

BIOC404 Biochemical Methods

Introduction to Biomolecular Simulations

Dr. Erich Kuechler

erich.kuechler@msl.ubc.ca

My Office Doesn't Matter Anymore

Introduction to biomolecular simulation

Lecture 1:
Molecular Dynamics
& Ensembles

Lecture 2:
Force Fields &
Solvation Models

Lecture 3:
Enhanced Sampling
Methods

Assignment:
Analyze simulation output and relate the findings to concepts introduced by the course.

Introduction to biomolecular simulation

Lecture 1:
Molecular Dynamics
& Ensembles

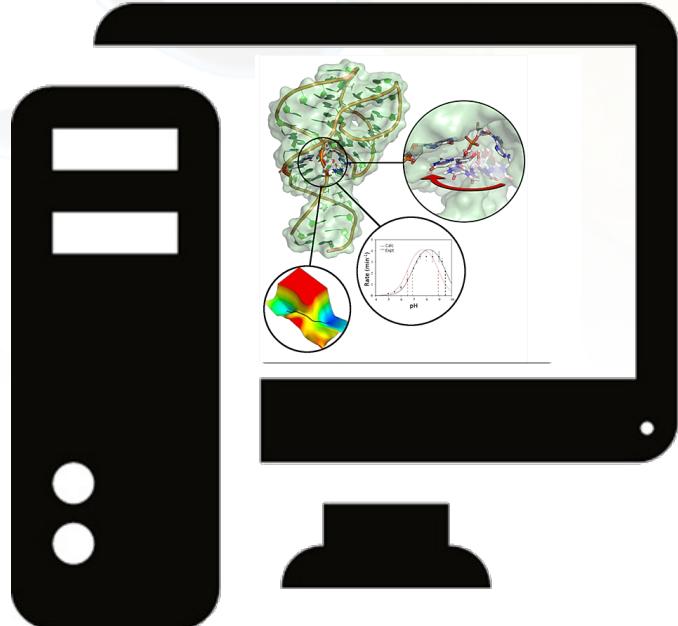
Lecture
2:

Analy-

Lecture 3:
Sampling
strategies

DISCLAIMER:
This is an introduction, I will not cover everything but will
instead give you a basic understanding of what molecular
dynamics is and how you can start to use it!
This will also lead to concepts introduced by the
course.

Introduction to biomolecular simulation: Lecture 1



Part 1: Computational techniques in biology

Part 2: Methods of molecular modeling

Part 3: Introduction to molecular dynamics

- Potential energy landscapes
- Sampling and phase space
- Ensembles and integrators

Computational biology and biochemistry

**Bioinformatics
and
Data Mining**

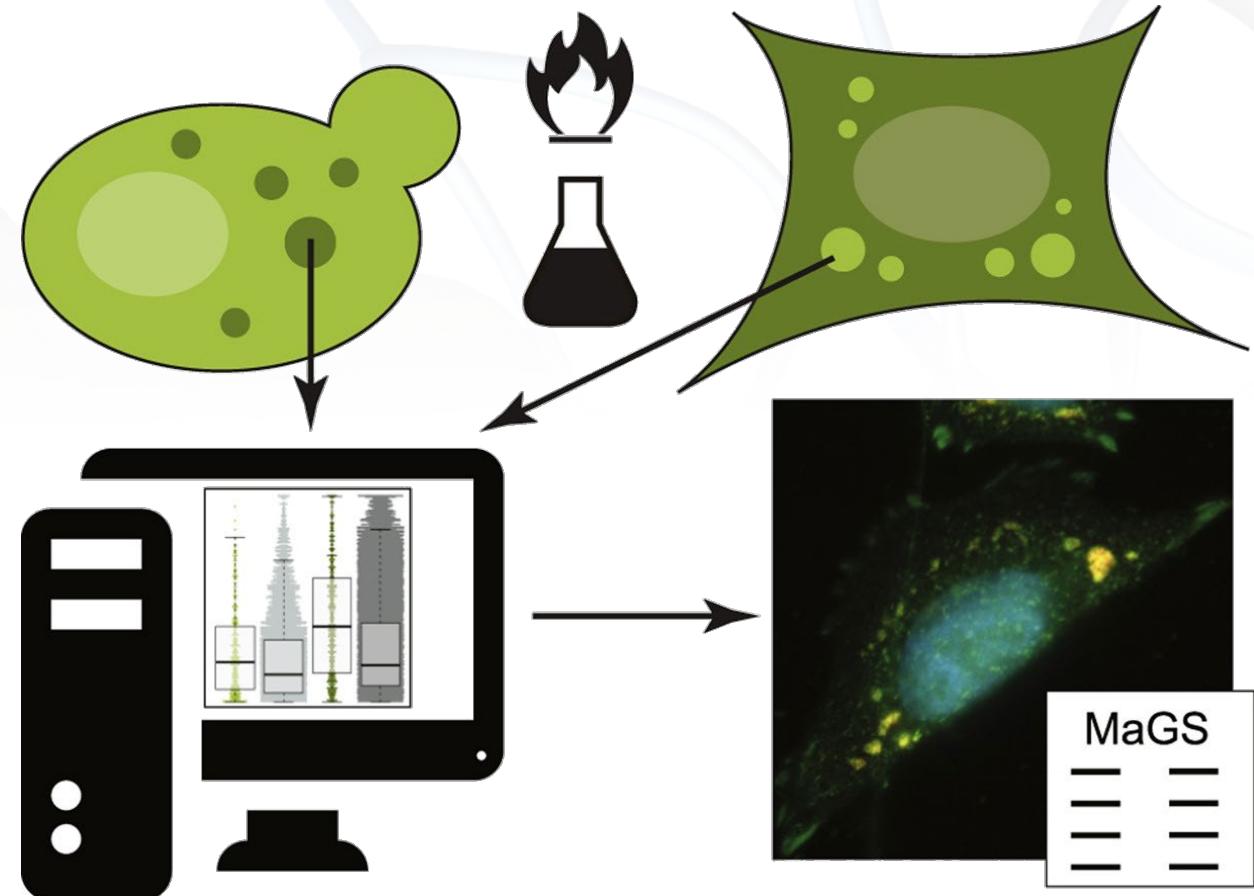
**Mathematical
Modeling and
Systems Bio**

**Structural
Modeling and
Simulation**

Computational biology and biochemistry

Bioinformatics and Data Mining

Uses **statistical** techniques to inform relationships or correlations in biological data sets.



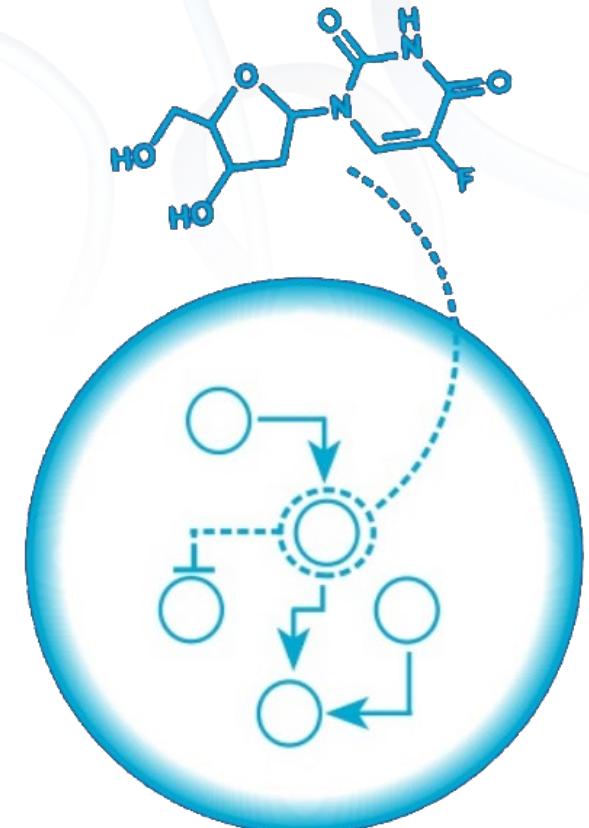
Computational biology and biochemistry

Bioinformatics and Data Mining

Uses **statistical** techniques to inform relationships or correlations in biological data sets.

Mathematical Modeling and Systems Bio

Uses **mathematical** frameworks to predict specific properties in complex biological systems



Computational biology and biochemistry

Bioinformatics and Data Mining

Uses **statistical** techniques to inform relationships or correlations in biological data sets.

Mathematical Modeling and Systems Bio

Uses **mathematical** frameworks to predict specific properties in complex biological systems

Structural Modeling and Simulation

Uses **physical and chemical** principles to obtain microscopic and macroscopic property quantification

Computational biology and biochemistry

Bioinformatics and Data Mining

Uses **statistical** techniques to inform relationships or correlations in biological data sets.

Mathematical Modeling and Systems Bio

Uses **mathematical** frameworks to predict specific properties in complex biological systems

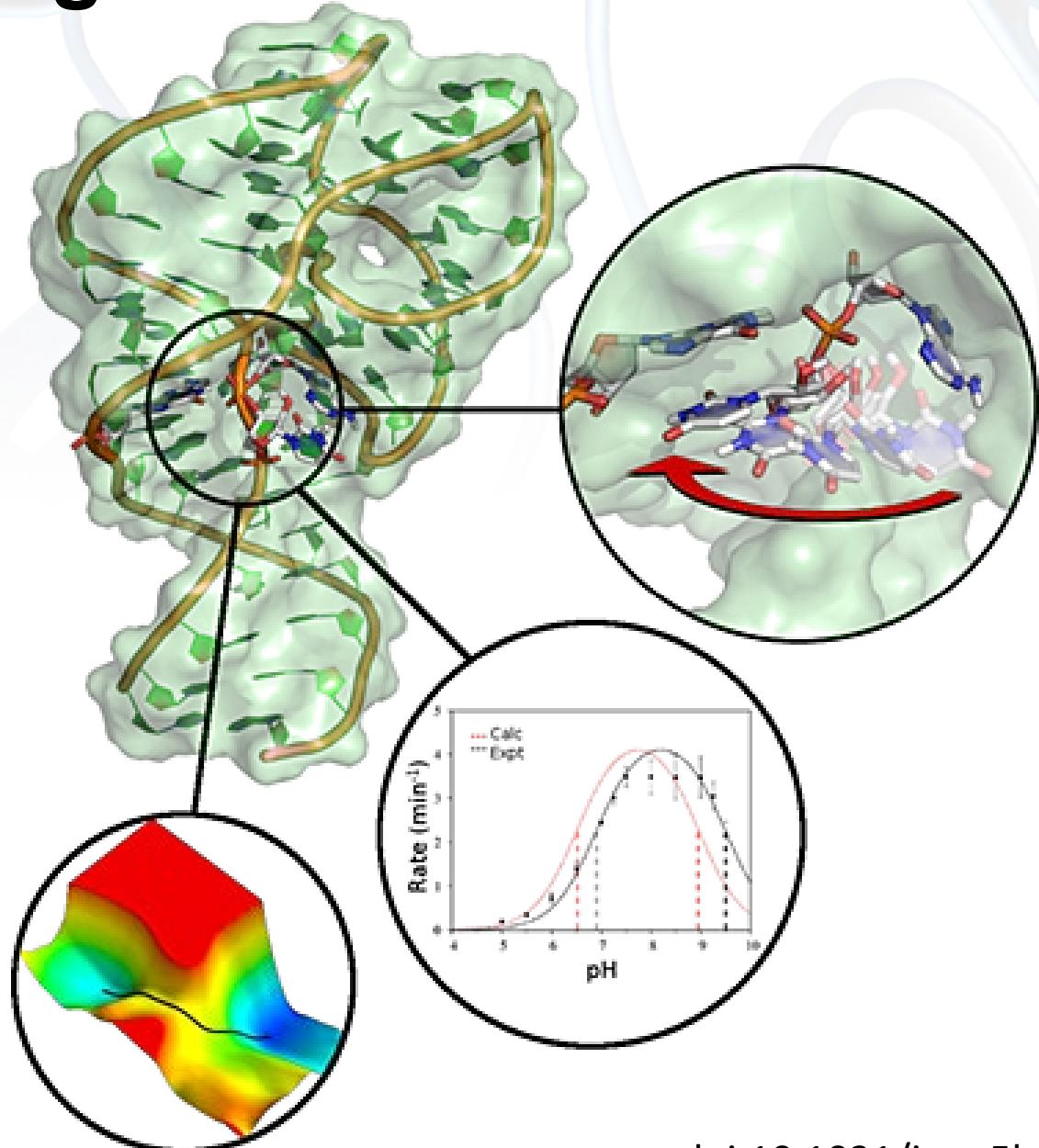
Structural Modeling and Simulation

Uses **physical and chemical** principles to obtain microscopic and macroscopic property quantification

Why use molecular modeling?

Leading reasons:

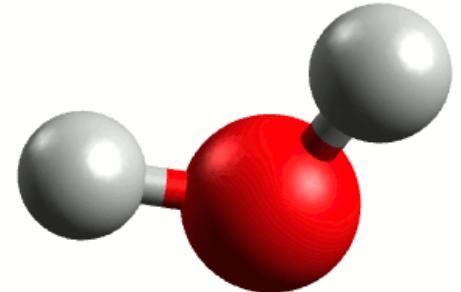
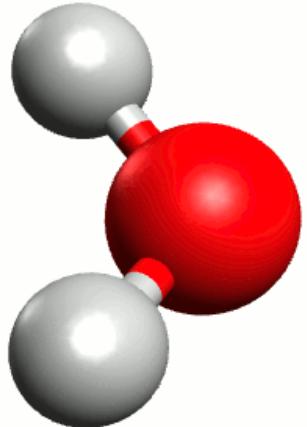
- Support experimental findings
- When you can't see something experimentally
- Prediction



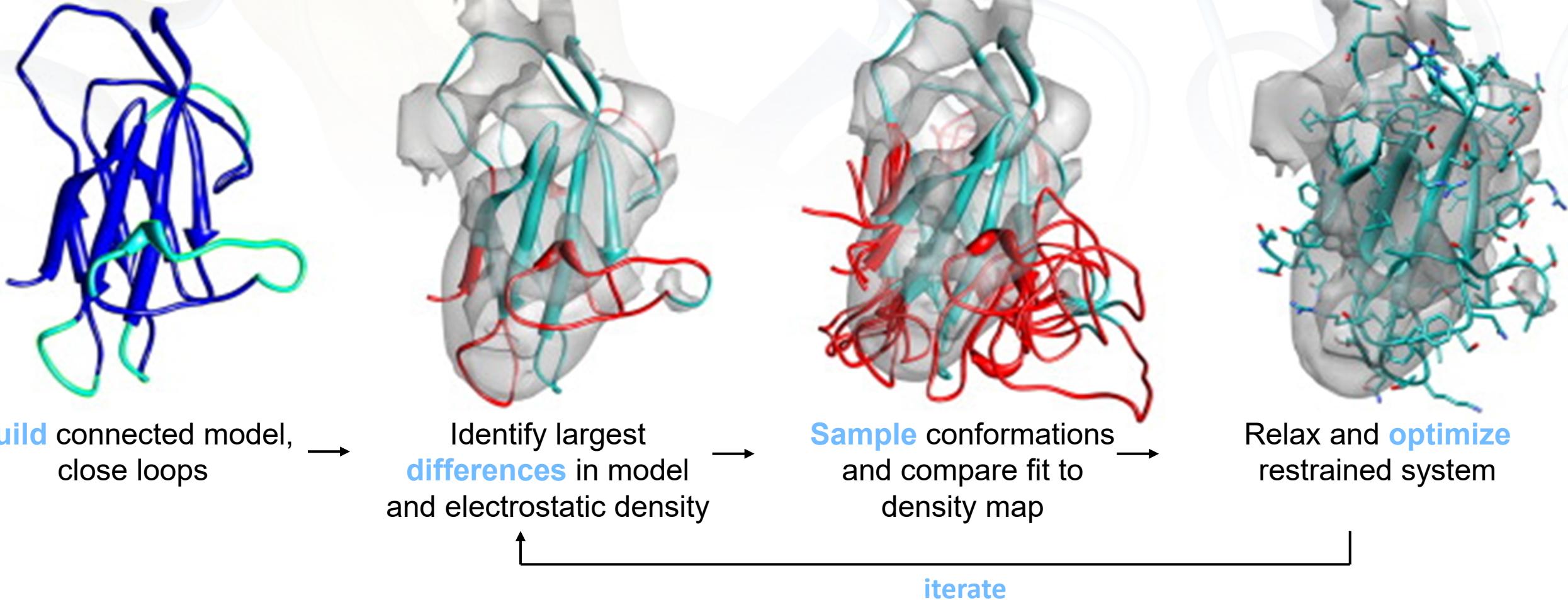
Example: Chemical reactions and intermediates

Things that are hard to see:

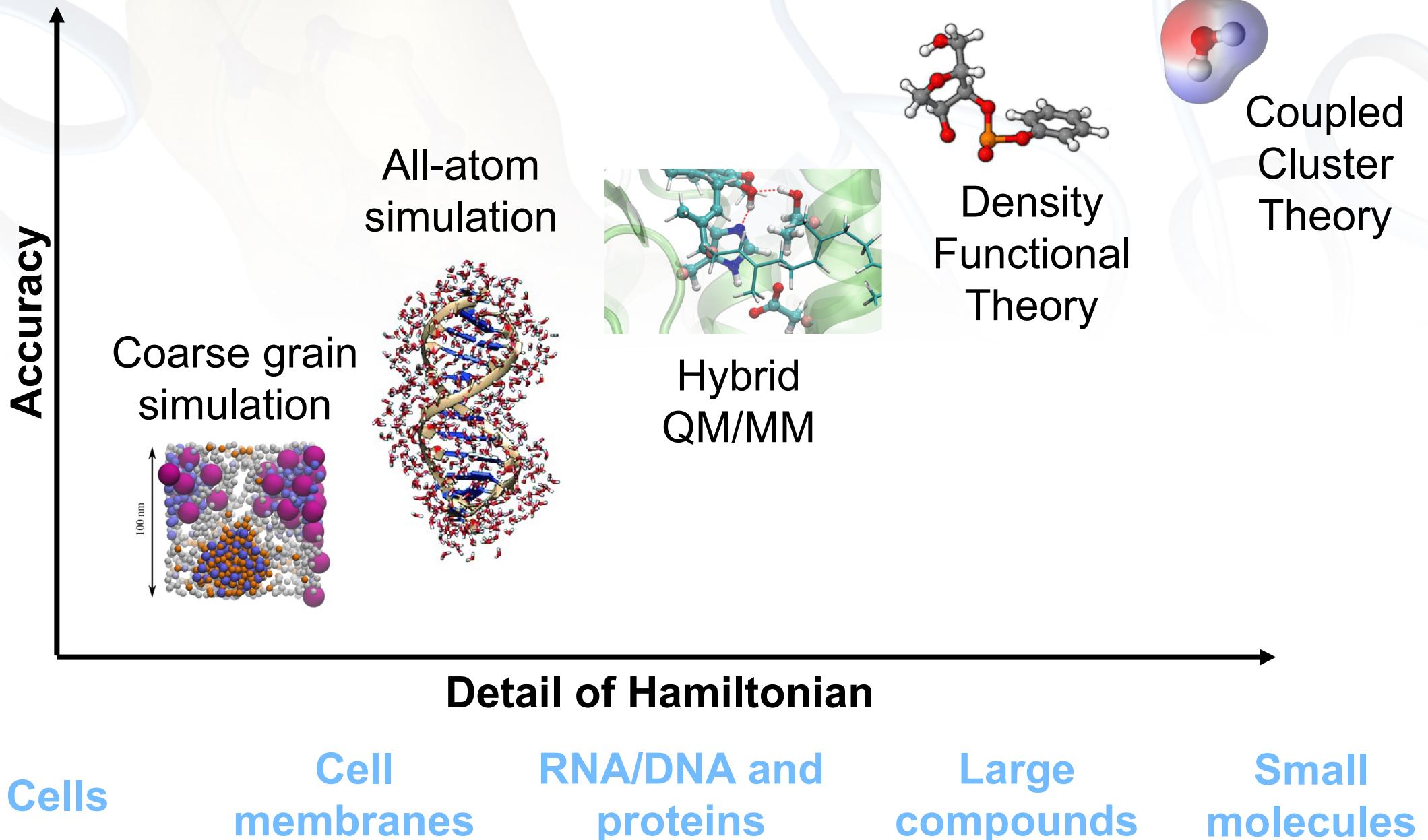
- High-energy configurations
- Transient reaction intermediates
- Unstable complexes and compounds



Example: Protein structure refinement

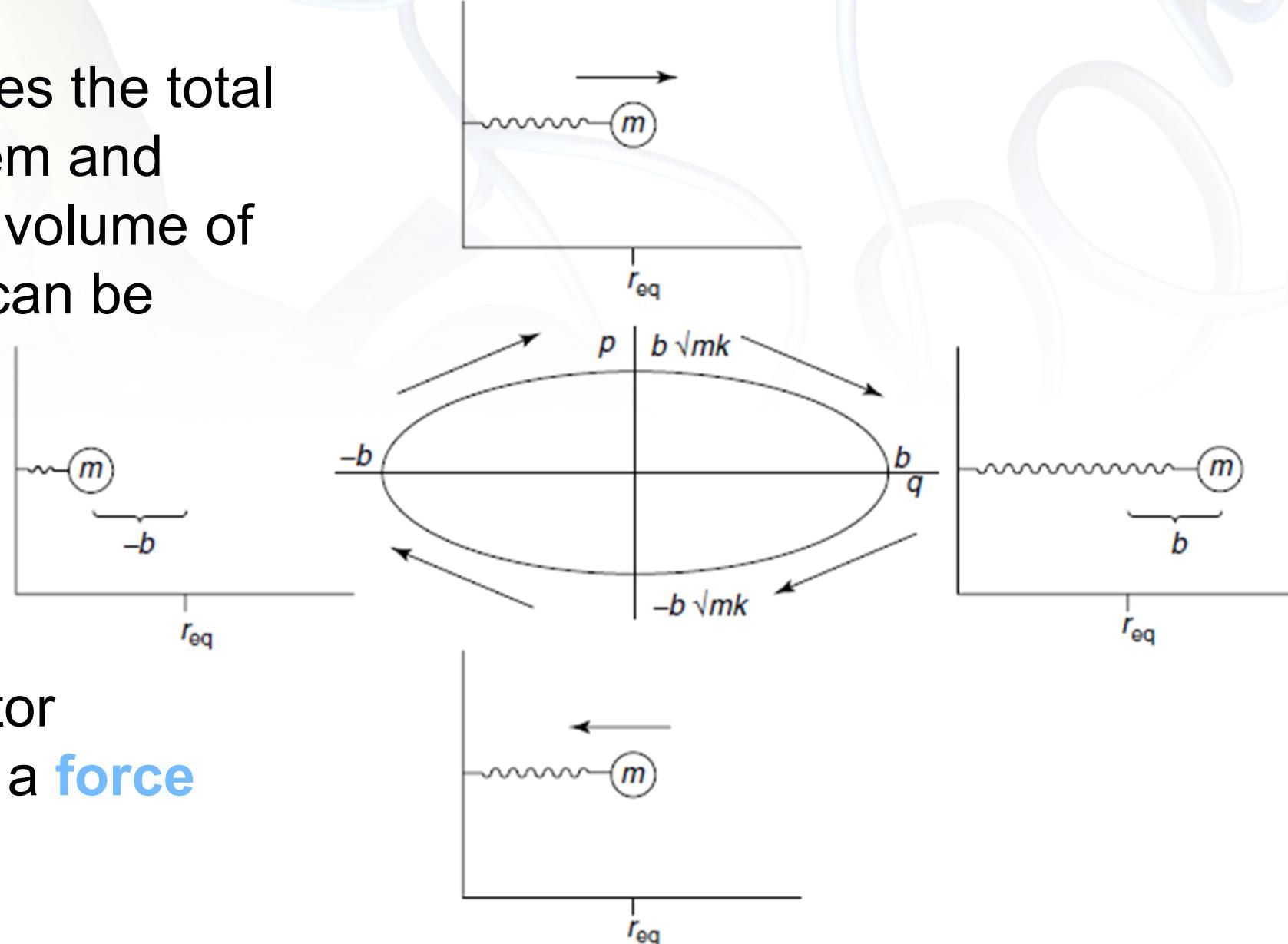


Key concept: Model selection



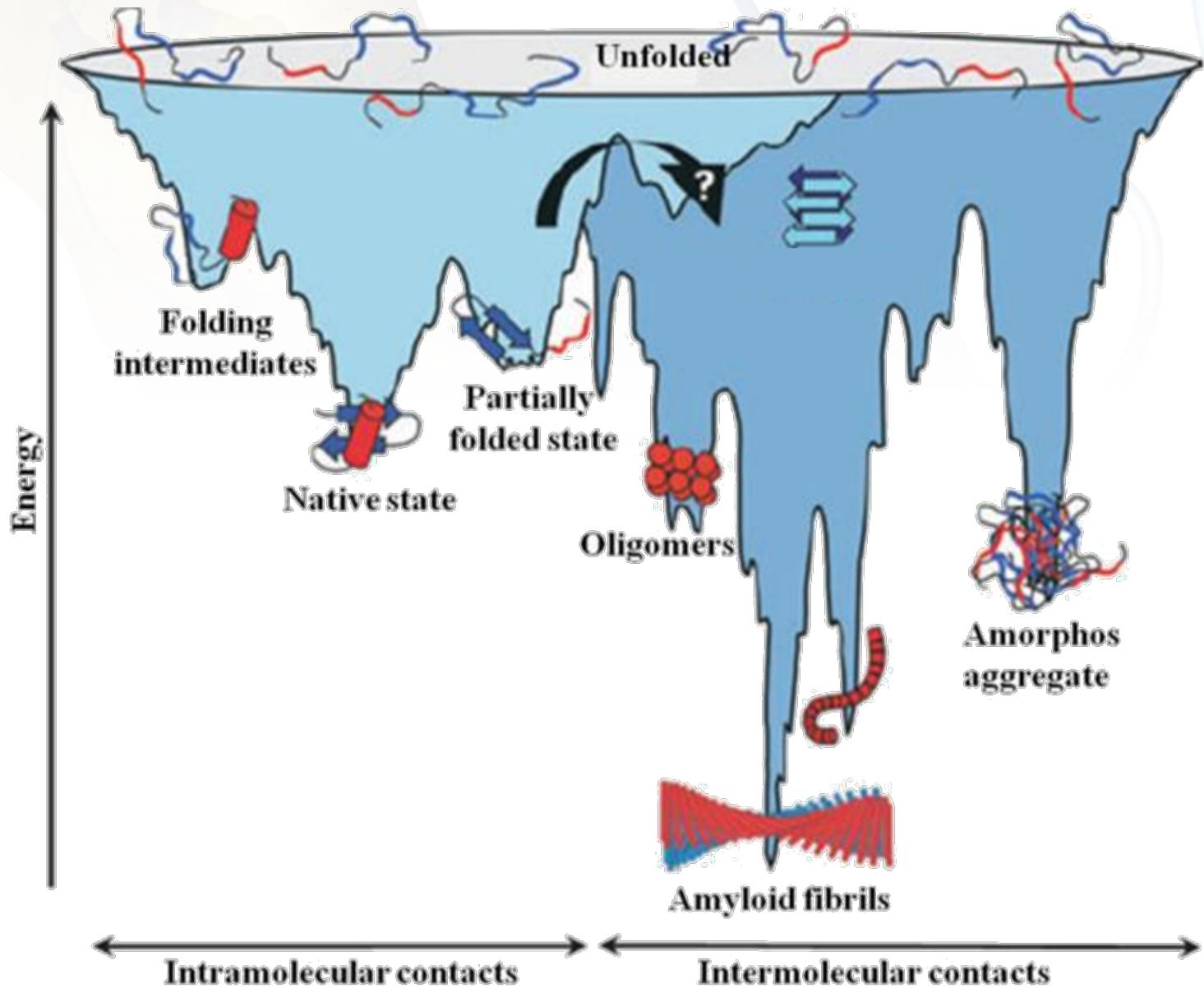
Key concept: What is a Hamiltonian?

A **Hamiltonian** describes the total energy of a given system and dictates the theoretical volume of the **phase space** that can be explored.

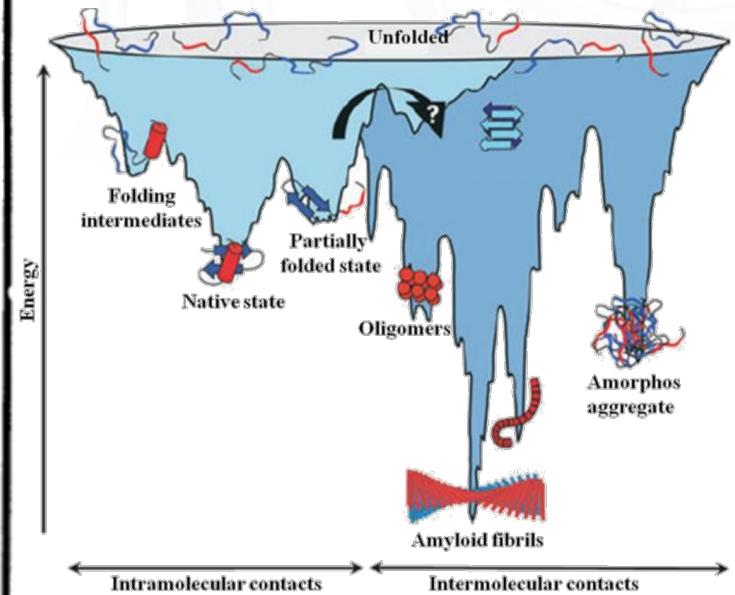
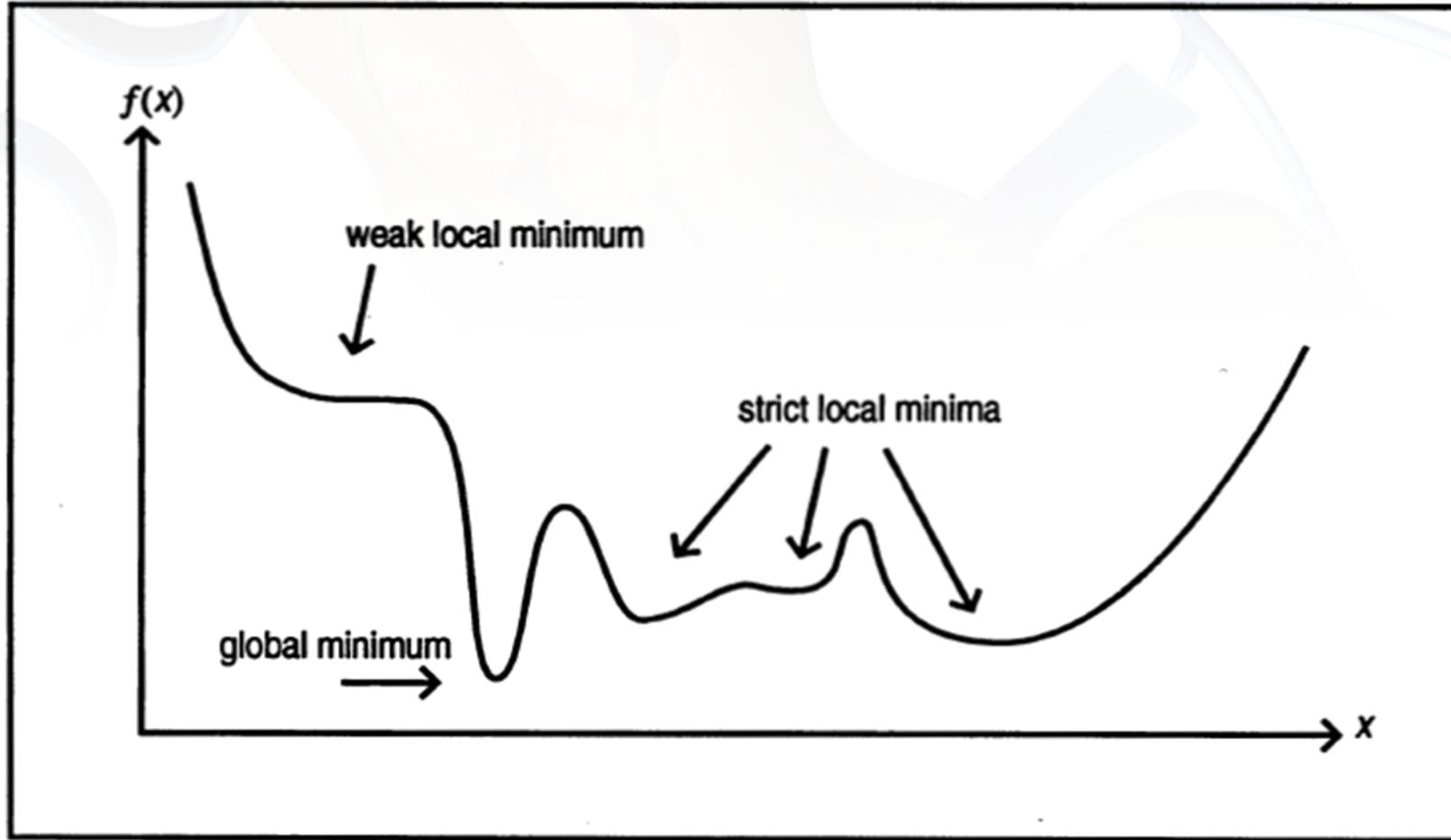


- In QM it is an operator
- In MD it is generally a **force field** (Lecture 2)

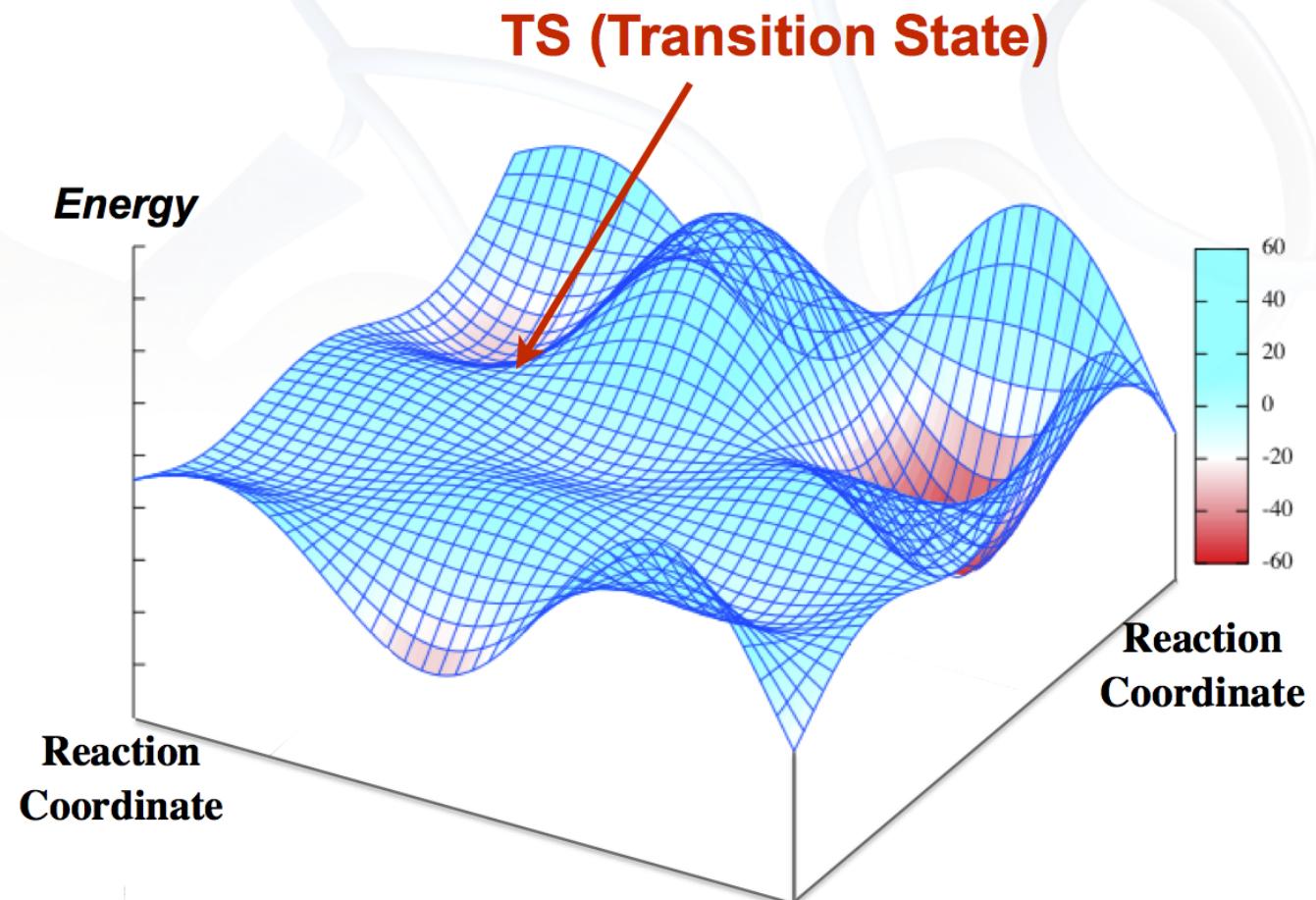
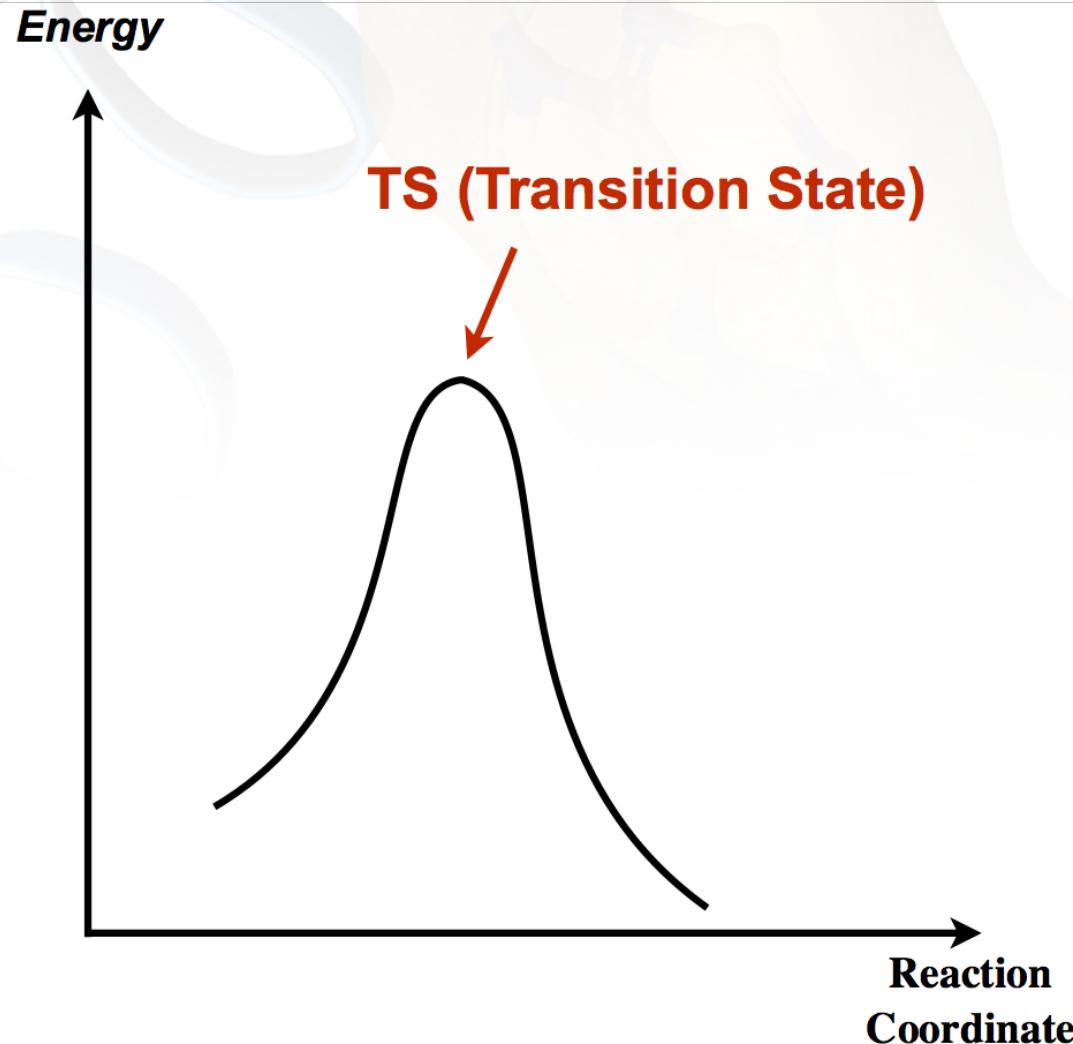
Key concept: Potential energy surfaces



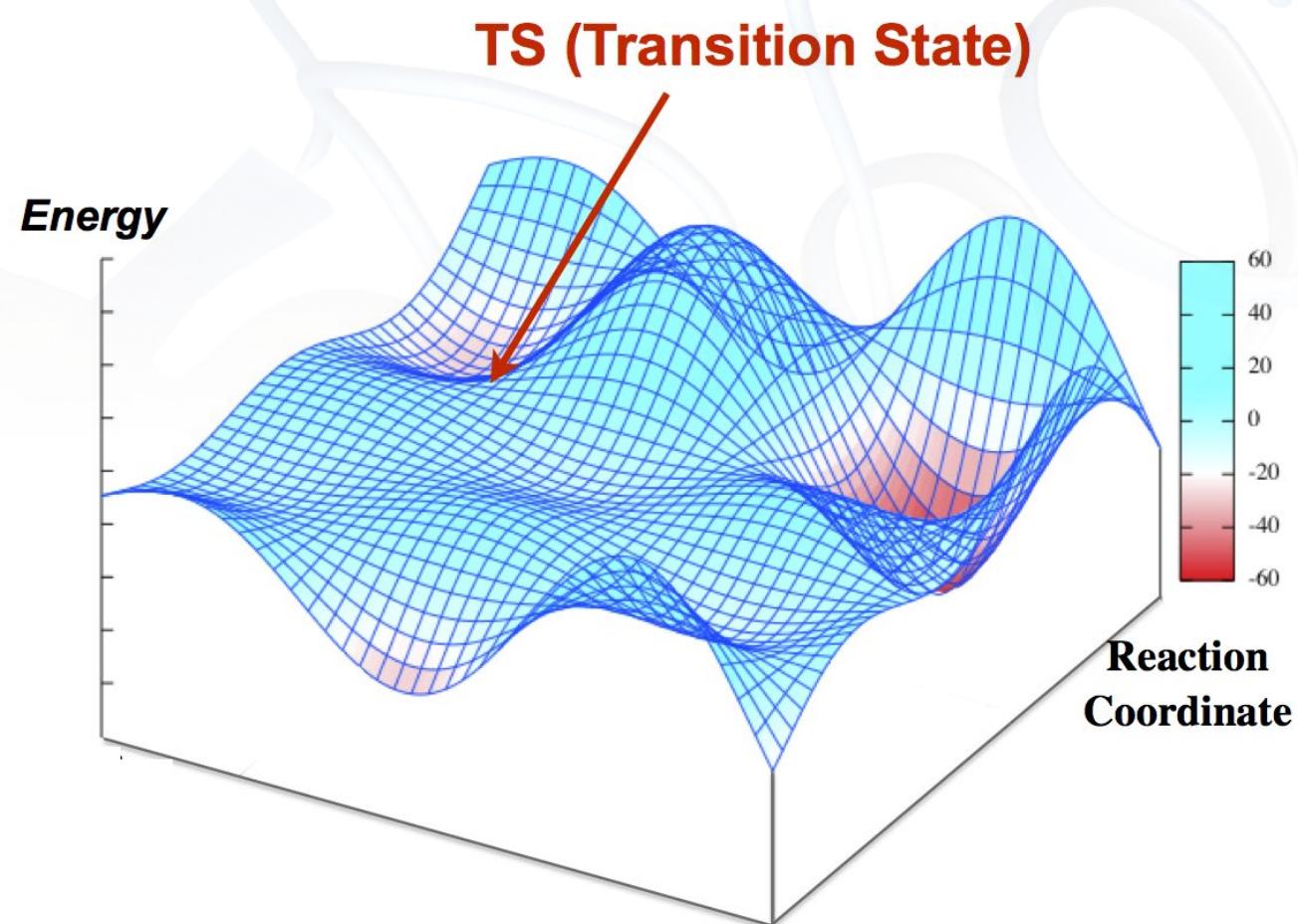
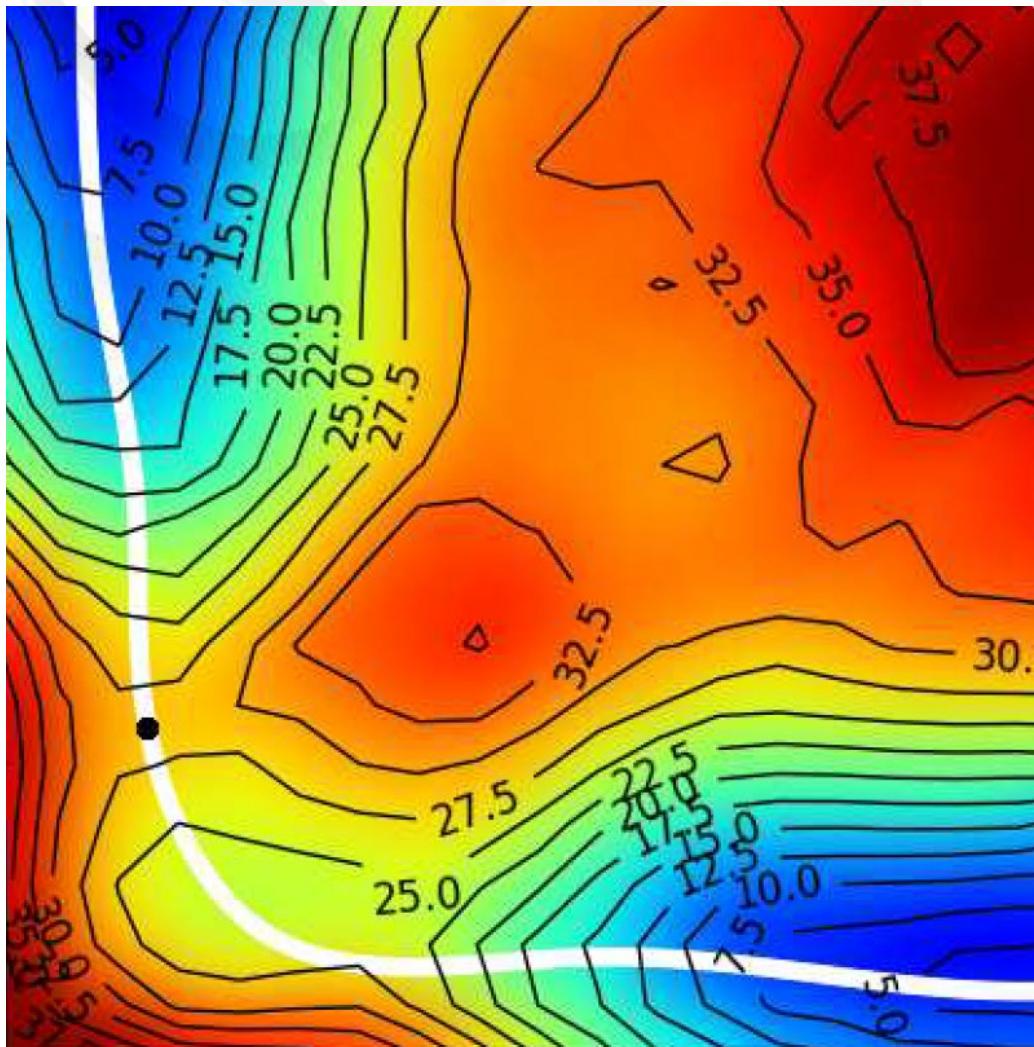
Key concept: Potential energy surfaces



Key concept: What is a transition state?



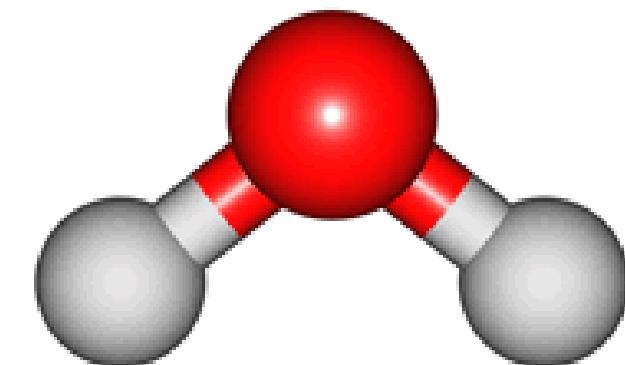
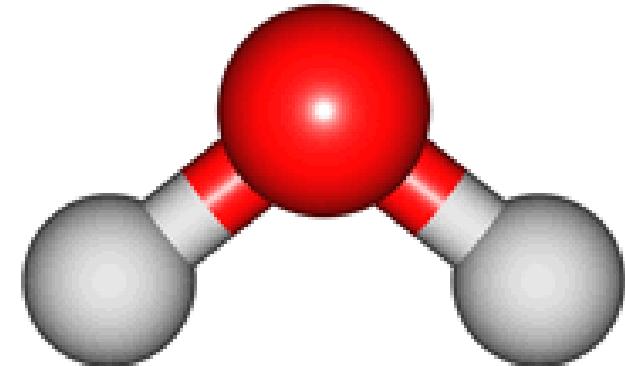
Key concept: What is a transition state?



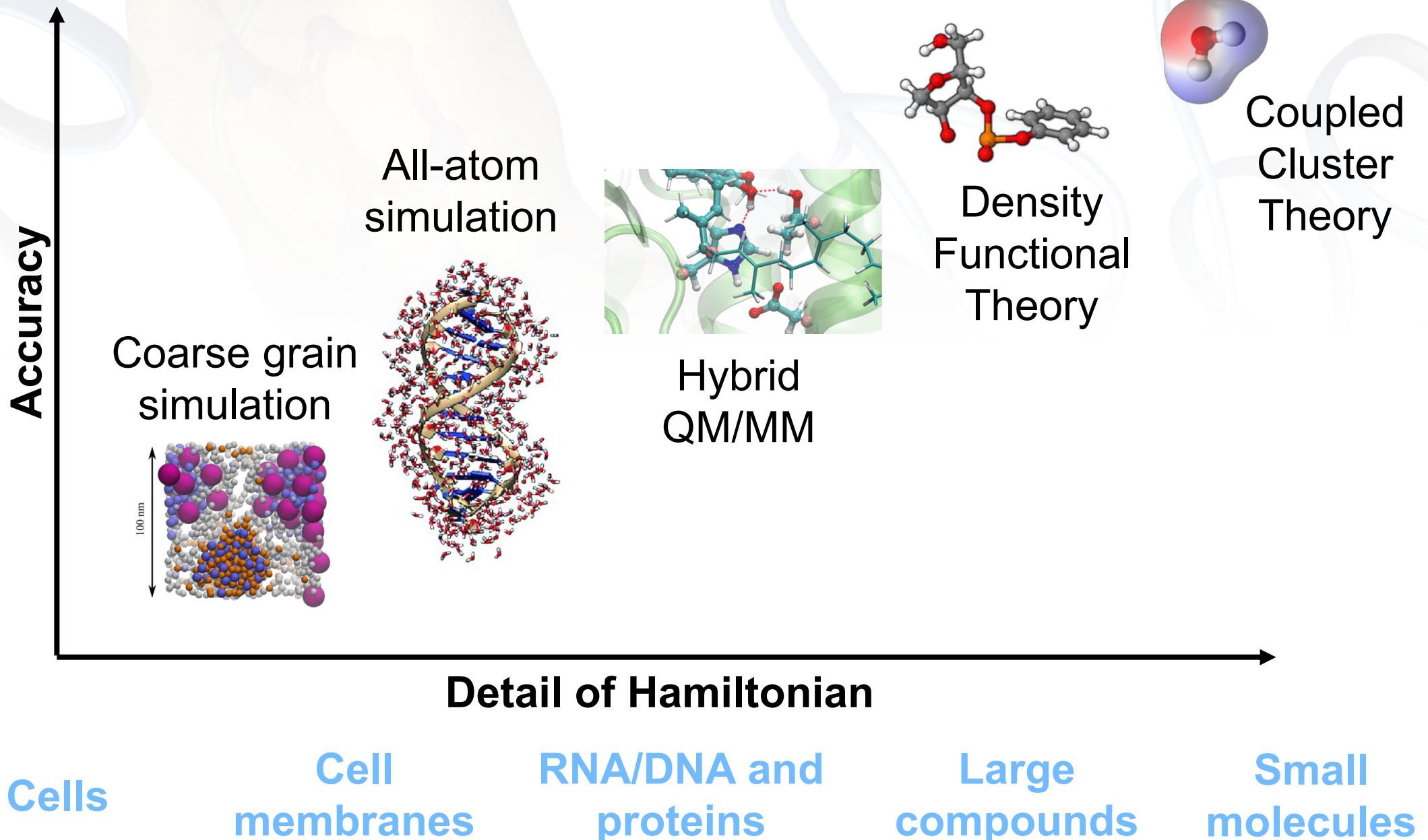
Another concept: Degrees of freedom

A **degree of freedom** is an *independent*, physical parameter in the model system.

The number degrees of freedom in your system dictates the volume of your **phase space**. Each degree of freedom is its own *dimension*.



Key concept: Model selection



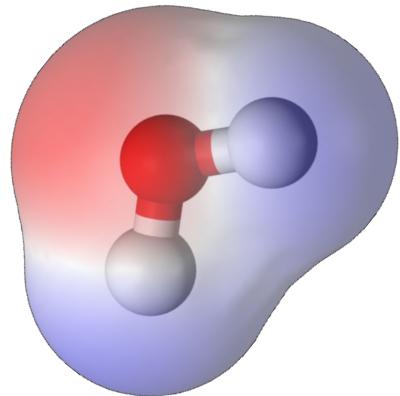
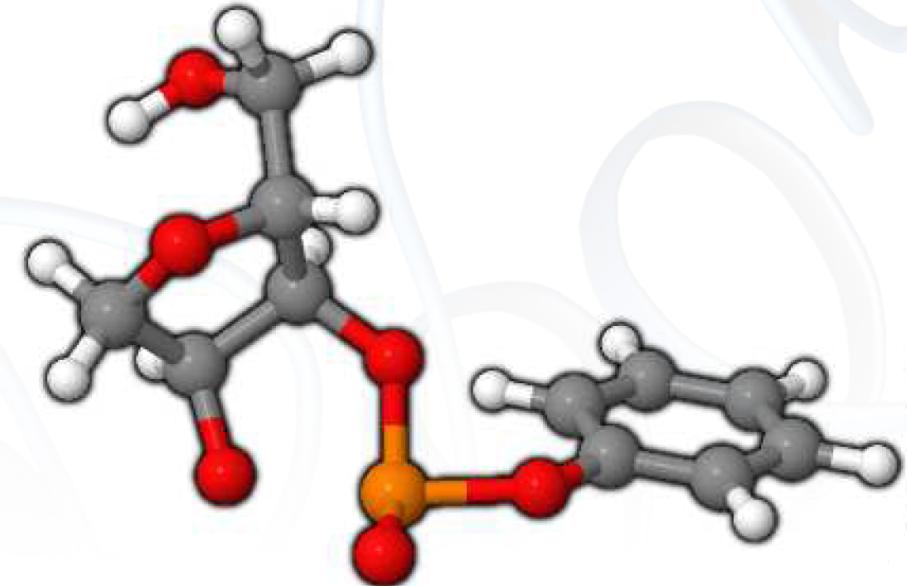
Key concept: Model selection

Models are tools for specific problems:

Key concept: Model selection

Models are tools for specific problems:

- **Quantum Mechanics:** highly accurate chemical reactions, detailed energy landscapes, and electronic configurations

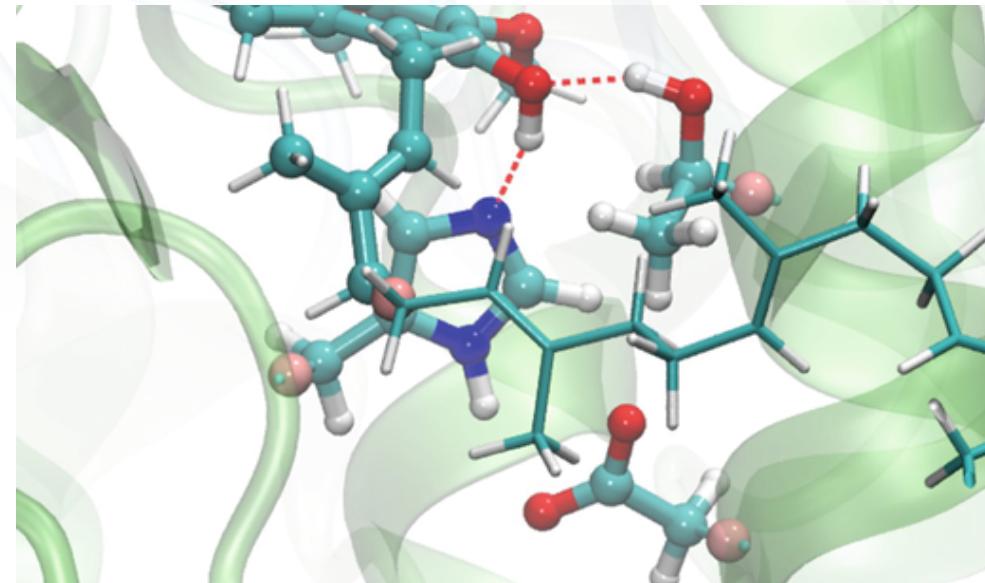


Degree of Freedom: Electrons

Key concept: Model selection

Models are tools for specific problems:

- **Quantum Mechanics:** highly accurate chemical reactions, detailed energy landscapes, and electronic configurations
- **Hybrid QM/MM:** chemical reactions happening in an environment, enzymes

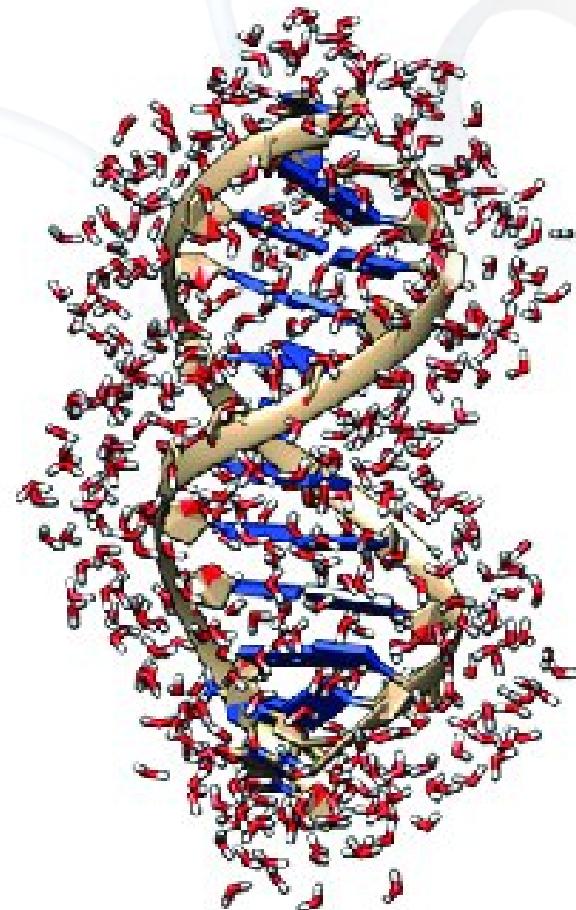


Degree of Freedom: Approximate electrons and atoms

Key concept: Model selection

Models are tools for specific problems:

- **Quantum Mechanics:** highly accurate chemical reactions, detailed energy landscapes, and electronic configurations
- **Hybrid QM/MM:** chemical reactions happening in an environment, enzymes
- **All-atom molecular mechanics:** biomolecular structure, biomolecular configurations, ligand interactions

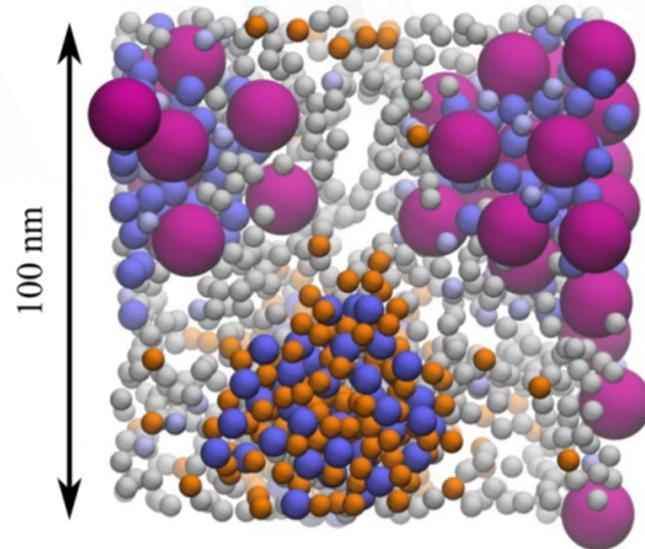


Degree of Freedom: Atoms and molecules

Key concept: Model selection

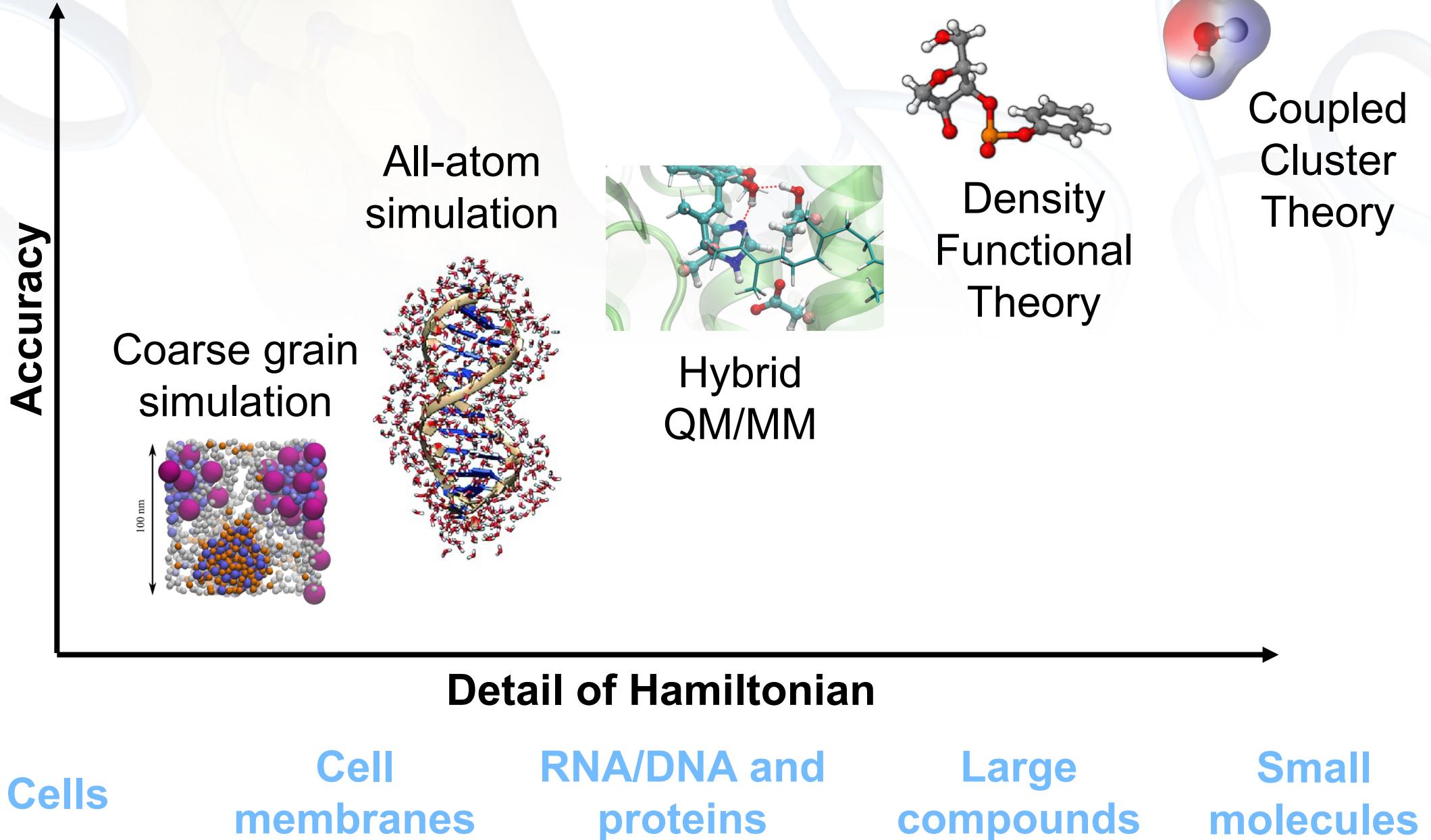
Models are tools for specific problems:

- **Quantum Mechanics:** highly accurate chemical reactions, detailed energy landscapes, and electronic configurations
- **Hybrid QM/MM:** chemical reactions happening in an environment, enzymes
- **All-atom molecular mechanics:** biomolecular structure, biomolecular configurations, ligand interactions
- **Coarse grain simulations:** large proteins or RNA/DNA, cellular environments, membranes, mixtures

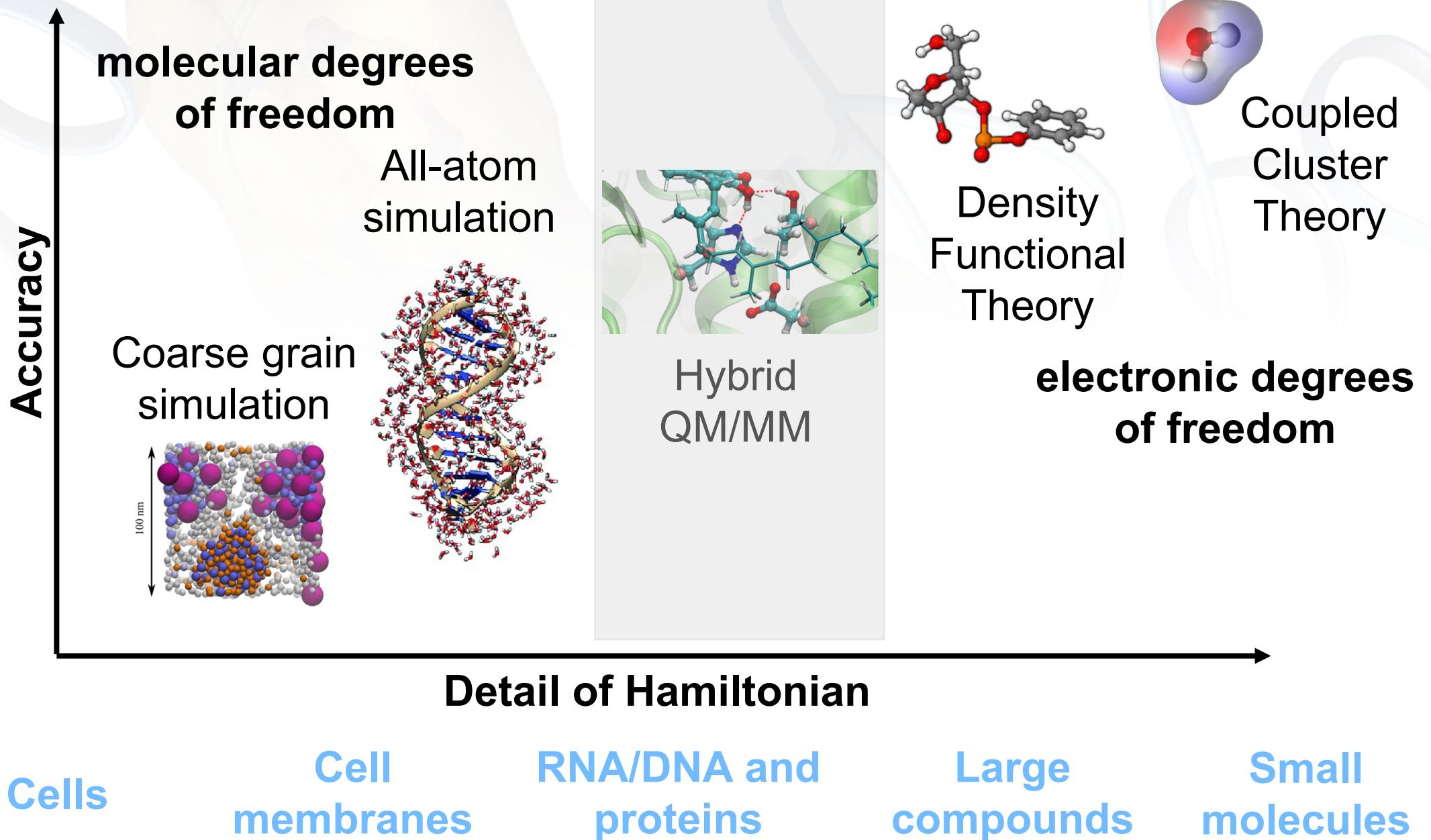


Degree of Freedom: Molecules and particles

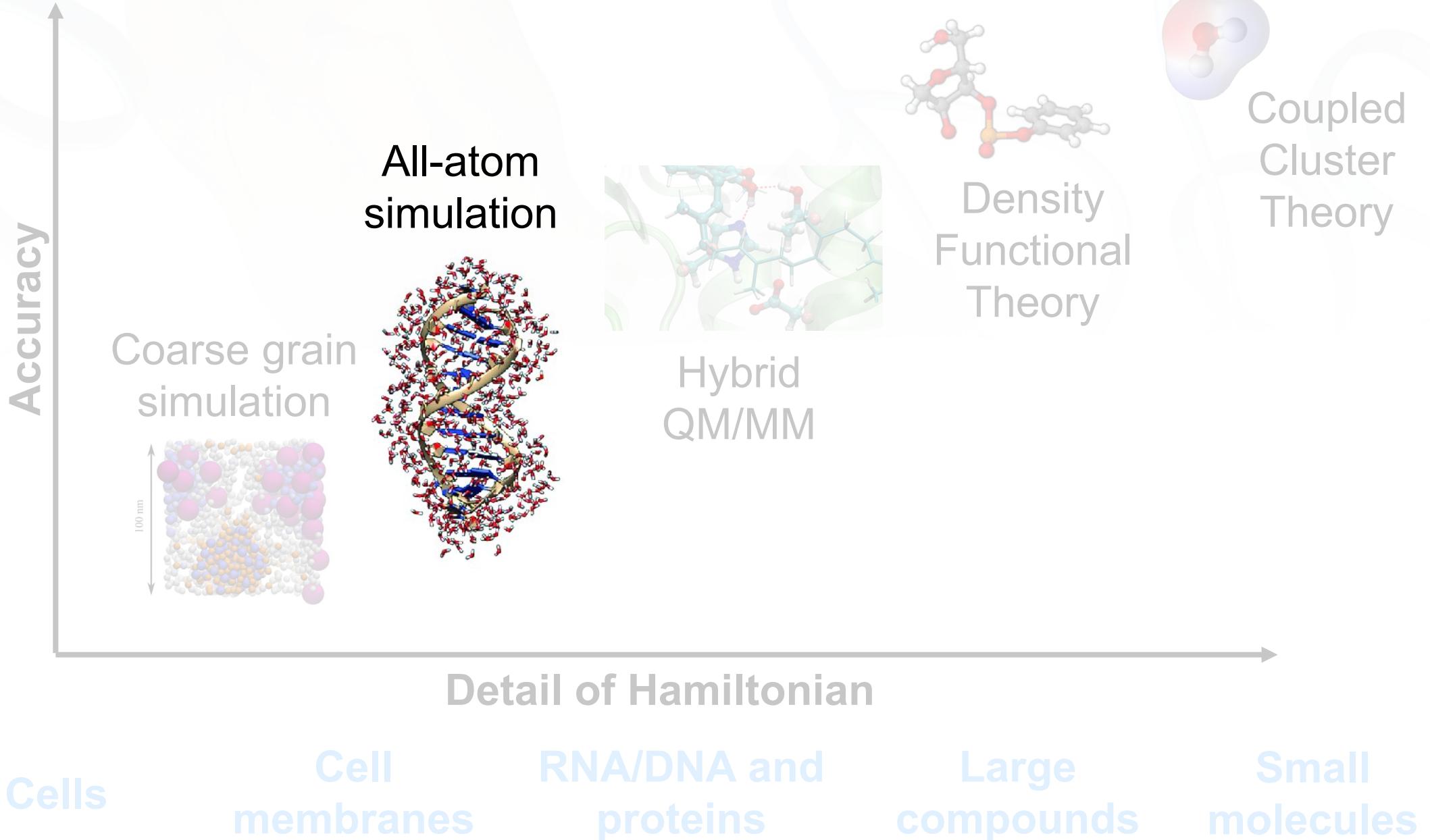
Key concept: Model selection



Key concept: Model selection



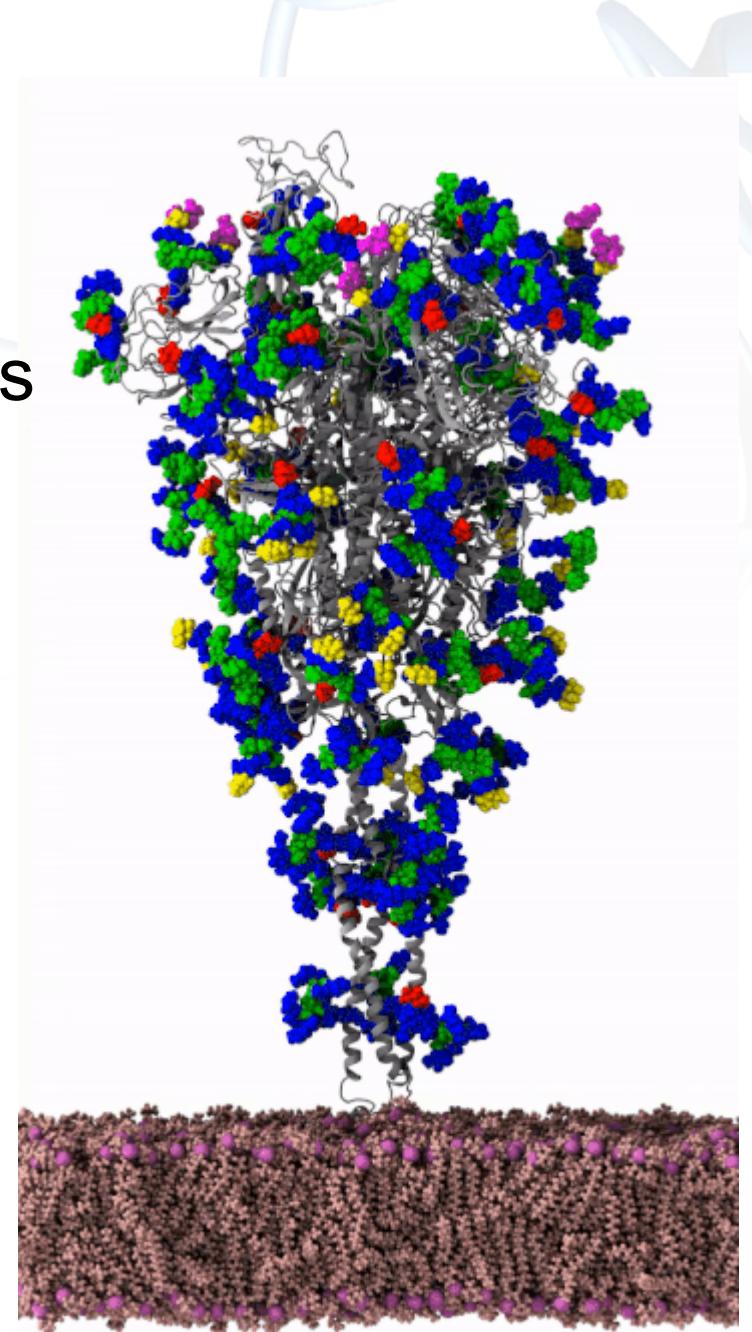
Key concept: Model selection



What is molecular dynamics?

“everything that living things do can be understood in terms of the jigglings and wigglings of atoms.” – **Richard P. Feynman**

$$F_{x,i} = m_i a_{x,i}$$

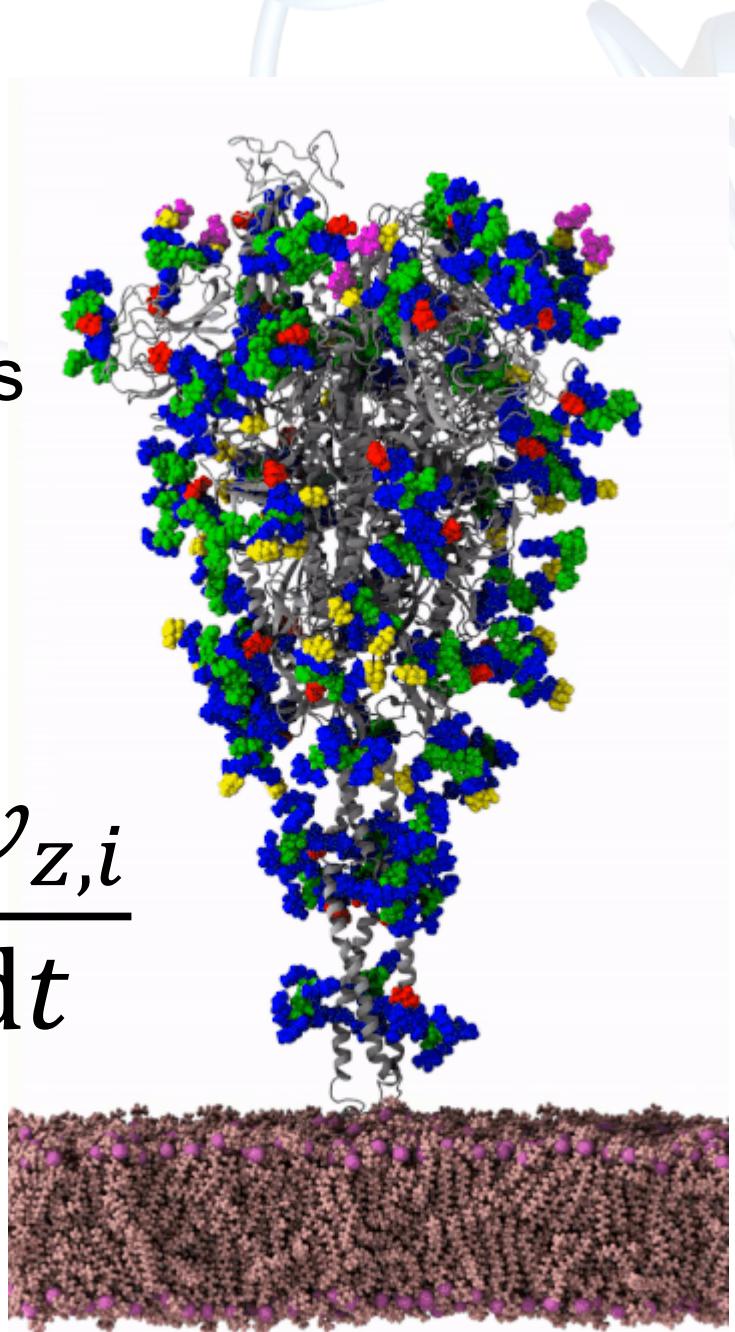


What is molecular dynamics?

“everything that living things do can be understood in terms of the jigglings and wigglings of atoms.” – **Richard P. Feynman**

$$F_{x,i} = m_i a_{x,i}$$

$$a_{x,i} = \frac{dv_{x,i}}{dt} \quad a_{y,i} = \frac{dv_{y,i}}{dt} \quad a_{z,i} = \frac{dv_{z,i}}{dt}$$

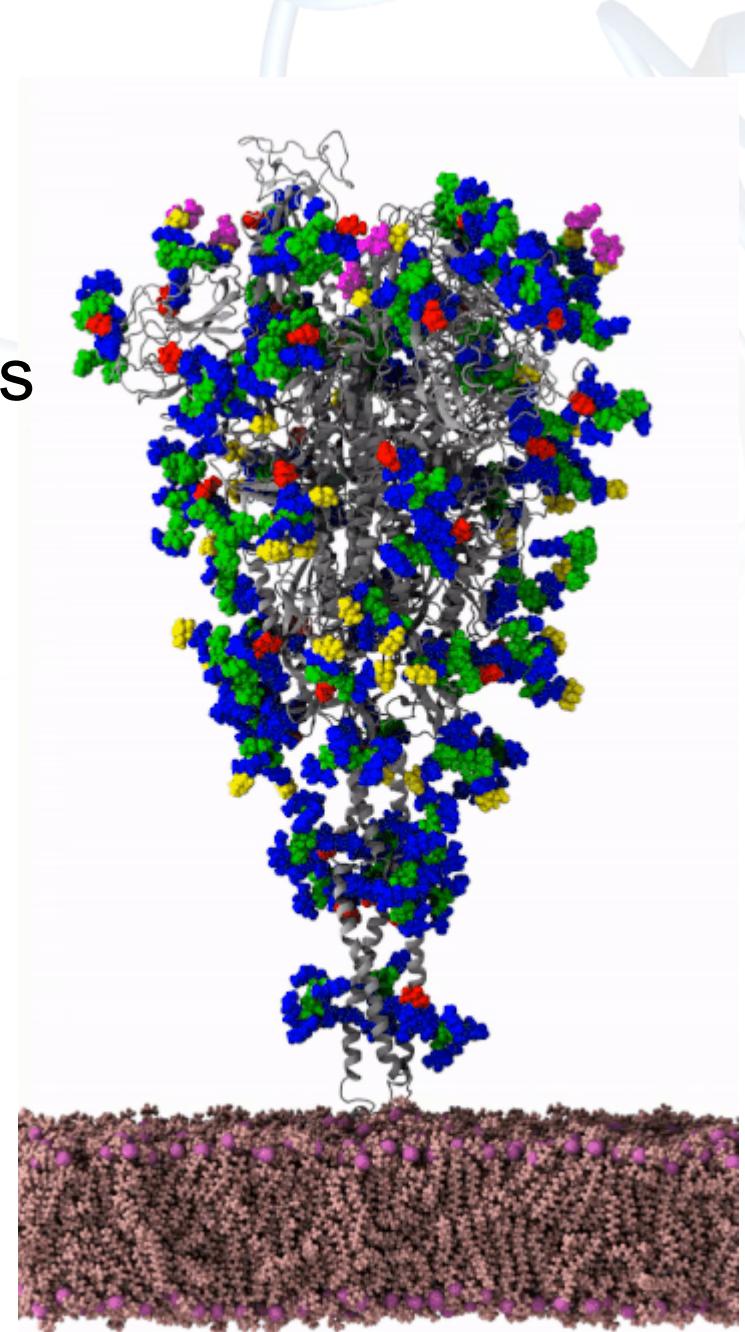


What is molecular dynamics?

“everything that living things do can be understood in terms of the jigglings and wigglings of atoms.” – **Richard P. Feynman**

$$F_{x,i} = m_i a_{x,i}$$

$$a_i = \frac{d\boldsymbol{v}_i}{dt}$$

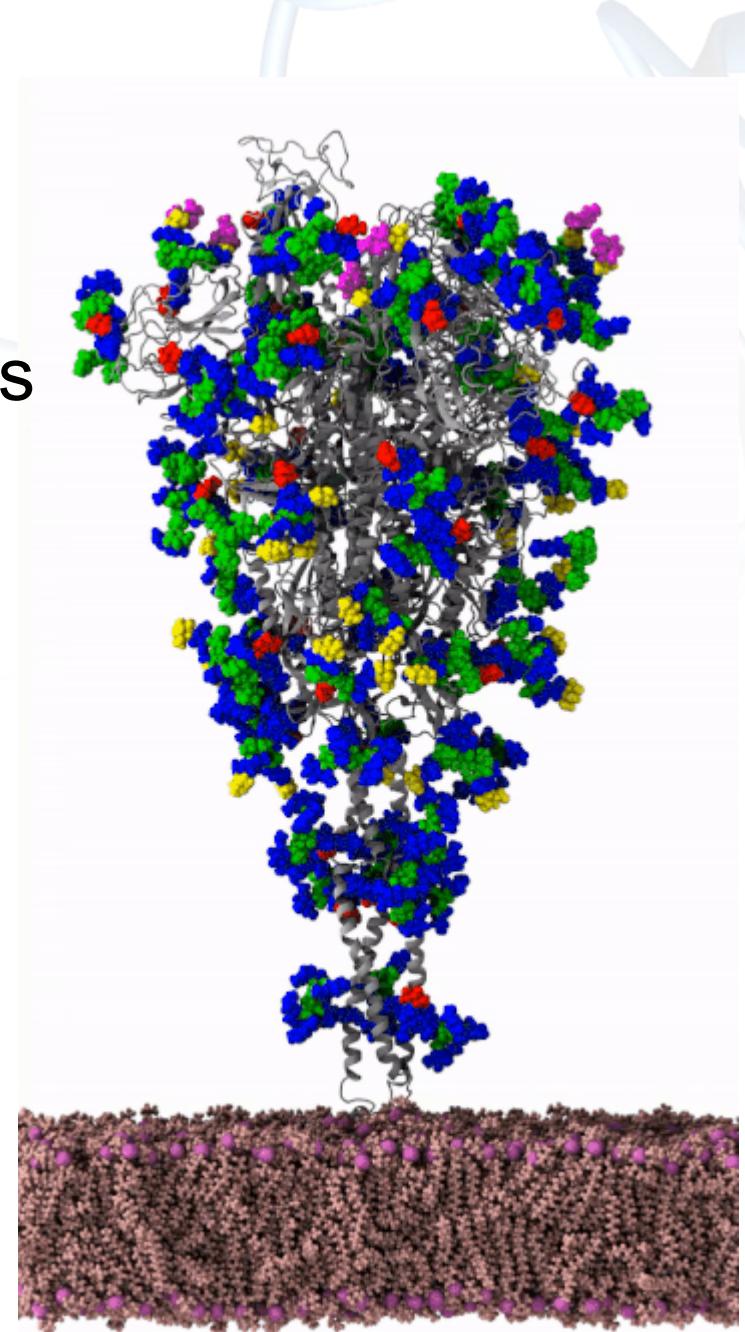


What is molecular dynamics?

“everything that living things do can be understood in terms of the jigglings and wigglings of atoms.” – **Richard P. Feynman**

$$F_{x,i} = m_i a_{x,i}$$

$$a_i = \frac{d\boldsymbol{v}_i}{dt} = \frac{d^2\boldsymbol{r}_i}{dt^2}$$

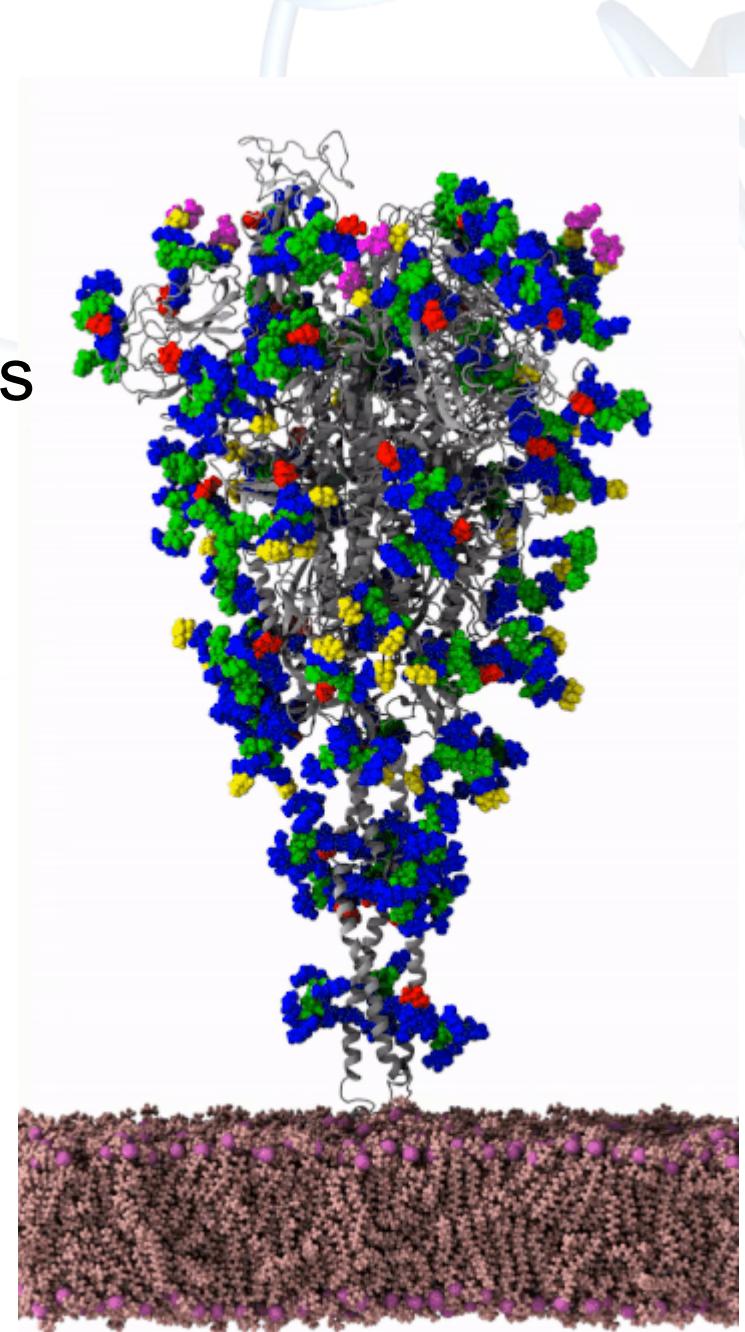


What is molecular dynamics?

“everything that living things do can be understood in terms of the jigglings and wigglings of atoms.” – **Richard P. Feynman**

$$F_{x,i} = m_i a_{x,i}$$

$$\ddot{\mathbf{r}}_i = \frac{d\dot{\mathbf{r}}_i}{dt} = \frac{d^2\mathbf{r}_i}{dt^2}$$

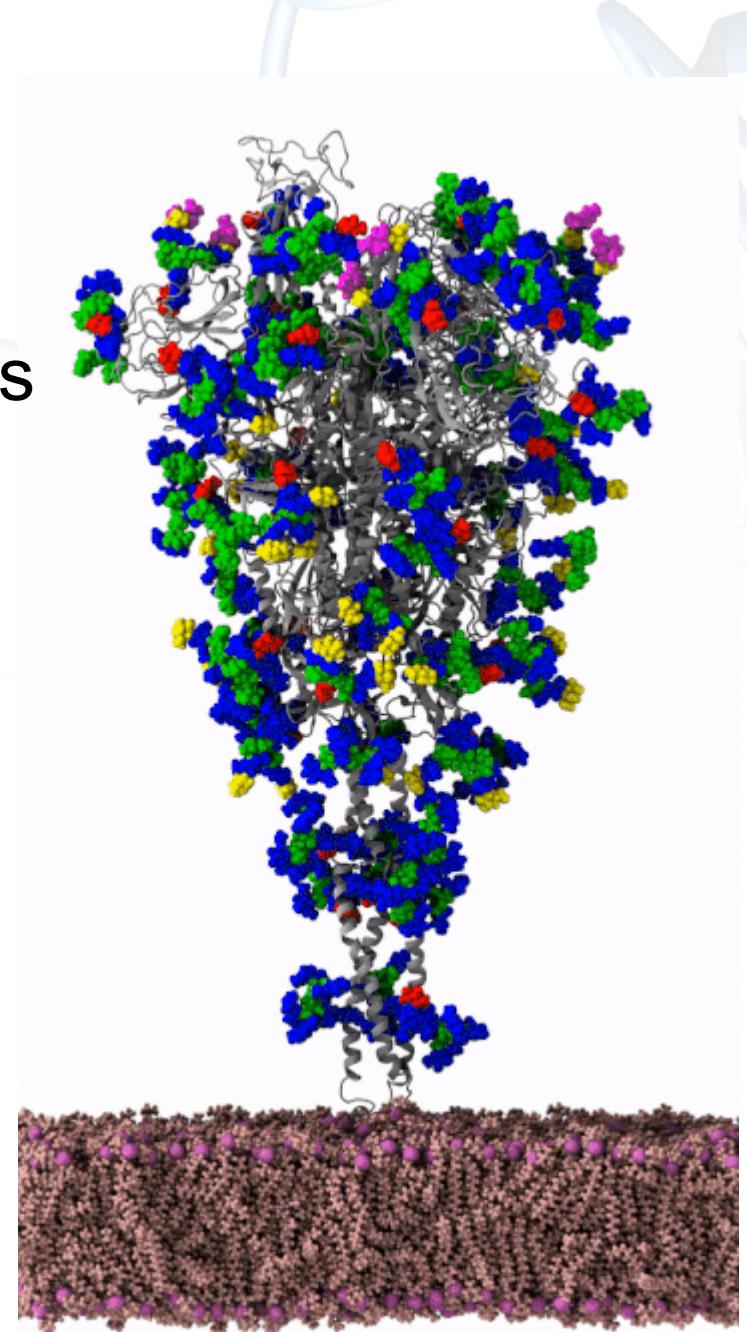


What is molecular dynamics?

“everything that living things do can be understood in terms of the jigglings and wigglings of atoms.” – **Richard P. Feynman**

$$F_{x,i} = m_i a_{x,i}$$

$$\mathbf{F}_i = m_i \ddot{\mathbf{r}}_i = \dot{\mathbf{p}}_i$$



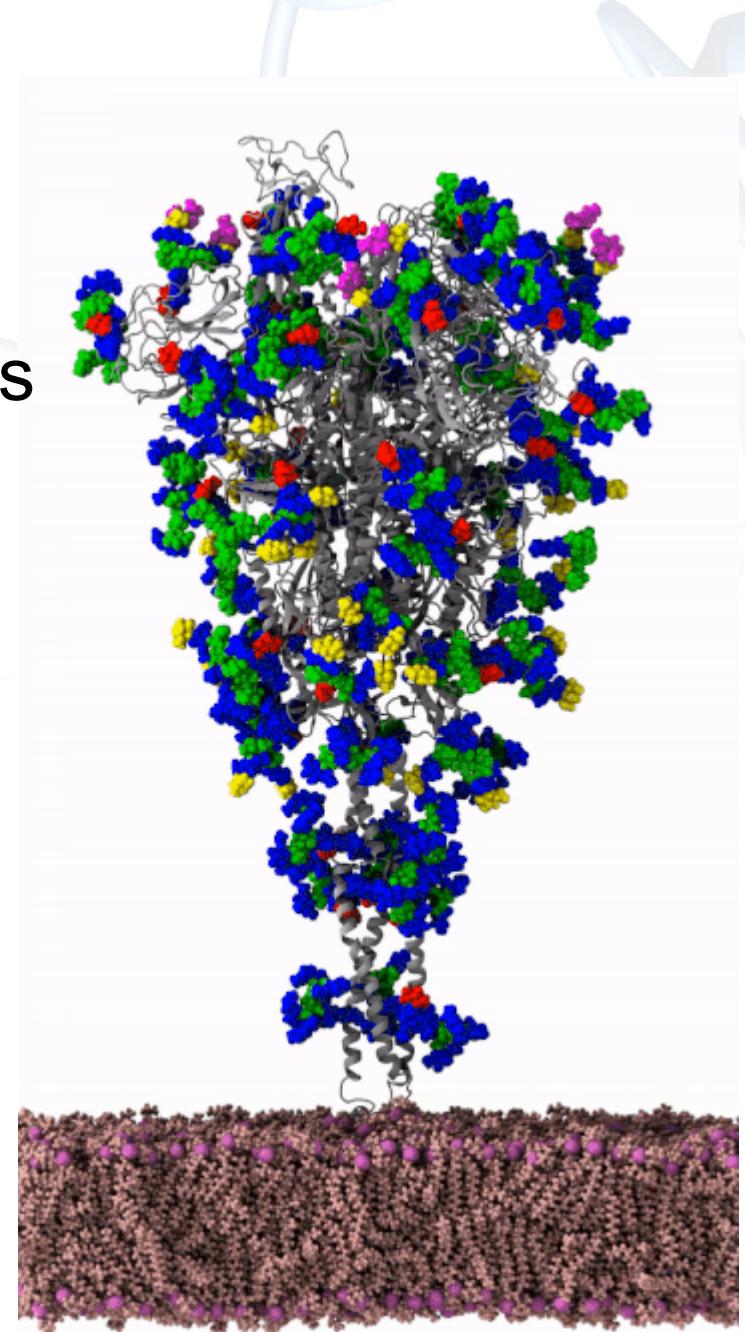
What is molecular dynamics?

“everything that living things do can be understood in terms of the jigglings and wigglings of atoms.” – **Richard P. Feynman**

$$F_{x,i} = m_i a_{x,i}$$

“Bruh, that's just kinetic.” – *Errybody*

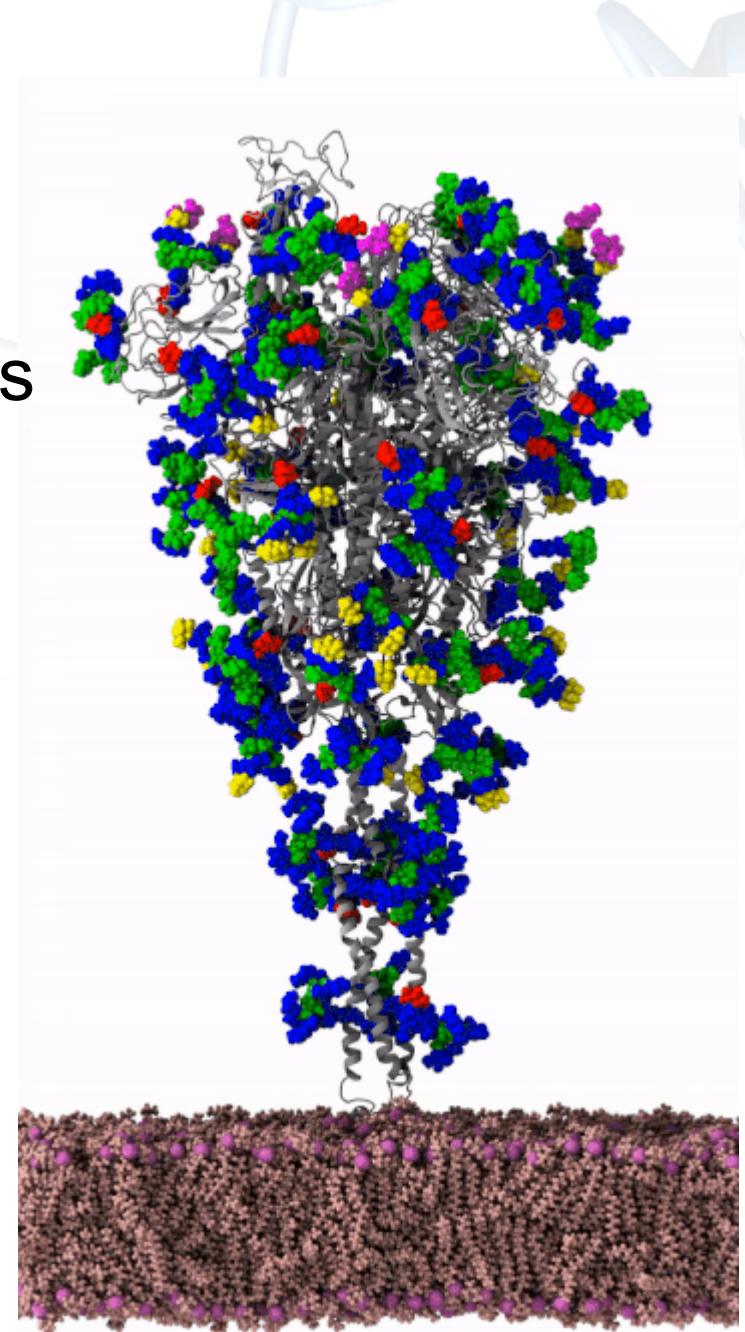
<https://twitter.com/RommieAmaro/status/1241810976866840577>



What is molecular dynamics?

“everything that living things do can be understood in terms of the jigglings and wigglings of atoms.” – **Richard P. Feynman**

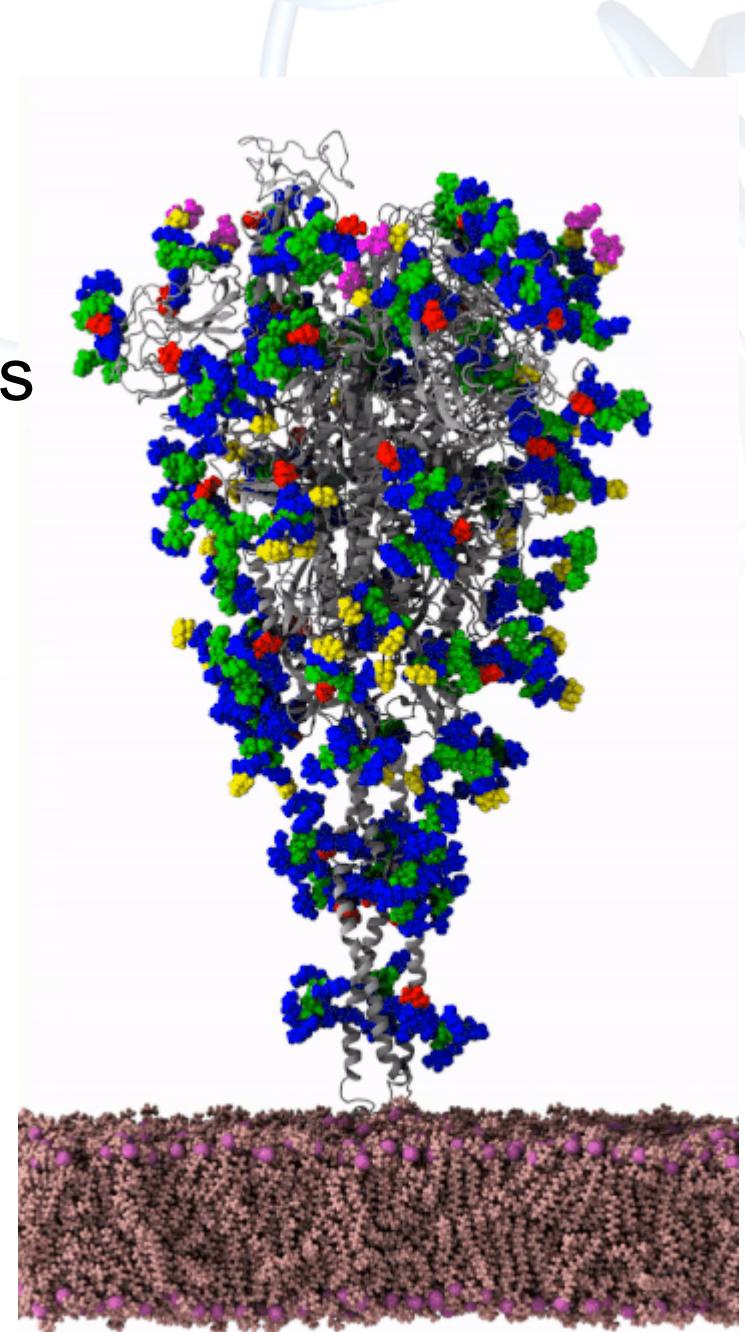
$$\mathbf{F}_i = \dot{\mathbf{p}}_i - \frac{\partial U(\mathbf{r}^N)}{\partial \mathbf{r}_i}$$



What is molecular dynamics?

“everything that living things do can be understood in terms of the jigglings and wigglings of atoms.” – **Richard P. Feynman**

$$\mathbf{F}_i = \dot{\mathbf{p}}_i - \frac{\partial U(\mathbf{r}^N)}{\partial \mathbf{r}_i}$$



Key concept: Ergodic hypothesis

The **ergodic hypothesis** is that the time average of sampling some observable is equal to the statistical ensemble average of that observable in the limit that time approaches infinity.

Key concept: Ergodic hypothesis

The **ergodic hypothesis** is that the time average of sampling some observable is equal to the statistical ensemble average of that observable in the limit that time approaches infinity.

$$A_{obs} = \langle A \rangle_{\text{time}} = \langle A \rangle_{\text{ensemble}}$$

Where obs is the ‘observed’ value of attribute A, $\langle A \rangle$ is the ‘average’ of attribute A. So ‘time average’ and ‘ensemble average’.

Key concept: Ergodic hypothesis

The **ergodic hypothesis** is that the time average of sampling some observable is equal to the statistical ensemble average of that observable in the limit that time approaches infinity.

$$A_{obs} = \langle A \rangle_{\text{time}} = \langle A \rangle_{\text{ensemble}}$$

Where:

$$\langle A \rangle_{\text{time}} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(\mathbf{p}^N(t), \mathbf{r}^N(t)) dt$$

And:

$$\langle A \rangle_{\text{ensemble}} = \iint A(\mathbf{p}^N, \mathbf{r}^N) \rho(\mathbf{p}^N, \mathbf{r}^N) d\mathbf{p}^N d\mathbf{r}^N$$

Key concept: Ergodic hypothesis

The **ergodic hypothesis** is that the time average of sampling some observable is equal to the statistical ensemble average of that observable in the limit that time approaches infinity.

$$A_{obs} = \langle A \rangle_{\text{time}} = \langle A \rangle_{\text{ensemble}}$$

Where:

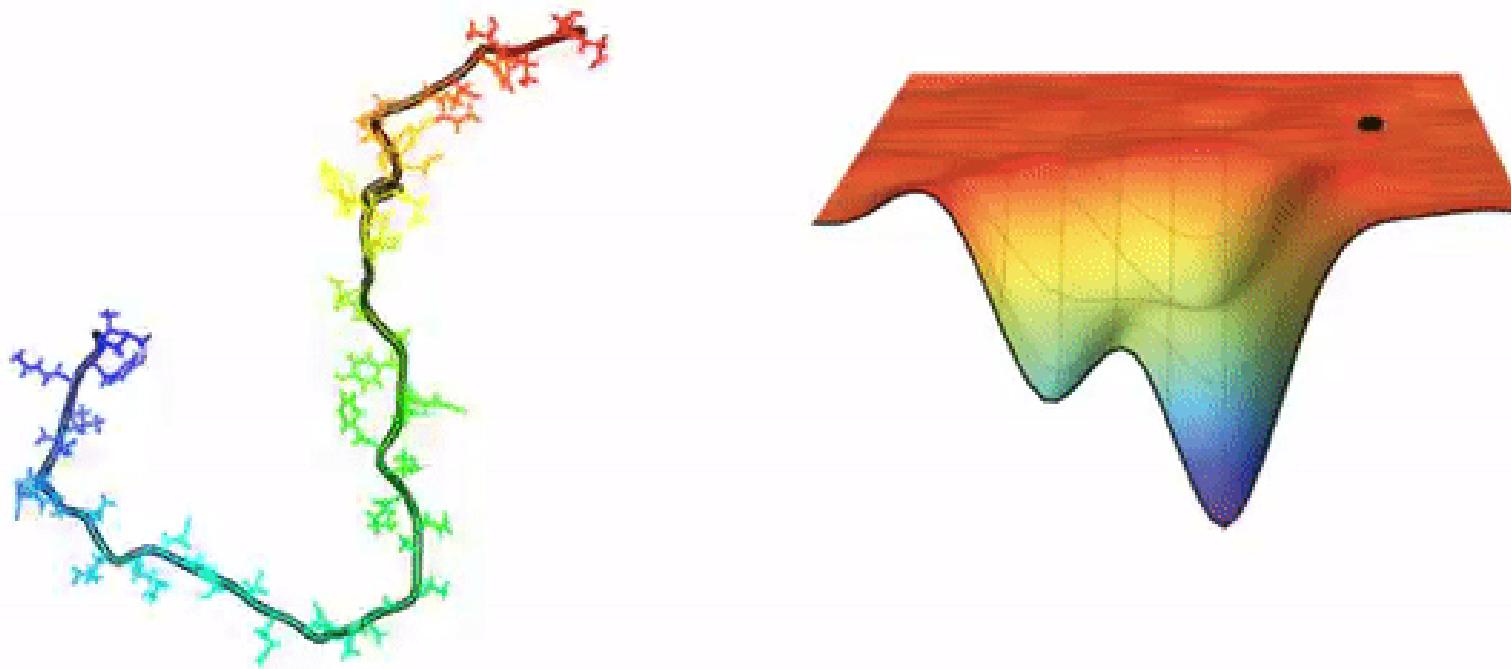
$$\langle A \rangle_{\text{time}} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A(\mathbf{p}^N(t), \mathbf{r}^N(t)) dt$$

And:

$$\langle A \rangle_{\text{ensemble}} = \iint A(\mathbf{p}^N, \mathbf{r}^N) \frac{\exp[-H(\mathbf{p}^N, \mathbf{r}^N)/k_B T]}{\iint \exp[-H(\mathbf{p}^N, \mathbf{r}^N)/k_B T] d\mathbf{p}^N d\mathbf{r}^N} d\mathbf{p}^N d\mathbf{r}^N$$

Key concept: Ergodic hypothesis

The **ergodic hypothesis** is that the time average of sampling some observable is equal to the statistical ensemble average of that observable in the limit that time approaches infinity.



What is an ensemble?

An **ensemble** is a series of system constraints imposed on the model system. It is important for the underlying assumptions used to run the simulation.

Canonical (N,V,T)

Temperature dependent properties, like folding

Isobaric-isothermal (N,P,T)

Biological systems where the volume is unknown

Microcanonical (N,V,E)

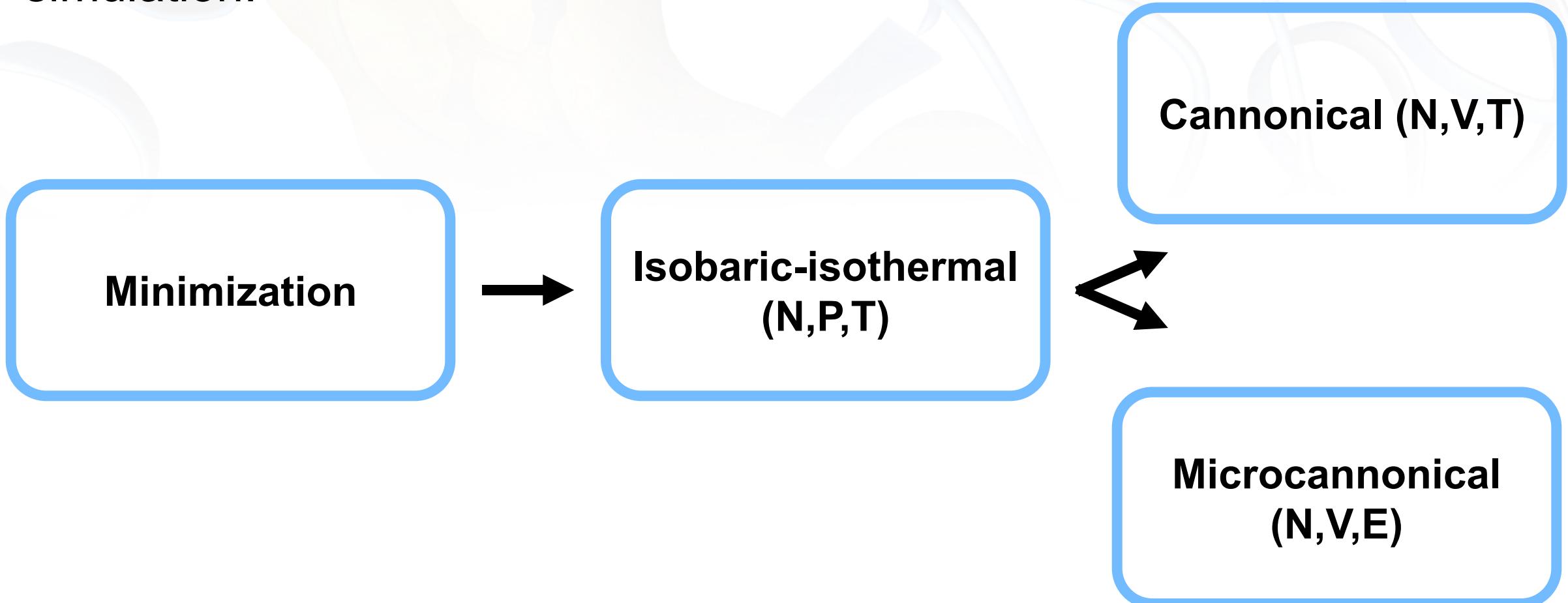
Reactions where bonds break and form or non-periodic

Grandcanonical (μ ,V,T)

Systems where the number of atoms can change (const. pH)

What is an ensemble?

It is okay, and common, to change **ensembles** for different parts of a simulation:

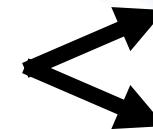


Basic molecular dynamics setup

Minimization



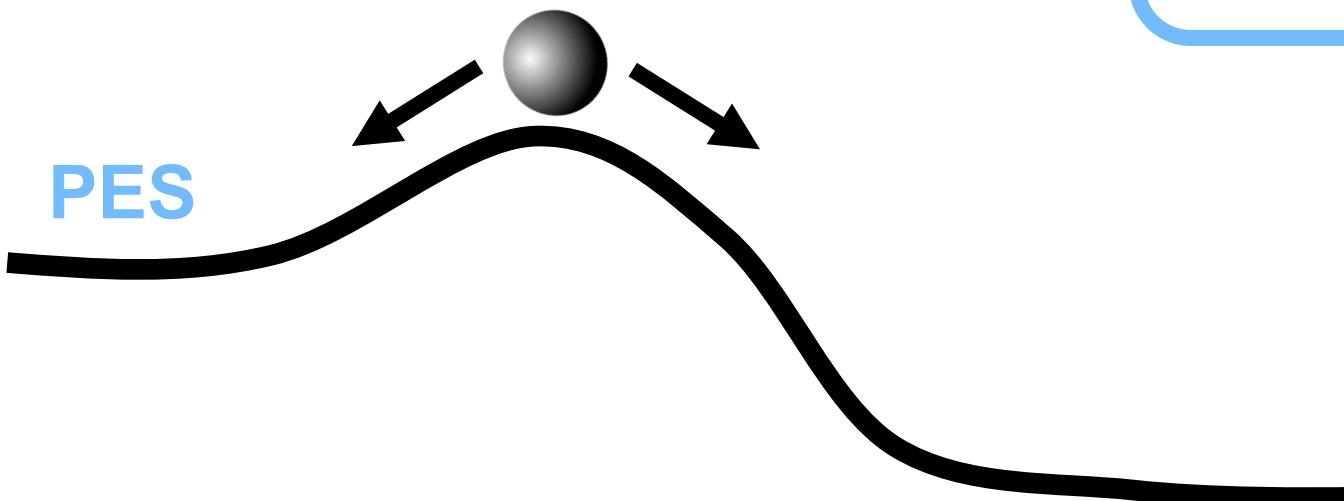
Isobaric-isothermal
(N,P,T)



Canonical (N,V,T)

Microcanonical
(N,V,E)

Performing a **minimization** removes bad contacts and relaxes highly strained geometries



PES

Basic molecular dynamics setup

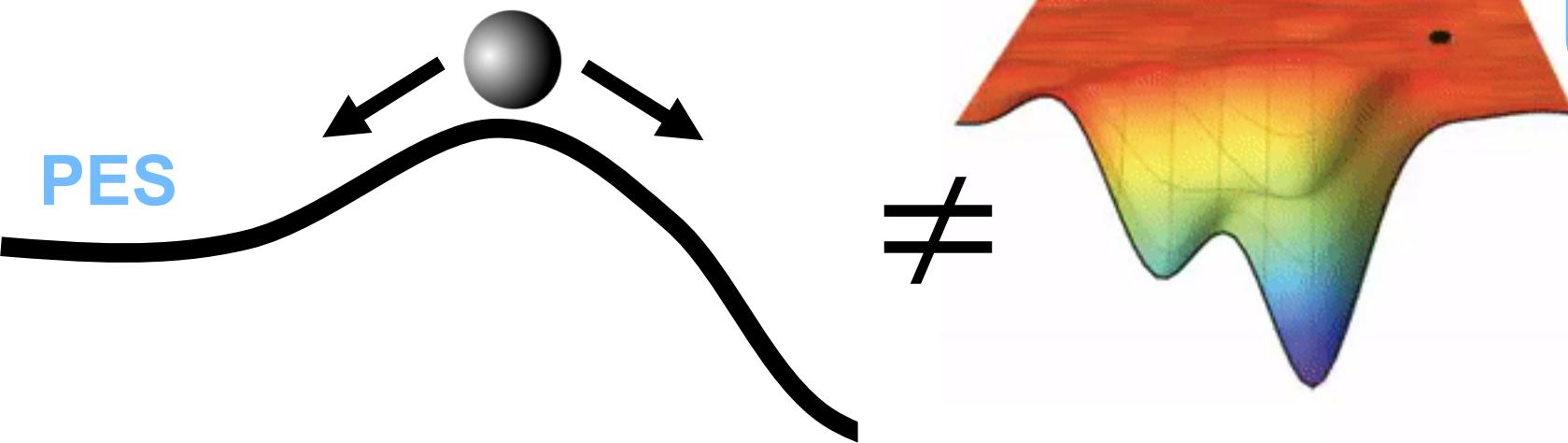
Minimization

Isobaric-isothermal
(N,P,T)

Canonical (N,V,T)

Microcanonical
(N,V,E)

Note minimization is not MD! (or vice versa)



Basic molecular dynamics setup

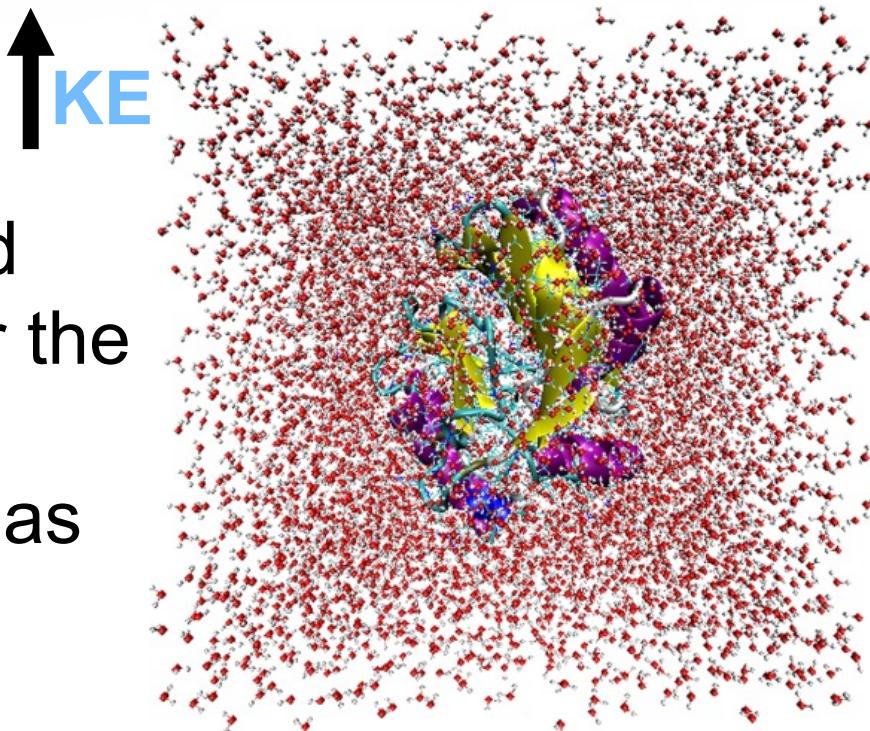
Minimization

Isobaric-isothermal
(N,P,T)

Canonical (N,V,T)

Microcanonical
(N,V,E)

Performing **heating** and **equilibration** allows for the ensemble to enter biologically relevant areas of phase space.



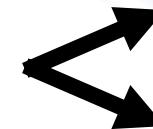
Basic molecular dynamics setup

Minimization



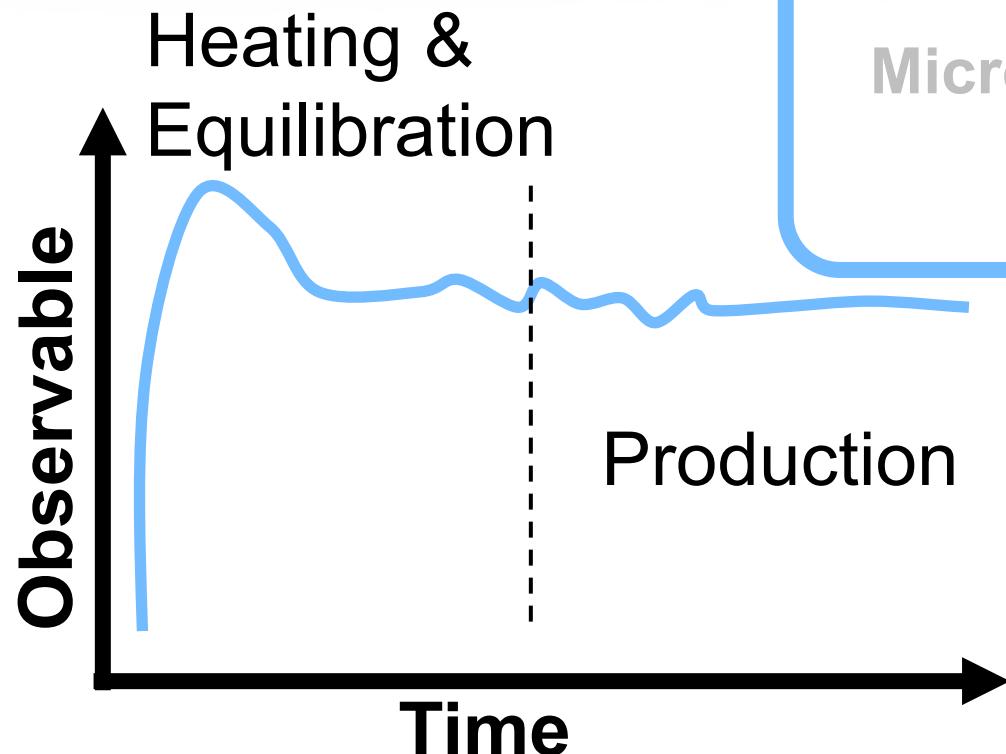
Isobaric-isothermal
(N,P,T)

Cannonical (N,V,T)

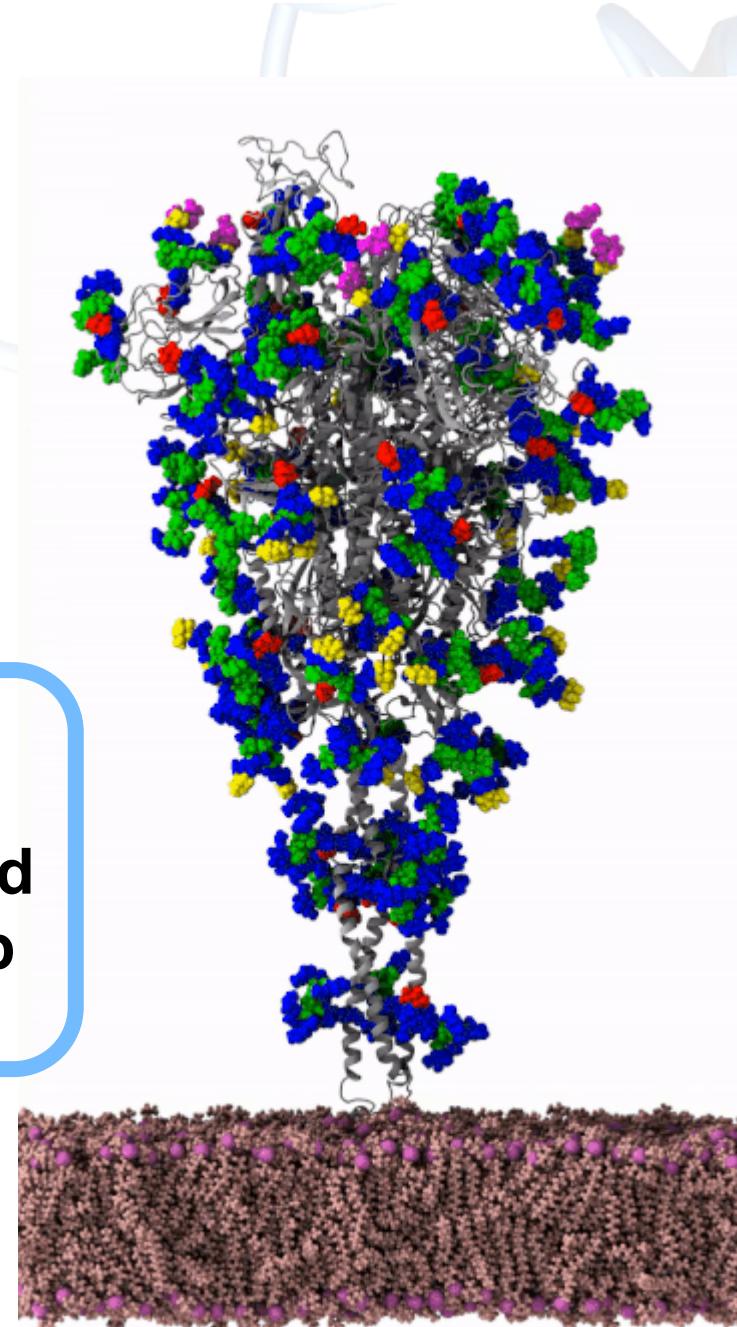
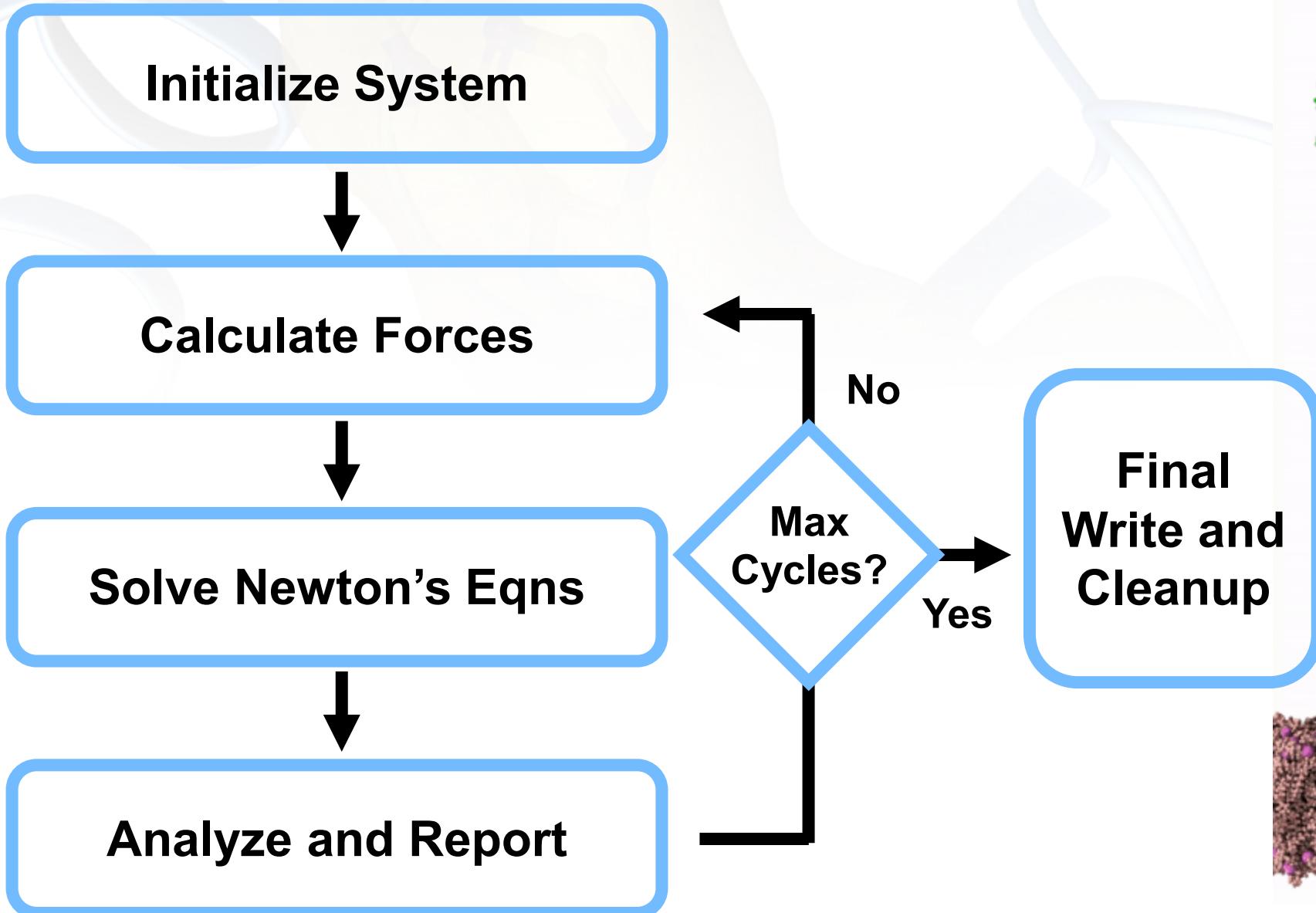


Microcanonical
(N,V,E)

Performing **heating** and **equilibration** allows for the ensemble to enter biologically relevant areas of phase space.



What is molecular dynamics?



Molecular dynamics integrators or: How I learned to start sampling and move through time

$$A_{obs} = \langle A \rangle_{\text{time}} = \langle A \rangle_{\text{ensemble}}$$

Molecular dynamics integrators or: How I learned to start sampling and move through time

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \dot{\mathbf{r}}(t)\delta t + \frac{1}{2}\ddot{\mathbf{r}}(t)\delta t^2 + \dots$$

Molecular dynamics integrators or: How I learned to start sampling and move through time

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \dot{\mathbf{r}}(t)\delta t + \frac{1}{2}\ddot{\mathbf{r}}(t)\delta t^2 + \dots$$

$$\mathbf{r}(t - \delta t) = \mathbf{r}(t) - \dot{\mathbf{r}}(t)\delta t + \frac{1}{2}\ddot{\mathbf{r}}(t)\delta t^2 + \dots$$

Molecular dynamics integrators or: How I learned to start sampling and move through time

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \dot{\mathbf{r}}(t)\delta t + \frac{1}{2}\ddot{\mathbf{r}}(t)\delta t^2 + \dots$$

$$\mathbf{r}(t - \delta t) = \mathbf{r}(t) - \dot{\mathbf{r}}(t)\delta t + \frac{1}{2}\ddot{\mathbf{r}}(t)\delta t^2 + \dots$$

Molecular dynamics integrators or: How I learned to start sampling and move through time

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \dot{\mathbf{r}}(t)\delta t + \frac{1}{2}\ddot{\mathbf{r}}(t)\delta t^2 + \dots$$

$$\mathbf{r}(t - \delta t) = \mathbf{r}(t) - \dot{\mathbf{r}}(t)\delta t + \frac{1}{2}\ddot{\mathbf{r}}(t)\delta t^2 + \dots$$

$$\mathbf{r}(t + \delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + \dot{\mathbf{r}}(t)\delta t^2$$

Molecular dynamics integrators or: How I learned to start sampling and move through time

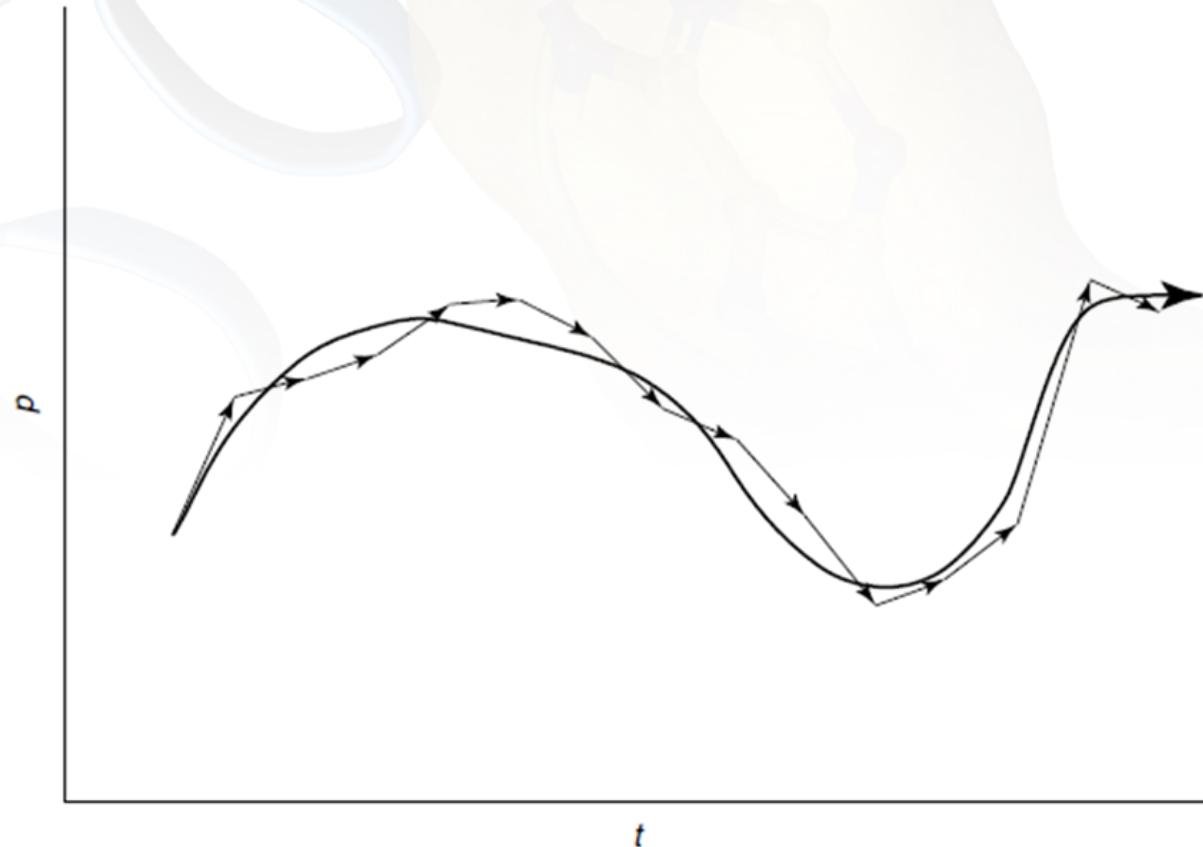
$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \dot{\mathbf{r}}(t)\delta t + \frac{1}{2}\ddot{\mathbf{r}}(t)\delta t^2 + \dots$$

$$\mathbf{r}(t - \delta t) = \mathbf{r}(t) - \dot{\mathbf{r}}(t)\delta t + \frac{1}{2}\ddot{\mathbf{r}}(t)\delta t^2 + \dots$$

$$\mathbf{r}(t + \delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + \dot{\mathbf{r}}(t)\delta t^2$$

This is the **Verlet algorithm**, useful because it is *cheap* and easy but has some inaccuracies!

Time step and molecular motions?

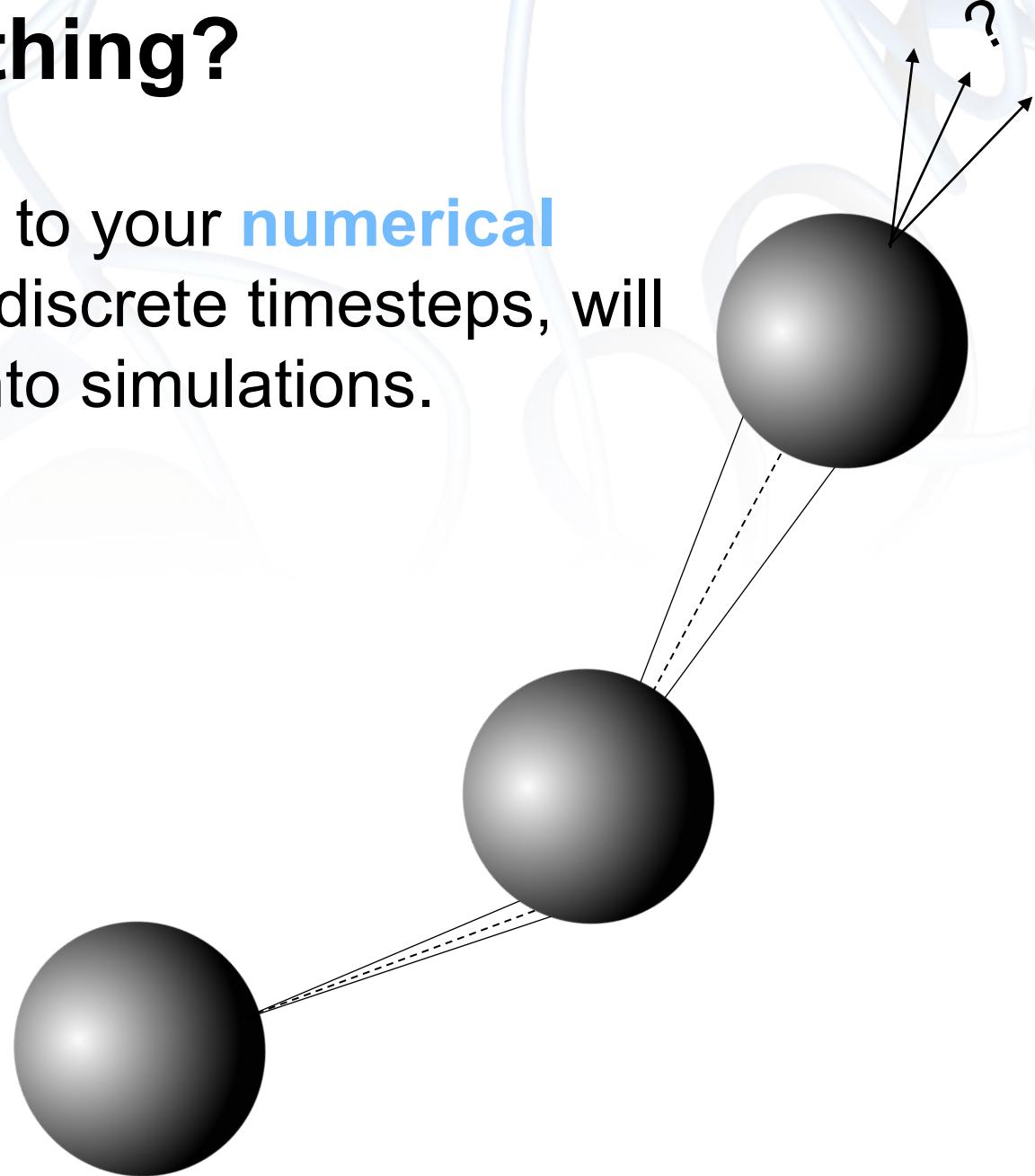


Values from
McCammon &
Harvey (1987) and
Eisenberg &
Kauzmann

Motion	length (Å)	time (fs)
bond vibration	0.1	10
water hindered rotation	0.5	1000
surface sidechain rotation	5	10^5
water diffusive motion	4	10^5
buried sidechain libration	0.5	10^5
hinge bending of chain	3	10^6
buried sidechain rotation	5	10^{13}
allosteric transition	3	10^{13}
local denaturation	7	10^{14}

Chaos and can we trust anything?

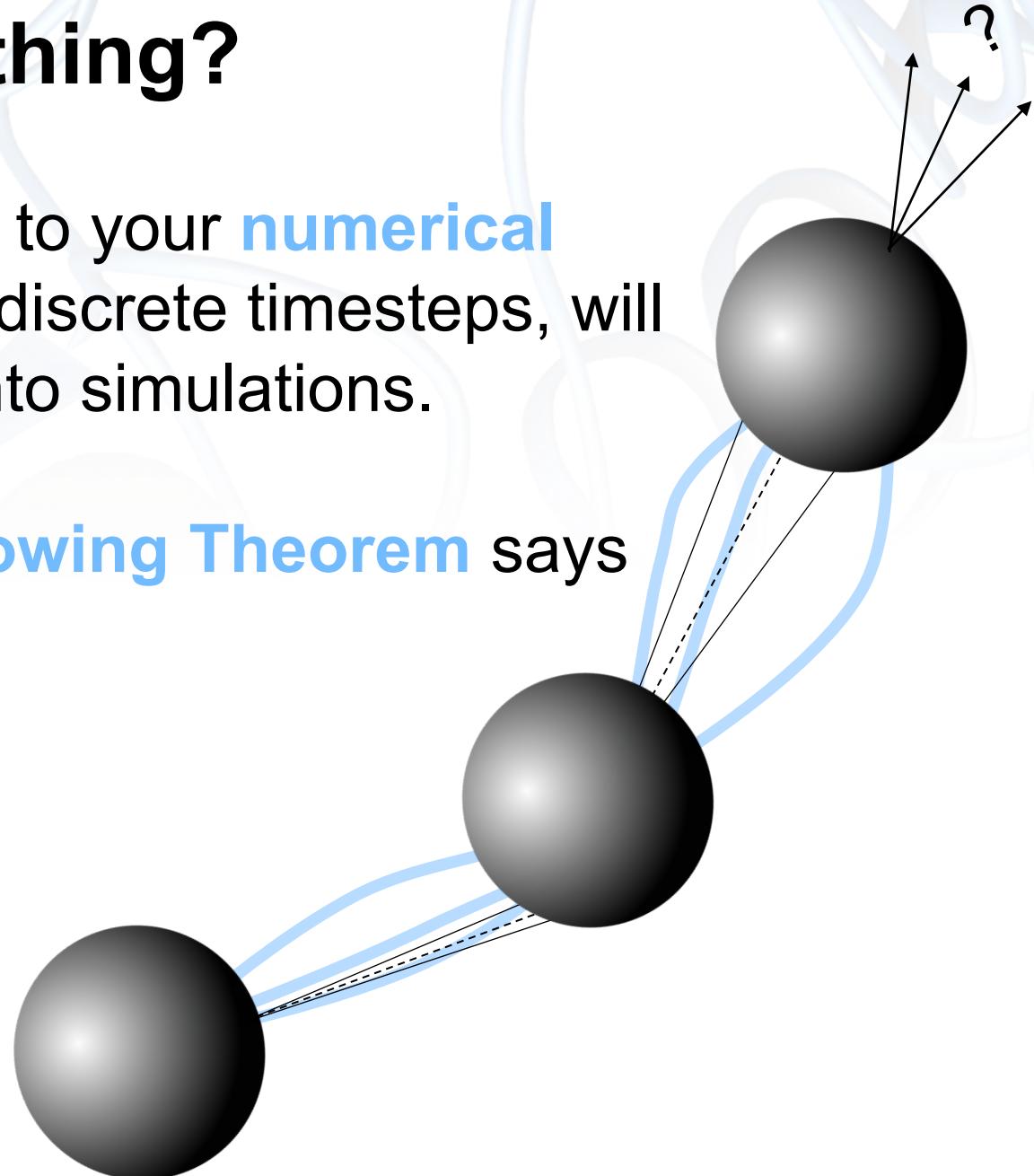
When using a computer, there is a limit to your **numerical precision** which, when combined with discrete timesteps, will introduce non-reversibility and chaos into simulations.



Chaos and can we trust anything?

When using a computer, there is a limit to your **numerical precision** which, when combined with discrete timesteps, will introduce non-reversibility and chaos into simulations.

However, does that matter? The **Shadowing Theorem** says probably not.

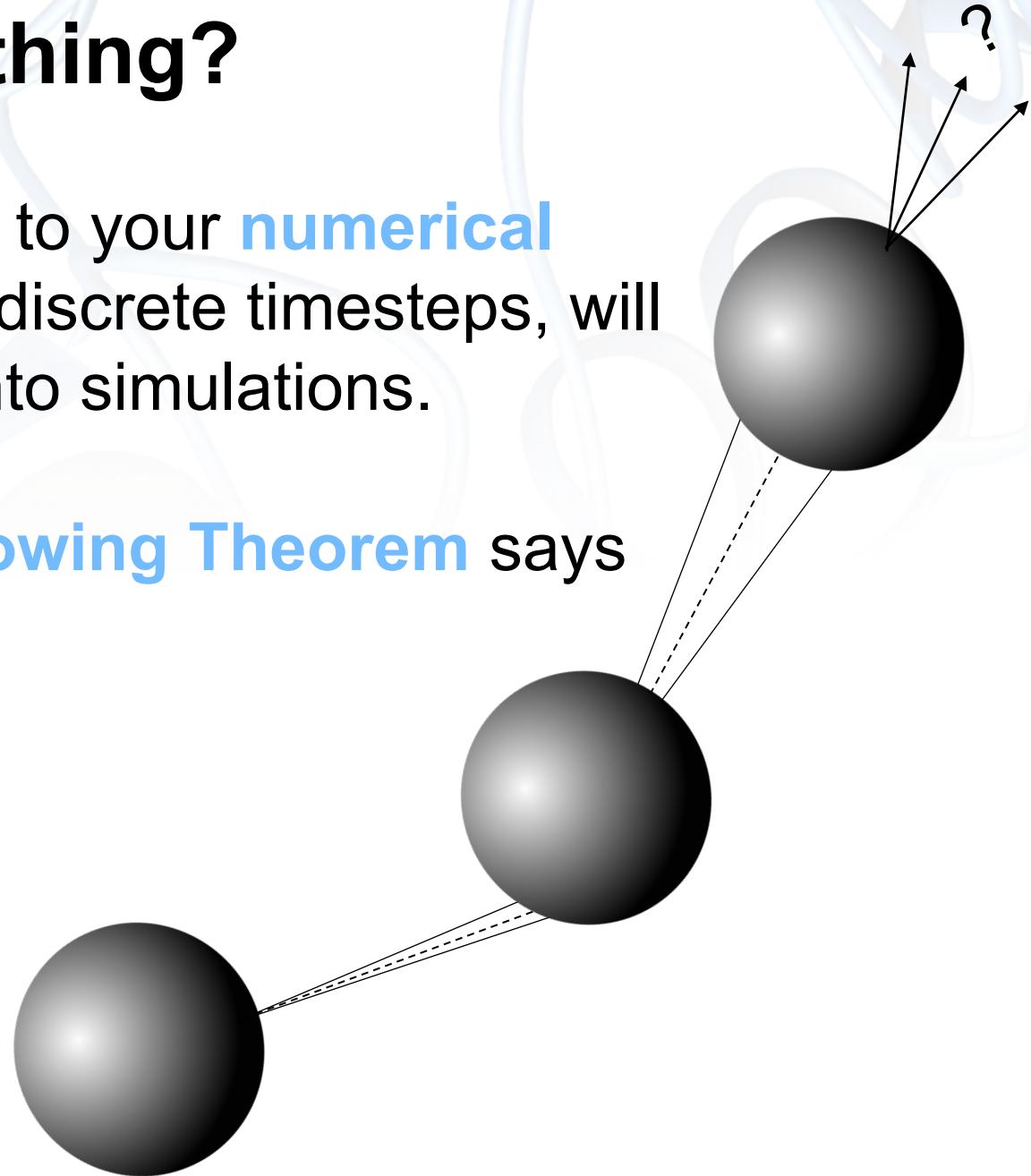


Chaos and can we trust anything?

When using a computer, there is a limit to your **numerical precision** which, when combined with discrete timesteps, will introduce non-reversibility and chaos into simulations.

However, does that matter? The **Shadowing Theorem** says probably not.

But, for this reason, it is better to think of the trajectories in molecular dynamics as **sampling** the ensemble instead of dynamics of the potential energy surface.



Types of all-atom simulation (yes there's more)

