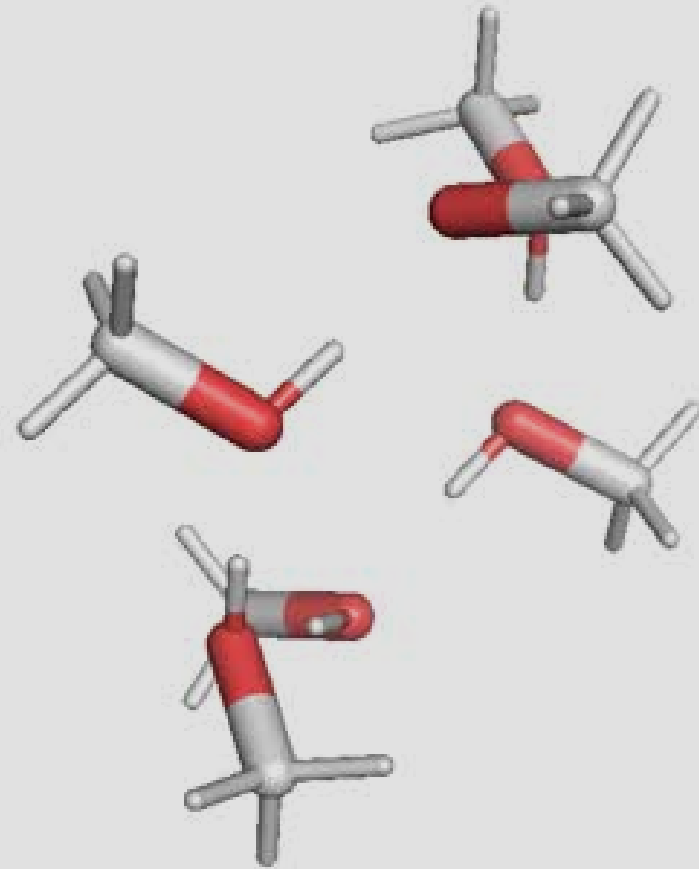
After today, you should be able to
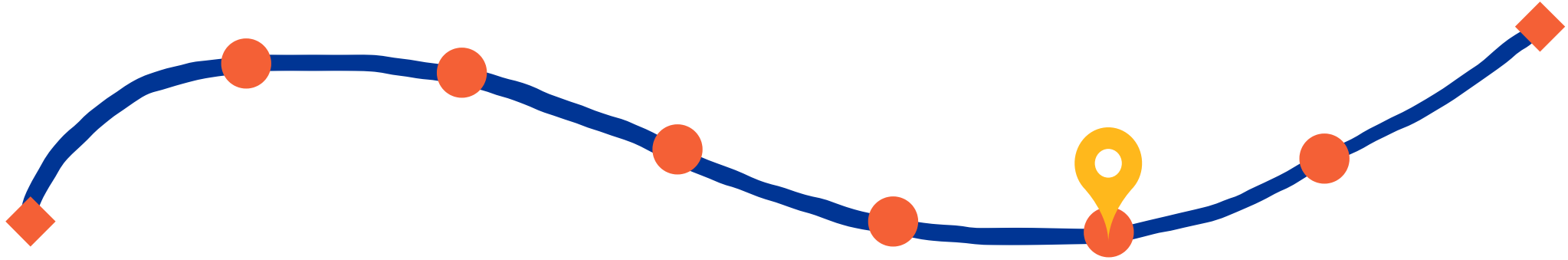
Understand noncovalent

contributions to force fields

Noncovalent Interactions Are Crucial for Simulating Multiple Molecules in MD

Role in Molecular Assembly:

- Facilitate the organization of molecules into complex structures.
- Determine the macroscopic properties of materials (e.g., solubility, melting points).

Importance in Biological Systems:

- Govern essential processes like enzyme-substrate binding, protein folding, and membrane formation.
- Critical for understanding biochemical pathways and drug design.

While covalent bonds define the primary structure of molecules, noncovalent interactions are pivotal in dictating how molecules interact

Dispersion Forces

- **Nature:** Weak, attractive forces arising from instantaneous dipoles in molecules.
- **Role:** Stabilize molecular assemblies by promoting close packing.

$$U = -\frac{C_6}{r^6}$$

C_6 Dispersion coefficient
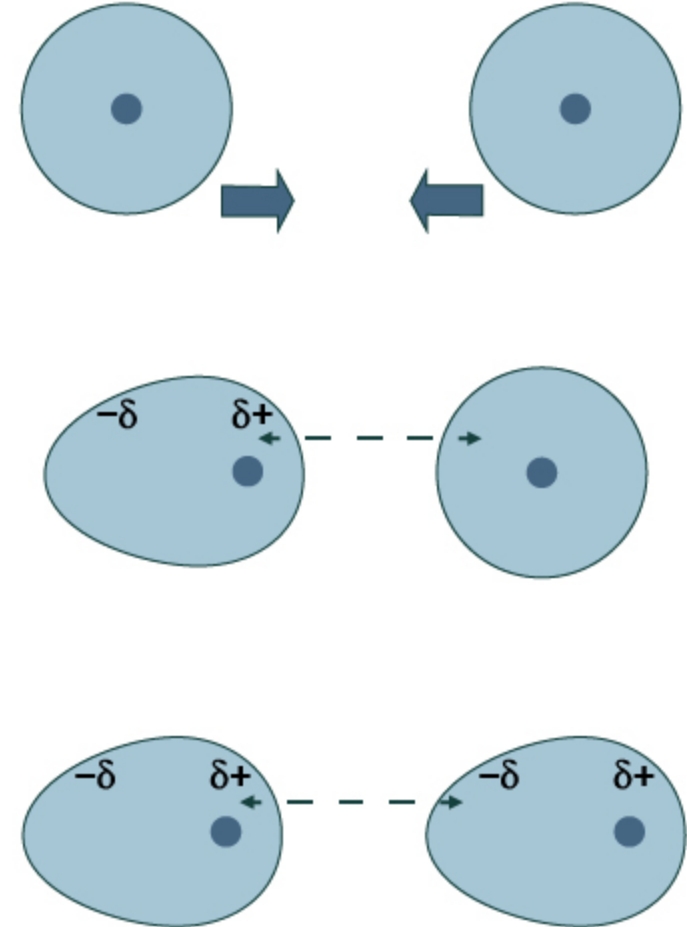
Repulsion Forces

- **Nature:** Strong, short-range forces due to overlapping electron clouds.
- **Role:** Prevent atoms from collapsing into each other, maintaining molecular integrity

$$U = \frac{C_{12}}{r^{12}}$$

C_{12} Repulsion coefficient
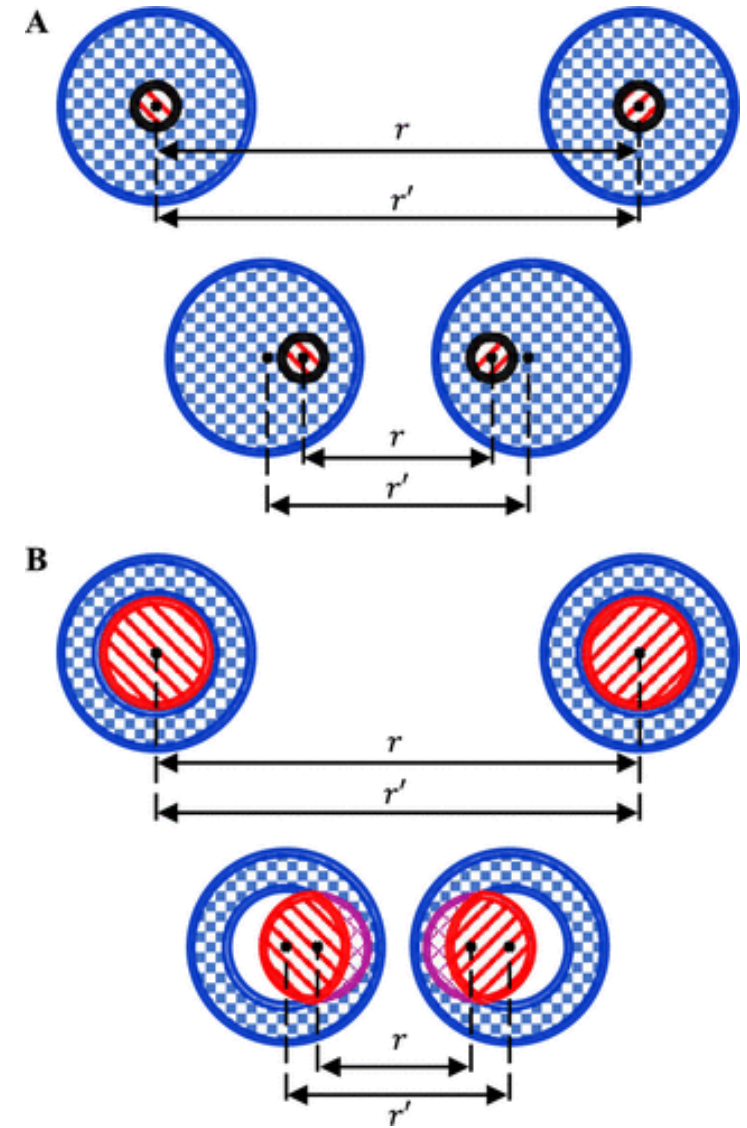
Combined van der Waals Potential

Van der Waals forces are modeled using the Lennard-Jones potential, which captures both the attractive and repulsive aspects of noncovalent interactions

$$U = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right]$$
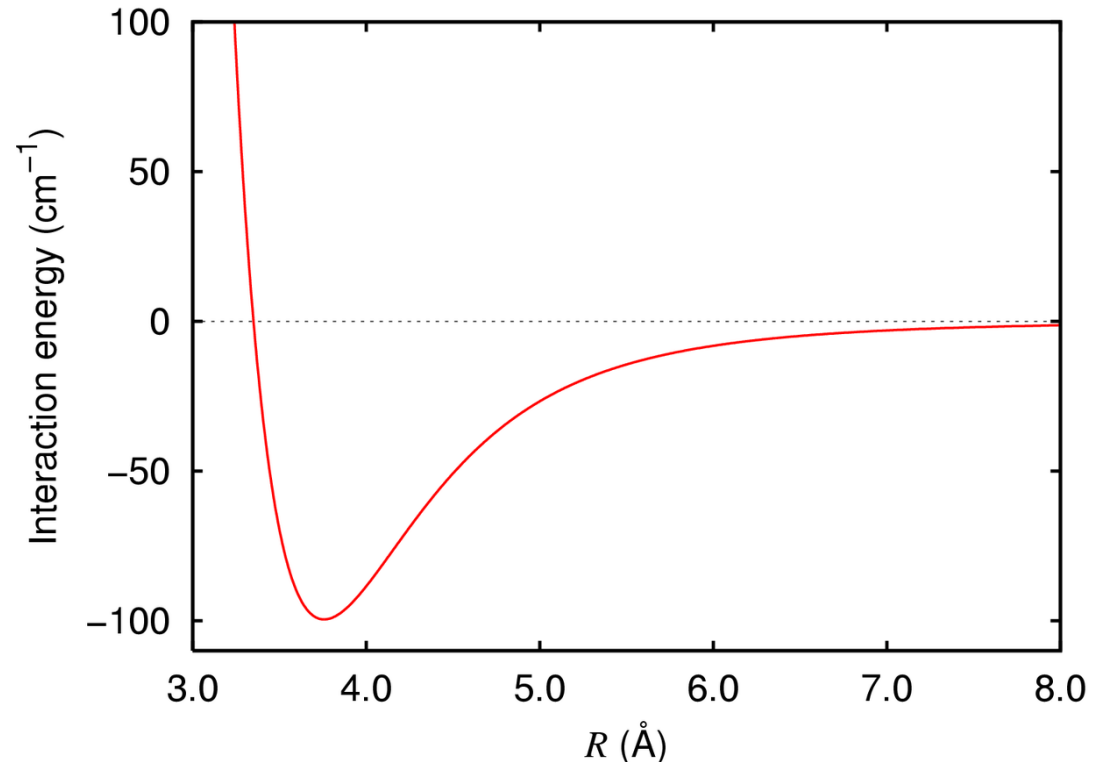
Electrostatic Interactions Drive Charged and Polar Molecule Behavior
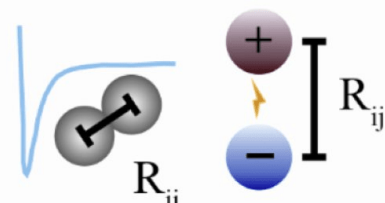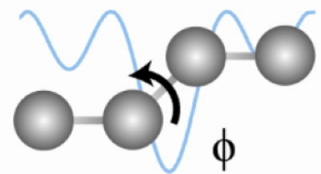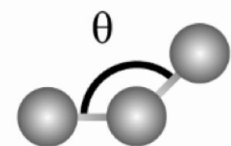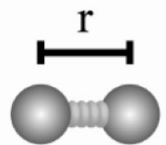
Electrostatic forces decay as *1/r*, making them significant over longer distances compared to van der Waals forces

$$U(r) = \frac{1}{4\pi\varepsilon_0}\frac{q_1 q_2}{r}$$
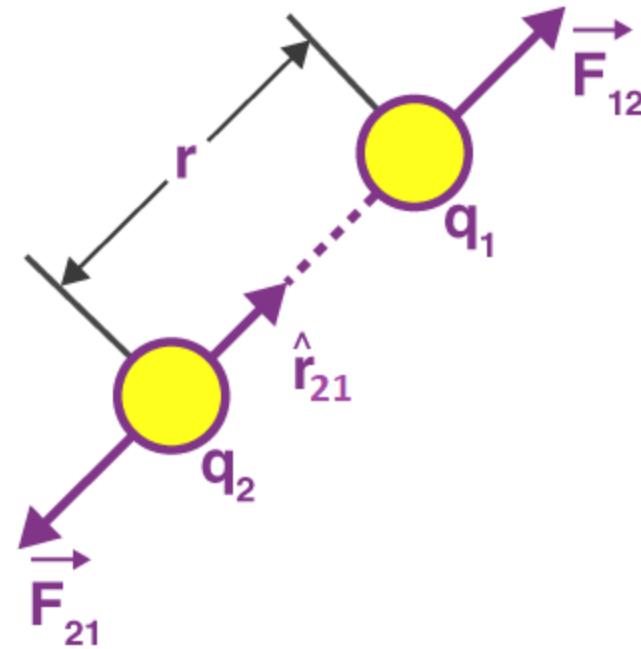
Bonded and nonbonded interactions make our complete force field
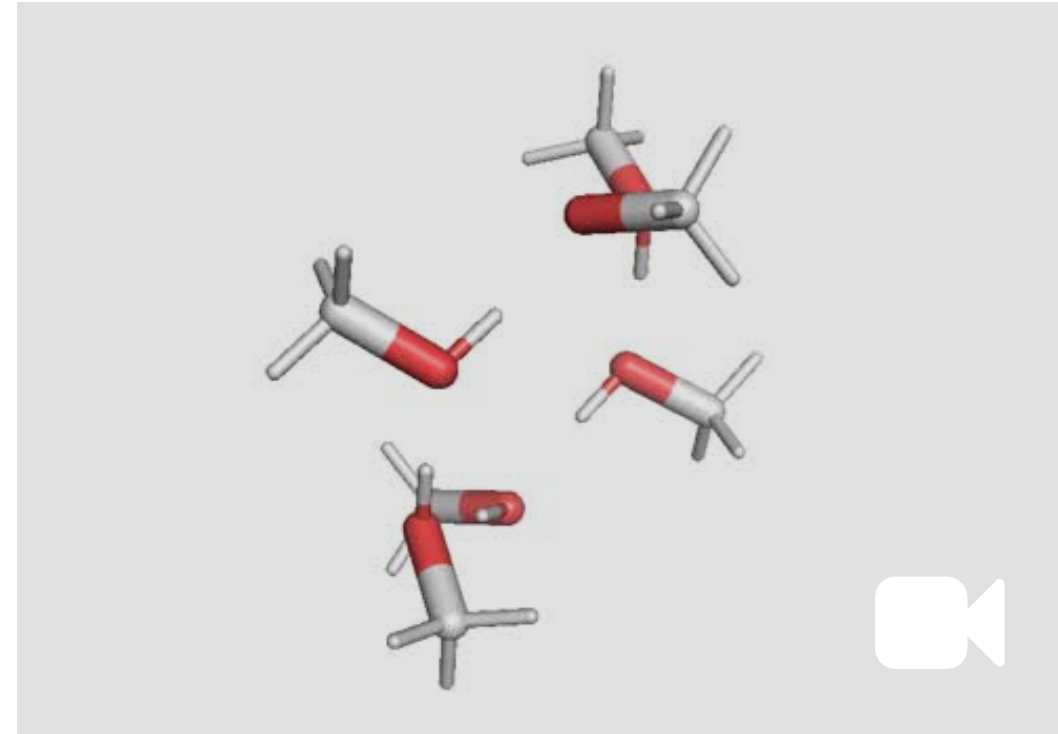
$$E_{total} = \sum_{bonds} K_r(r - r_{eq})^2 + \sum_{angles} K_\theta(\theta - \theta_{eq})^2 + \sum_{dihedrals} \frac{V_n}{2}[1 + \cos(n\phi - \gamma)] + \sum_{i<j}\left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\epsilon R_{ij}}\right]$$

Bonded — — — — — — — — — — — — — — — — — — **Non-bonded**

After today, you should be able to
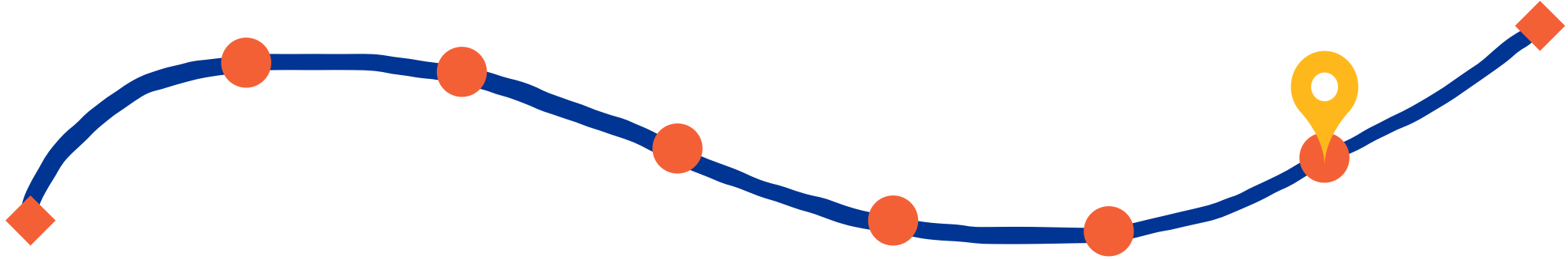
Identify data for force field

parameterization

Parameterizing Force Fields Begins with Quantum Mechanical Data for Small Molecules

Role of Quantum Mechanics:

- **QM Calculations:** Provide high-accuracy data on molecular geometries, energetics, and electronic distributions.
- **Data Utilization:** QM data inform the selection and tuning of force field parameters to ensure they reflect true molecular behavior.

Small Molecule Focus:

- **Simplicity:** Smaller molecules have fewer atoms and simpler interactions, making QM calculations more manageable.
- **Accuracy:** QM methods (e.g., Density Functional Theory, Hartree-Fock) yield precise information essential for initial parameterization.

From Small Molecules to Proteins: Increasing Complexity in Force Field Parameterization

Complexity of Proteins:

- **Size and Structure:** Proteins consist of hundreds to thousands of atoms with intricate three-dimensional structures.
- **Diverse Interactions:** Include a variety of noncovalent interactions, such as hydrogen bonds, ionic bonds, hydrophobic interactions, and van der Waals forces.

Limitations of QM for Large Systems:

- **Computational Cost:** QM calculations become computationally prohibitive for large biomolecules like proteins.
- **Alternative Strategies:** Utilize QM data from representative small segments or use empirical and semi-empirical methods.

Experimental Data Are Crucial for Refining Force Field Parameters

Types of Experimental Data:

- **Spectroscopic Data:** Infrared (IR), Nuclear Magnetic Resonance (NMR), and Raman spectroscopy provide insights into bond vibrations and molecular geometries.
- **Crystallography:** X-ray crystallography offers precise information on atomic positions and molecular conformations.
- **Thermodynamic Measurements:** Data on melting points, boiling points, and solvation energies inform interaction strengths.

Parameter Optimization:

- **Fitting Process:** Adjust force field parameters to minimize discrepancies between simulation results and experimental observations.
- **Validation Metrics:** Use root-mean-square deviation (RMSD), binding affinities, and structural stability as benchmarks.

Fitting Force Field Parameters to Experimental Data Ensures Realistic Simulations

Parameter Adjustment:

- **Process:** Fine-tune force field parameters to minimize discrepancies between simulation outcomes and experimental observations.
- **Techniques:** Use of optimization algorithms and statistical methods to achieve best-fit parameters.

Iterative Refinement:

- **Feedback Loop:** Use simulation results to identify parameter inaccuracies and iteratively adjust them based on experimental data.
- **Continuous Improvement:** Enhance force field accuracy through ongoing comparisons and adjustments.

Challenges in Parameterizing Force Fields for Proteins

High Dimensionality:

- **Issue:** Proteins possess numerous degrees of freedom, making comprehensive parameterization computationally intensive.
- **Solution:** Utilize advanced optimization techniques and high-performance computing resources.

Diverse Chemical Environments:

- **Issue:** Different regions of a protein (e.g., active sites, hydrophobic cores) experience varied chemical environments.
- **Solution:** Develop region-specific parameters or use adaptive force fields that can account for environmental variations.

Challenges in Parameterizing Force Fields for Proteins

Dynamic Conformational Changes:

- **Issue:** Proteins frequently undergo conformational shifts that must be accurately captured by the force field.
- **Solution:** Incorporate flexible dihedral terms and ensure that parameters support a wide range of conformational states.

Long-Range Electrostatic Interactions:

- **Issue:** Accurate modeling of electrostatics in large, charged systems is computationally demanding.
- **Solution:** Implement efficient algorithms like Particle Mesh Ewald (PME) and use approximations where appropriate.

Summary of Force Field Parameterization Process

Step-by-Step Process:

- **Quantum Mechanical Calculations:** Obtain high-accuracy data for small molecules and representative fragments.
- **Empirical Data Integration:** Incorporate experimental measurements to validate and refine parameters.
- **Parameter Optimization:** Adjust force field parameters through iterative simulations and comparisons.
- **Advanced Techniques:** Utilize machine learning, multi-scale modeling, and automated pipelines to enhance parameter accuracy and efficiency.

Different force fields are tailored for specific types of molecules and applications

- **Common Force Fields:**
  - **AMBER:** Optimized for proteins and nucleic acids.
  - **CHARMM:** Versatile, used for a wide range of biomolecules.
  - **OPLS:** Focuses on liquids and organic molecules.
- **Selection Criteria:**
  - Compatibility with the system being studied.
  - Availability of parameters for the molecules of interest.
- **Limitations:**
  - Force fields are approximations and may not capture all interactions.
  - Ongoing development aims to improve accuracy and transferability.

Before the next class, you should

Lecture 13:
Molecular simulation princples

Lecture 14:
Molecular system representations

Today

Tuesday

- Work on A05
- Review material