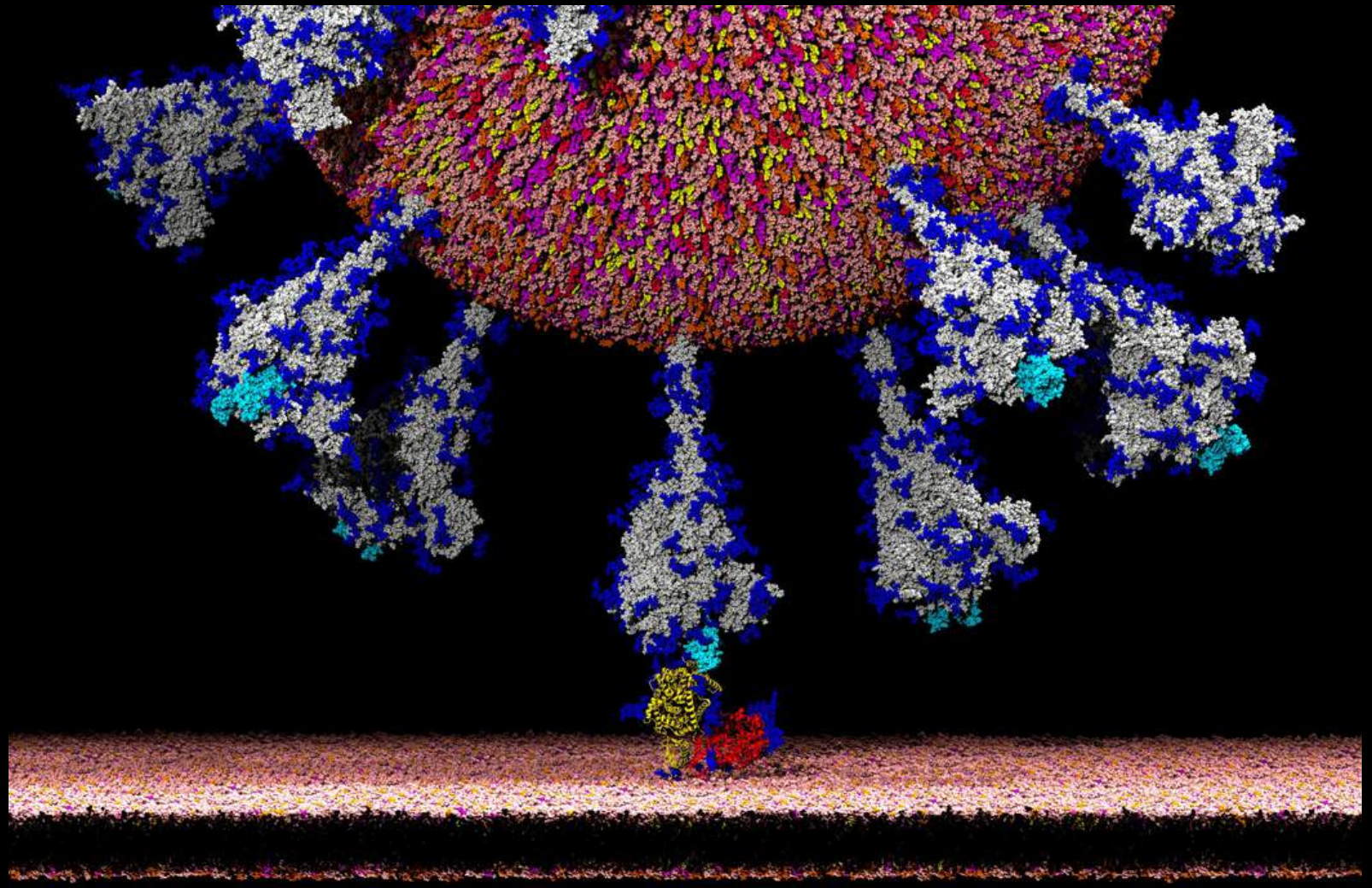
# Simulation techniques for sampling biomolecular conformations and free energies

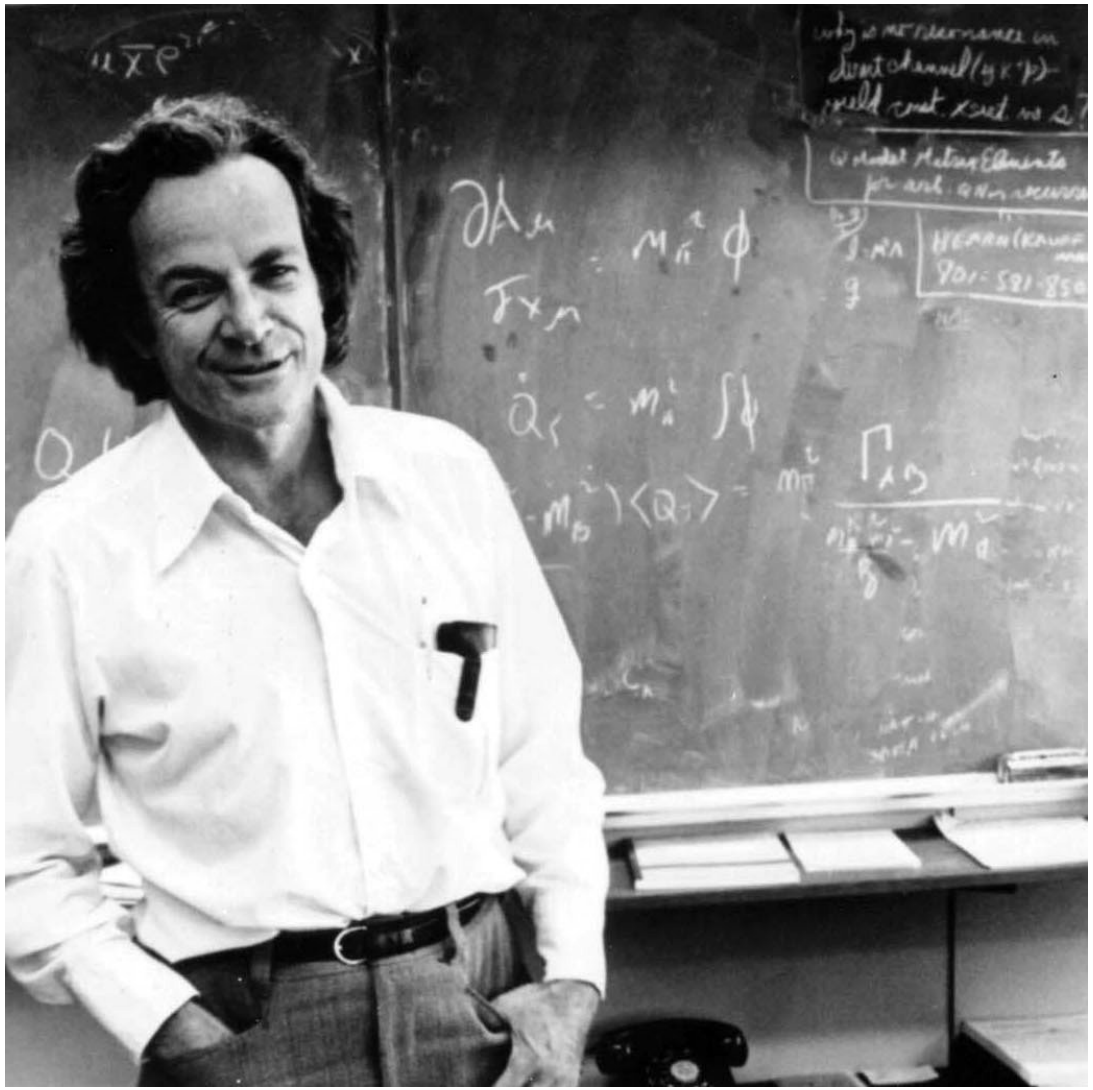
Glen M. Hocky, Experimental Biochem

December 4, 2020



Lorenzo Casalino, Amaro Lab (*Colloquium Feb 5*). https://www.nytimes.com/interactive/2020/health/coronavirus-unveiled.html

# INTRODUCTION

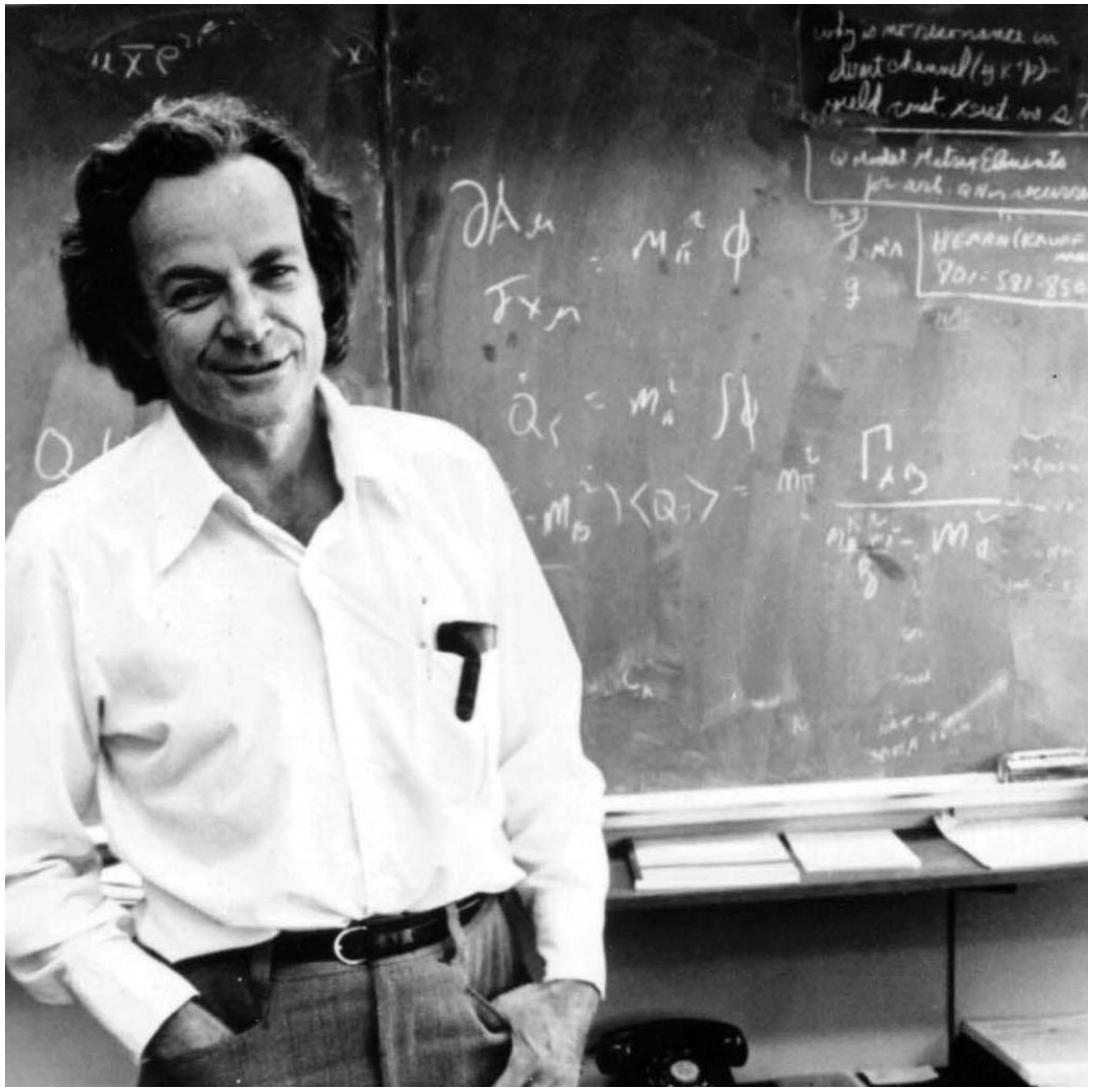


"...if we were to name the most powerful assumption of all, which leads one on and on in an attempt to understand life, it is that all things are made of atoms, and that everything that living things do can be understood in terms of the jigglings and wigglings of atoms."
– Feynman Lectures in Physics, 1964

**"…all things are made of atoms" = Chemistry**

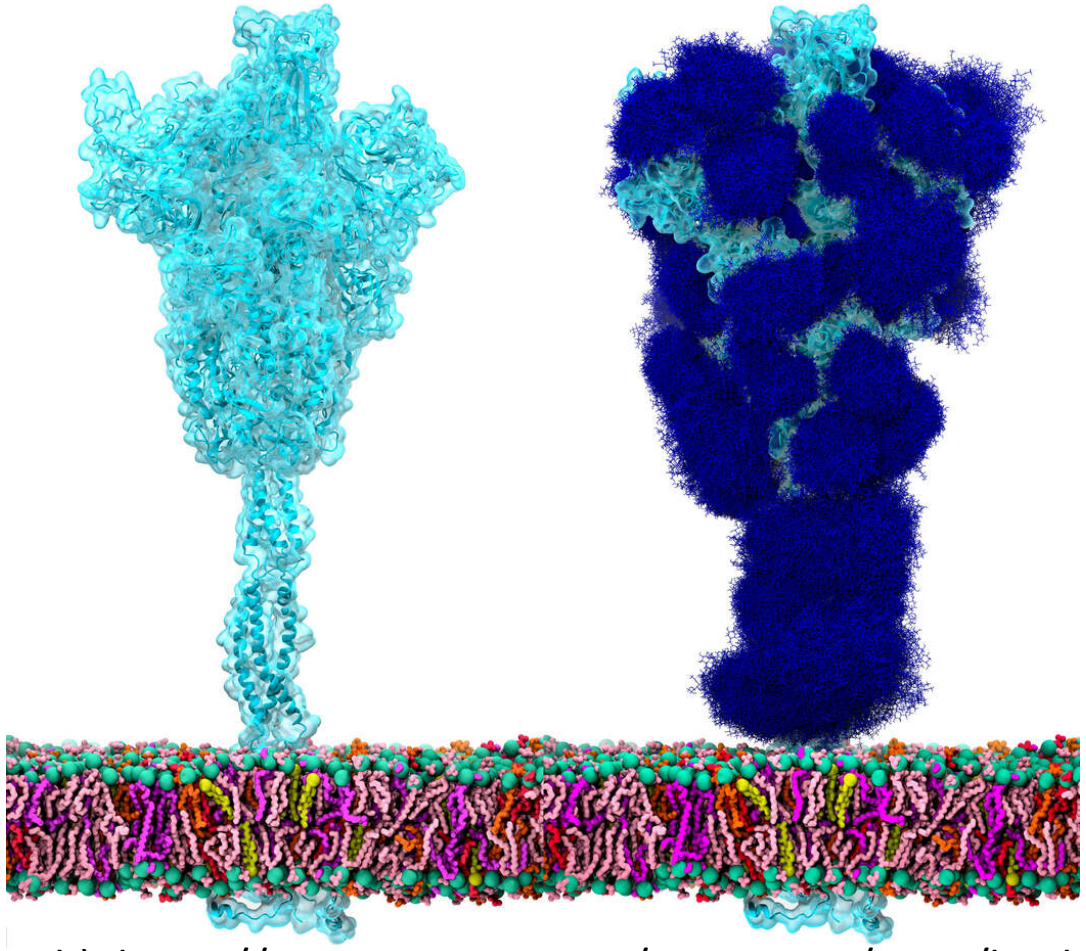
**"…jigglings and wigglings" = Physics**

# Outline



1. What computational research is all about
2. Computational techniques for studying biomolecular problems
3. Example problem from my past work
4. One research problem from our department (Alida Besch)

# USING COMPUTERS TO SOLVE PROBLEMS IN BIOCHEMISTRY



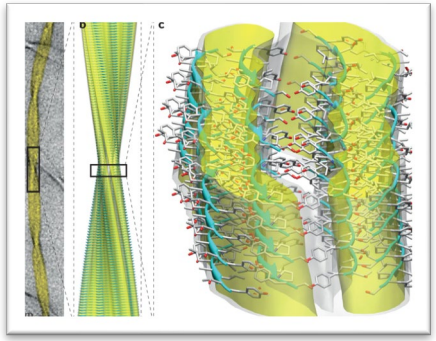
Lorenzo Casalino, Zied Gaieb (Amaro Lab). https://www.nytimes.com/interactive/2020/health/coronavirus-unveiled.html

Goal: Use the laws of physics to understand and predict molecular properties and interactions between molecules

# DIFFERENT RULES AT DIFFERENT LENGTH SCALES

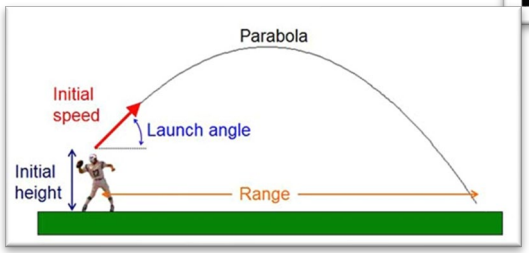
Theory

Relativity
$\sim 10^{21}$m

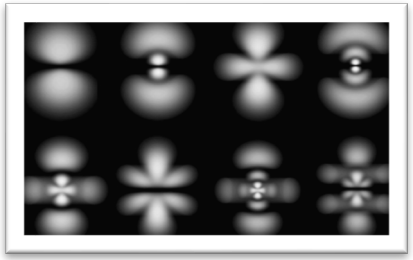


Classical Mechanics
$10^{-9}$ m

Classical Mechanics
$10^{-6}$ m




Classical Mechanics
1 m



Quantum Mechanics
$10^{-12}$ m

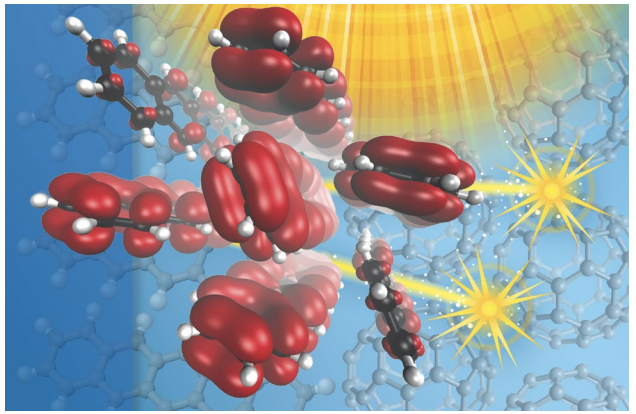
Particle Physics
$10^{-18}$ m

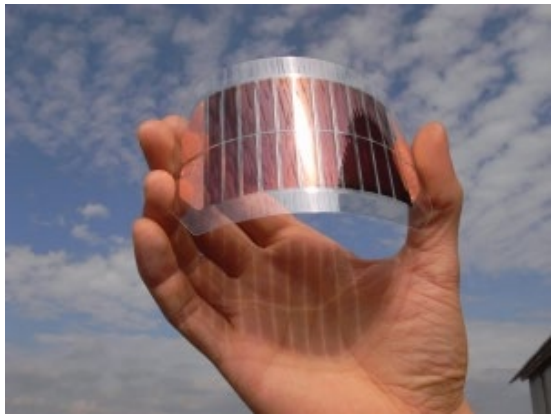
Size/Mass

# TWO KINDS OF THEORETICAL CHEMISTRY

## Quantum Mechanics

What is the behavior of the electrons?

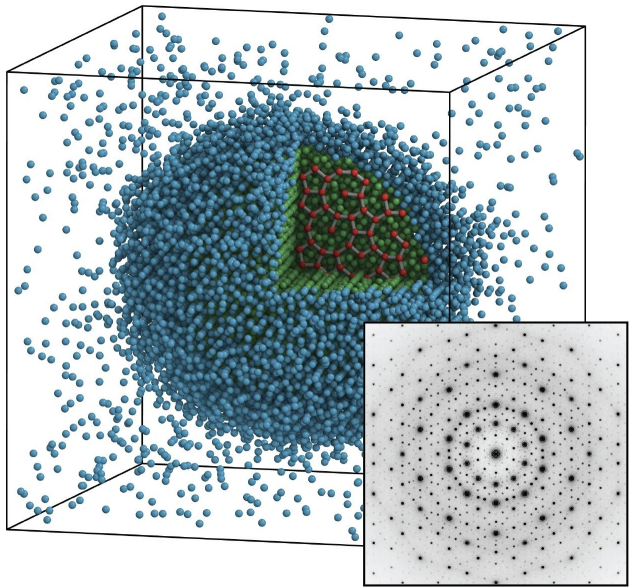


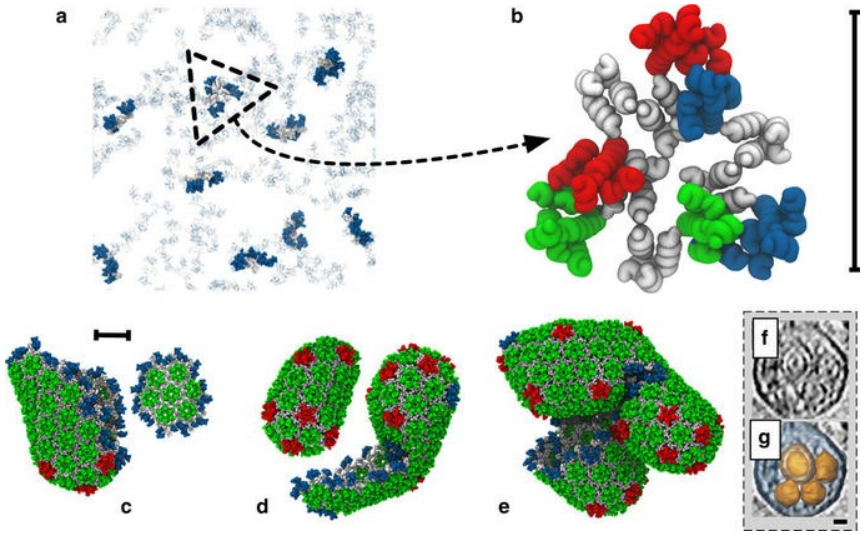


Important for the study of enzymes and understanding spectroscopy (not covering today)

## Statistical Mechanics

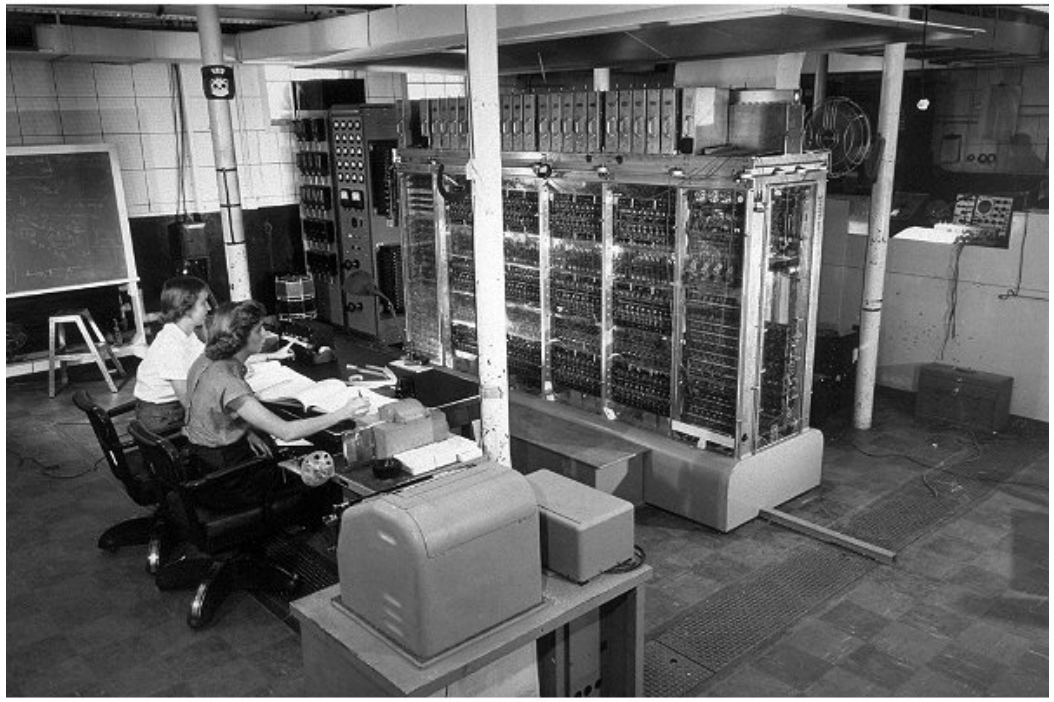
How do large collections of atoms/molecules behave?



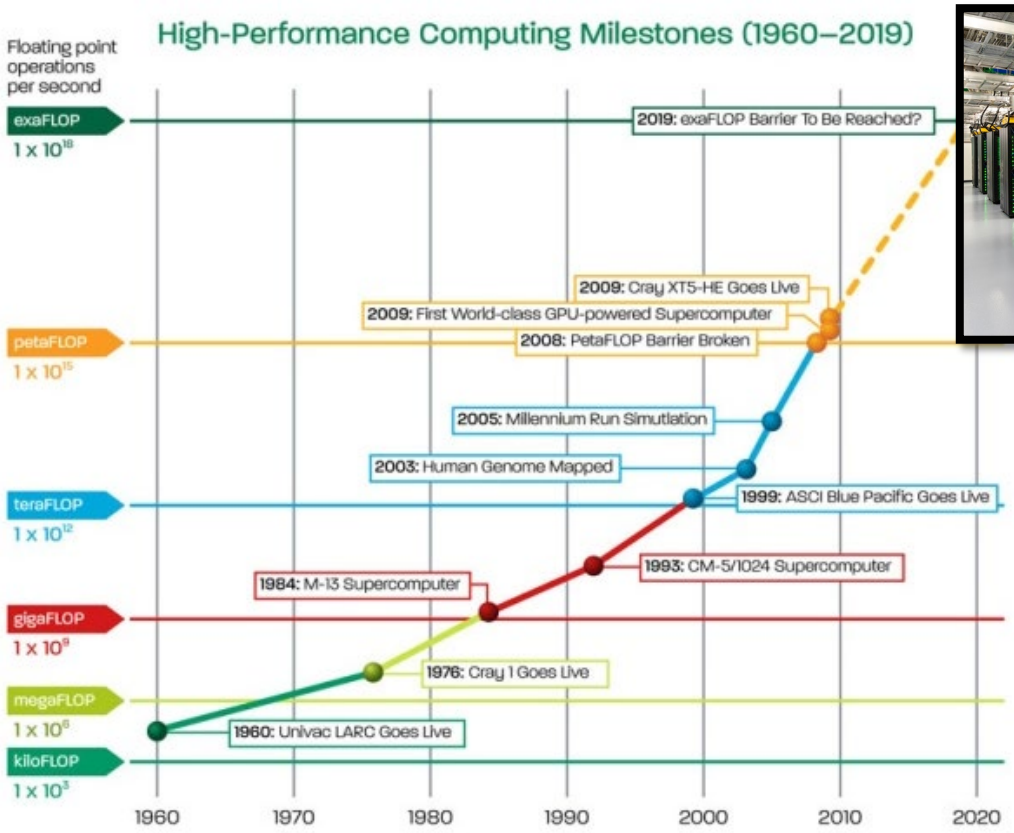


# Need for computers

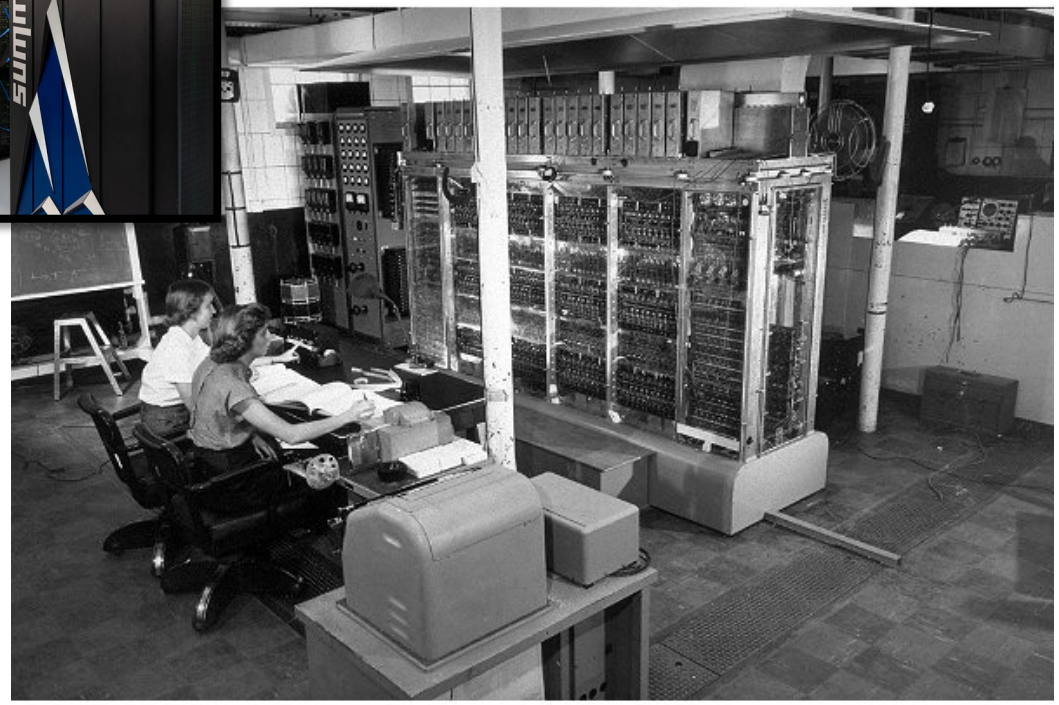
- Equations of quantum mechanics and of statistical mechanics are too complicated to solve by hand in most cases
- Used to make the most approximation that seemed reasonable, then sometimes use computers as calculators
- Computers first applied in chemistry during the Manhattan Project to predict nuclear properties



# Advances in computing power



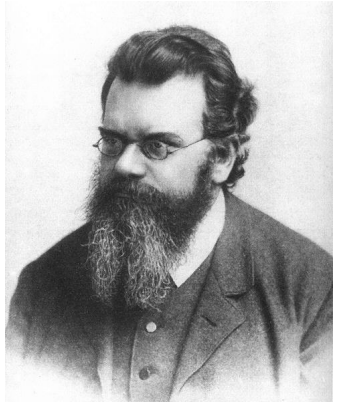


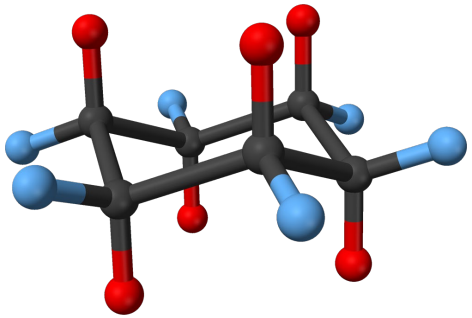


# Simulation Goal: Computing Weighted Averages

Real goal of simulations is to produce representative "samples" $X_i$

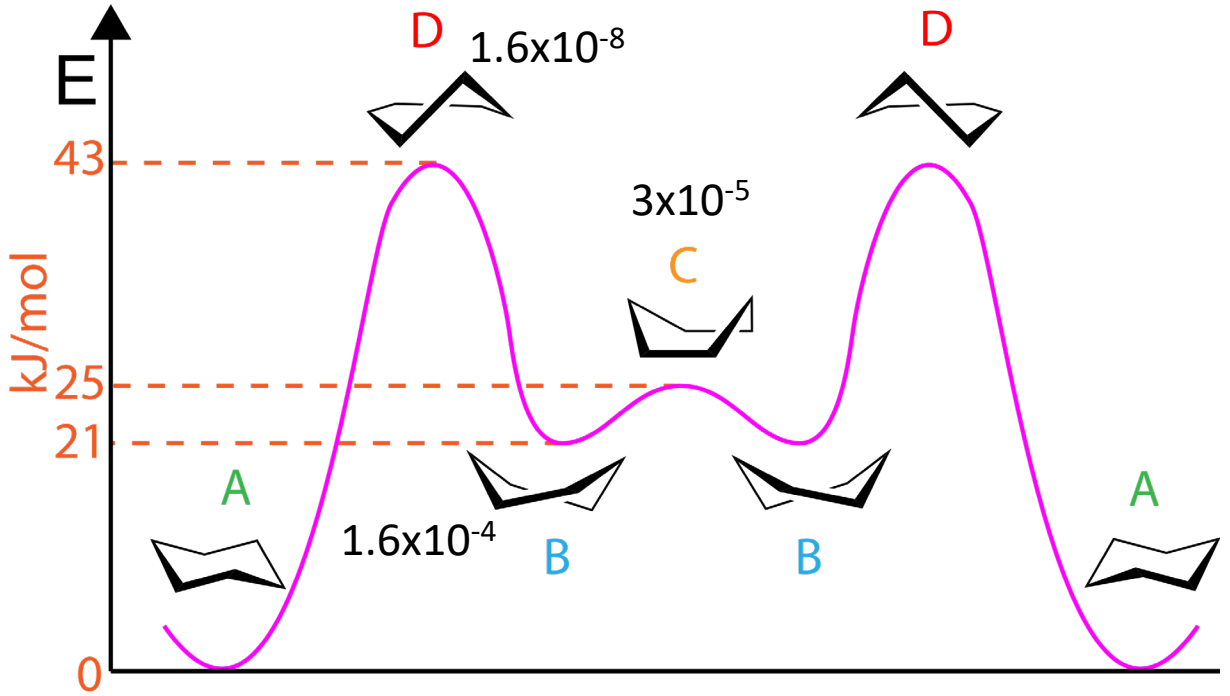
Then, compute averages of some quantity $A$ you are interested in: $\frac{1}{N}\sum_{i=1}^{N} A(X_i)$





**Probability of a molecular conformation**

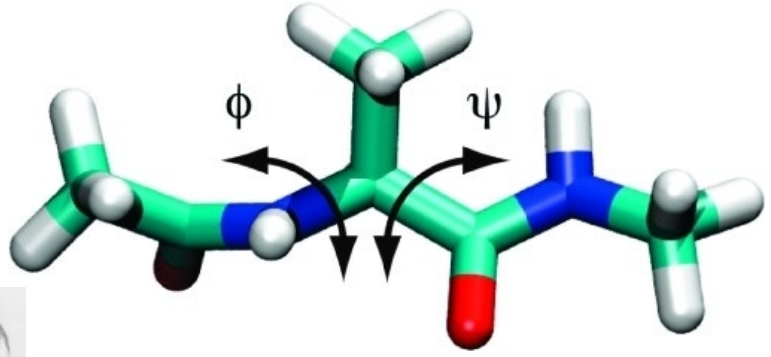
$$P\ (X) \propto e^{-\frac{E(X)}{k_B T}}$$

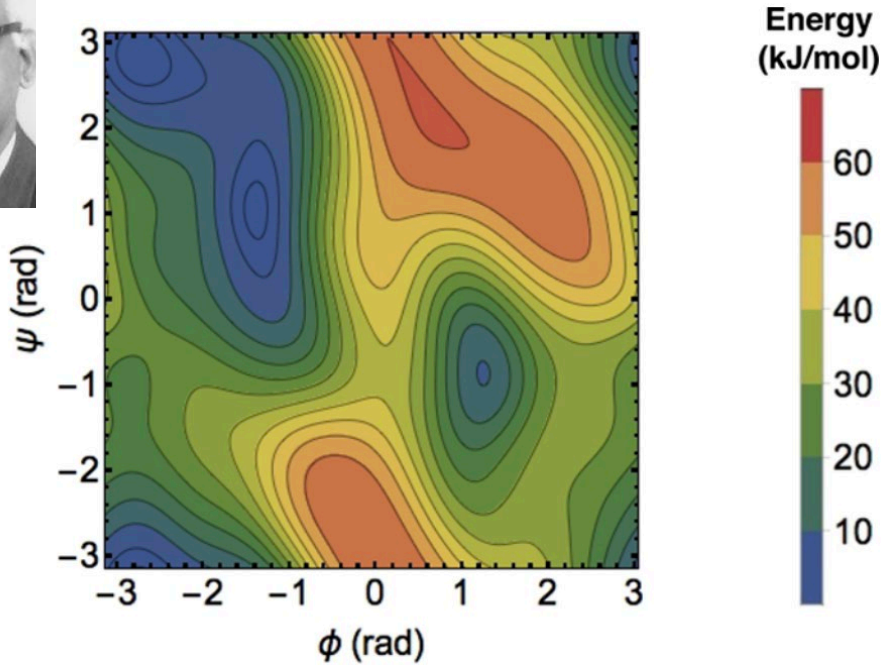



**Example "A":**
Average distance between C1 and C6?

# Atomistic Molecular Dynamics Basics

- Simple idea:
  F = m a → positions, velocities of atoms
  - Where do forces come from?
- Molecular mechanics "forcefield" built to reproduce experimental and quantum mechanical data
  - Atoms
    - Mass
    - Charge
    - Excluded volume
    - van der Waal's interactions
  - Bonds
    - Stretch
    - Bend
  - Torsion



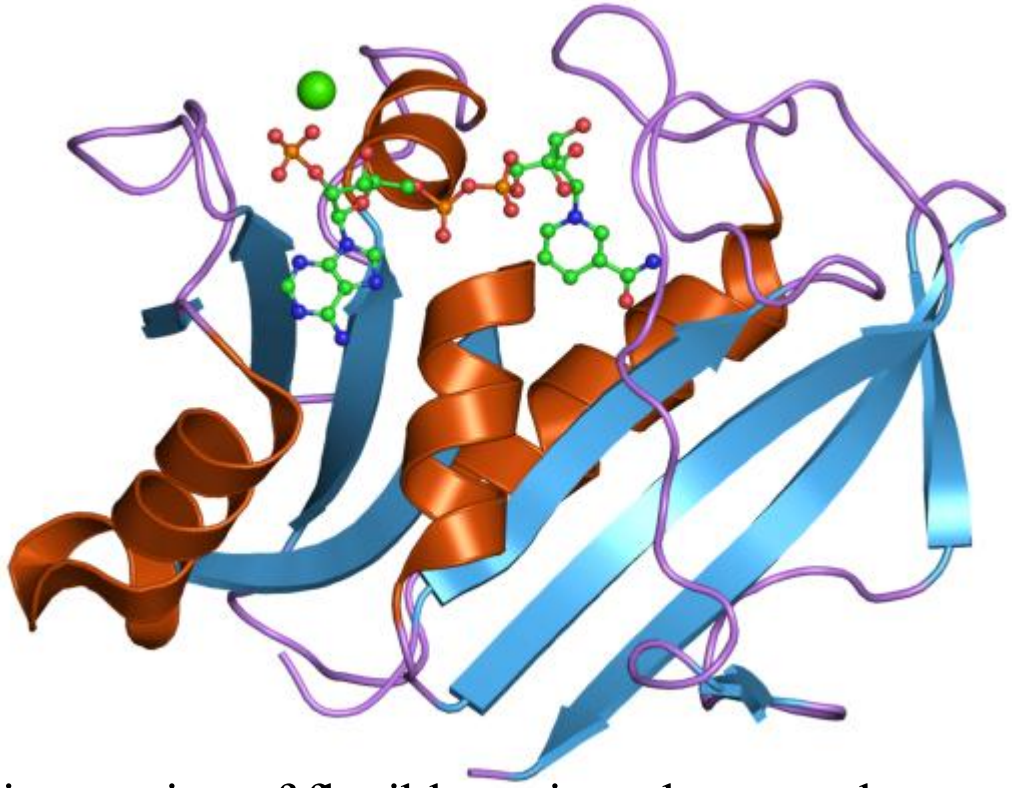




**Put molecule into a "temperature bath", and get** $P(X) \propto e^{-\frac{E(X)}{k_B T}}$

# EXAMPLE: DHFR

## What did we do for this lab?

1) Download structures from the PDB
2) Pick parameters for protein and water from standard set of options (Amber 14SB + TIP3P water)
3) Parameterize ligand (find reasonable atomic partial charges and spring constants using Generalized Amber Forcefield)
4) "Minimize" and "Equilibrate" initial structure
5) Run 100s of nanoseconds of MD simulations using a 2fs timestep and constant pressure of 1 atm and temperature of 300 K

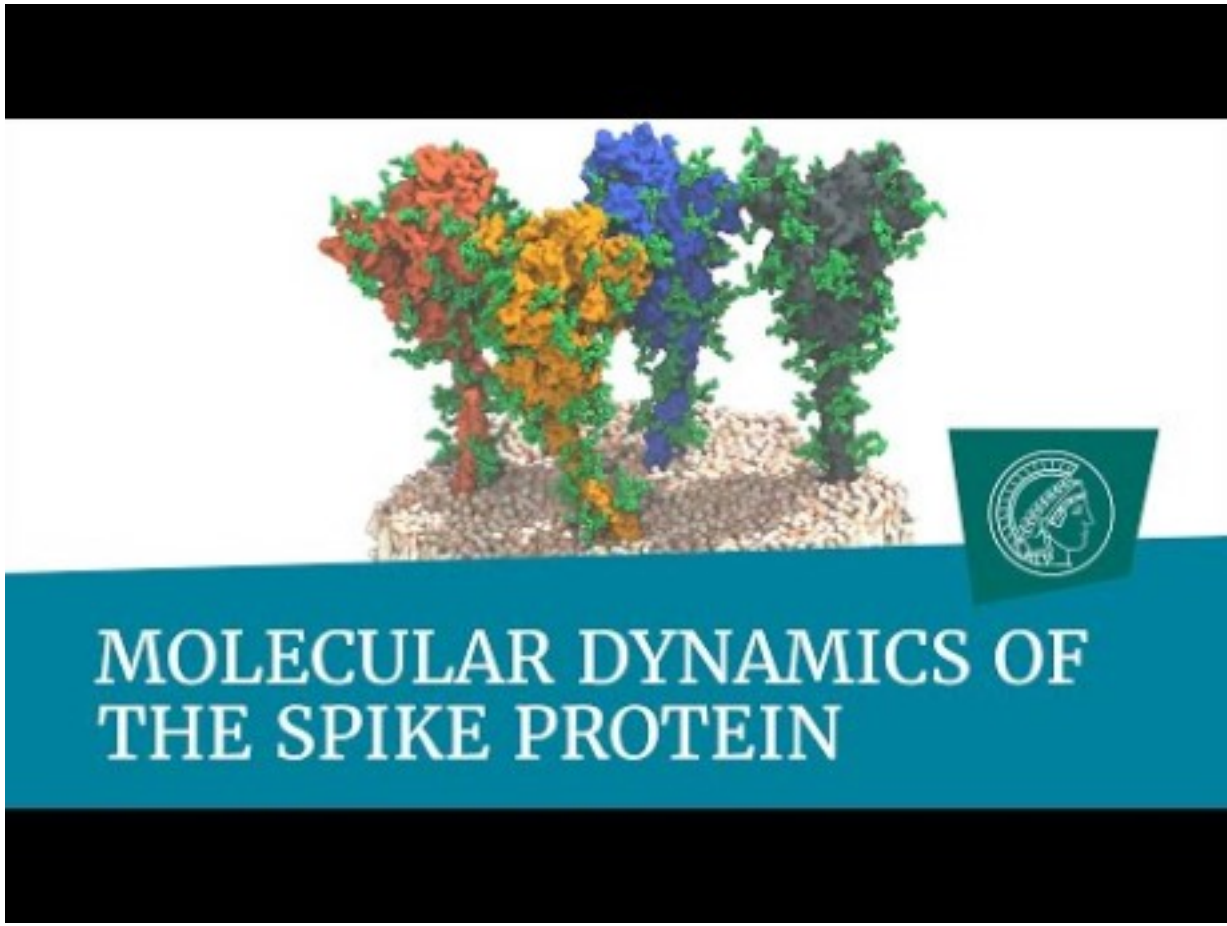


## What can we look at/do?

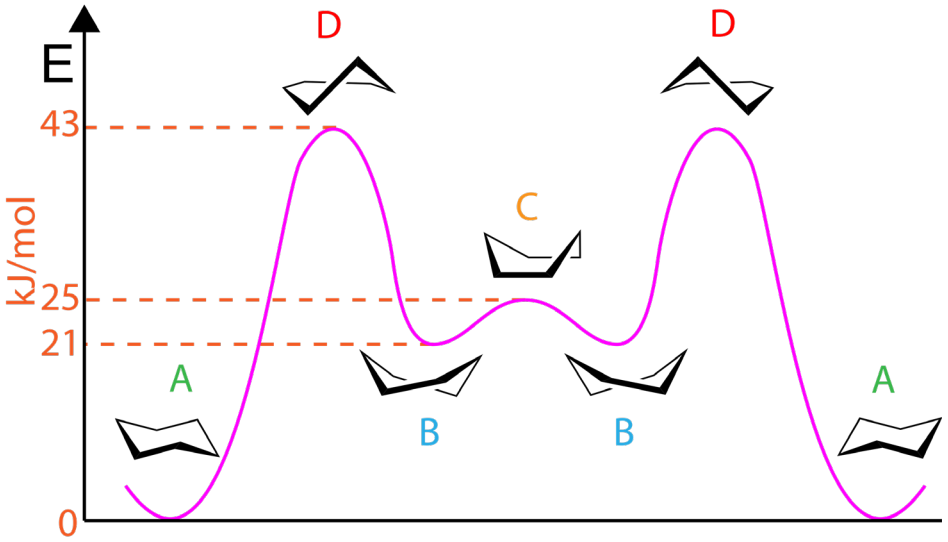
1) Can see properties that go beyond static structures, including motion of flexible regions, how much ligand moves in pockets, if different regions of the protein are correlated…
2) Can compare some structural or dynamical properties to experimental observables for validation (NMR observables, flexibility <-> B-factors, interaction of key residues to ligands)
3) Can make new predictions by e.g. introducing mutations, and using more advanced techniques to calculate free energies (later)

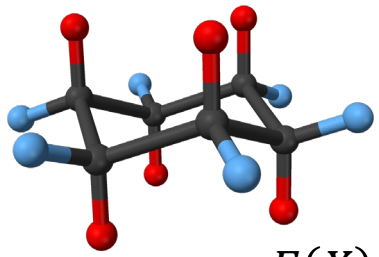
# What Limits Atomistic Molecular Dynamics?

- "Time step" shorter than fastest motion in the system (bond vibrations) – 1-2 fs
- System size (1->10 million atoms)
  - Solvating a system with water and ions makes most useful systems to study 100k-10 million atoms (very big system (30nm)^3)
- Computing force field for this many interactions
  - Split across many processor cores on a parallel machine
  - 100-10k cores for these sizes, 50-500ns/day
- Good starting structures!
- (Forcefield accuracy)

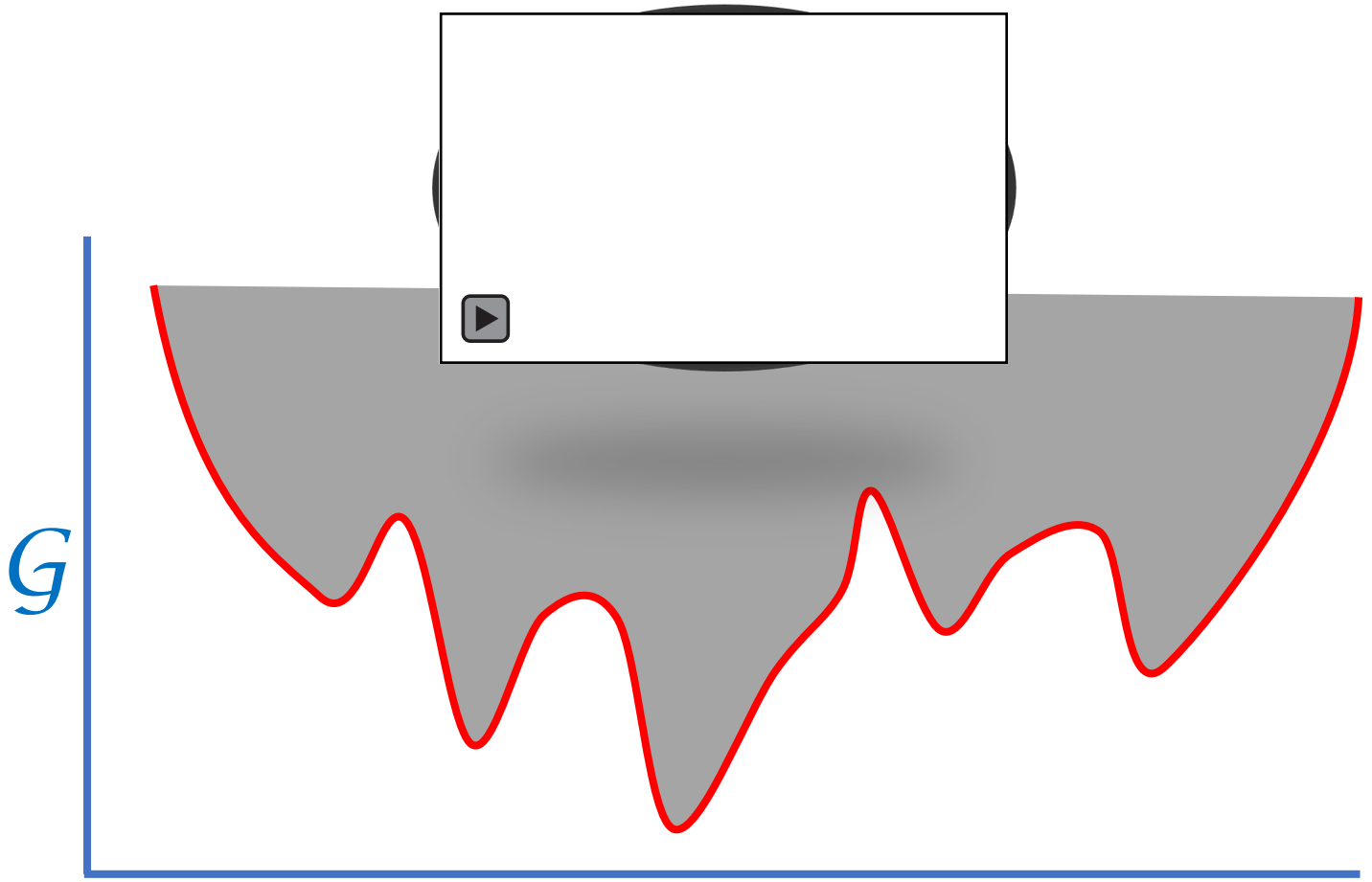


# Free Energy Landscape



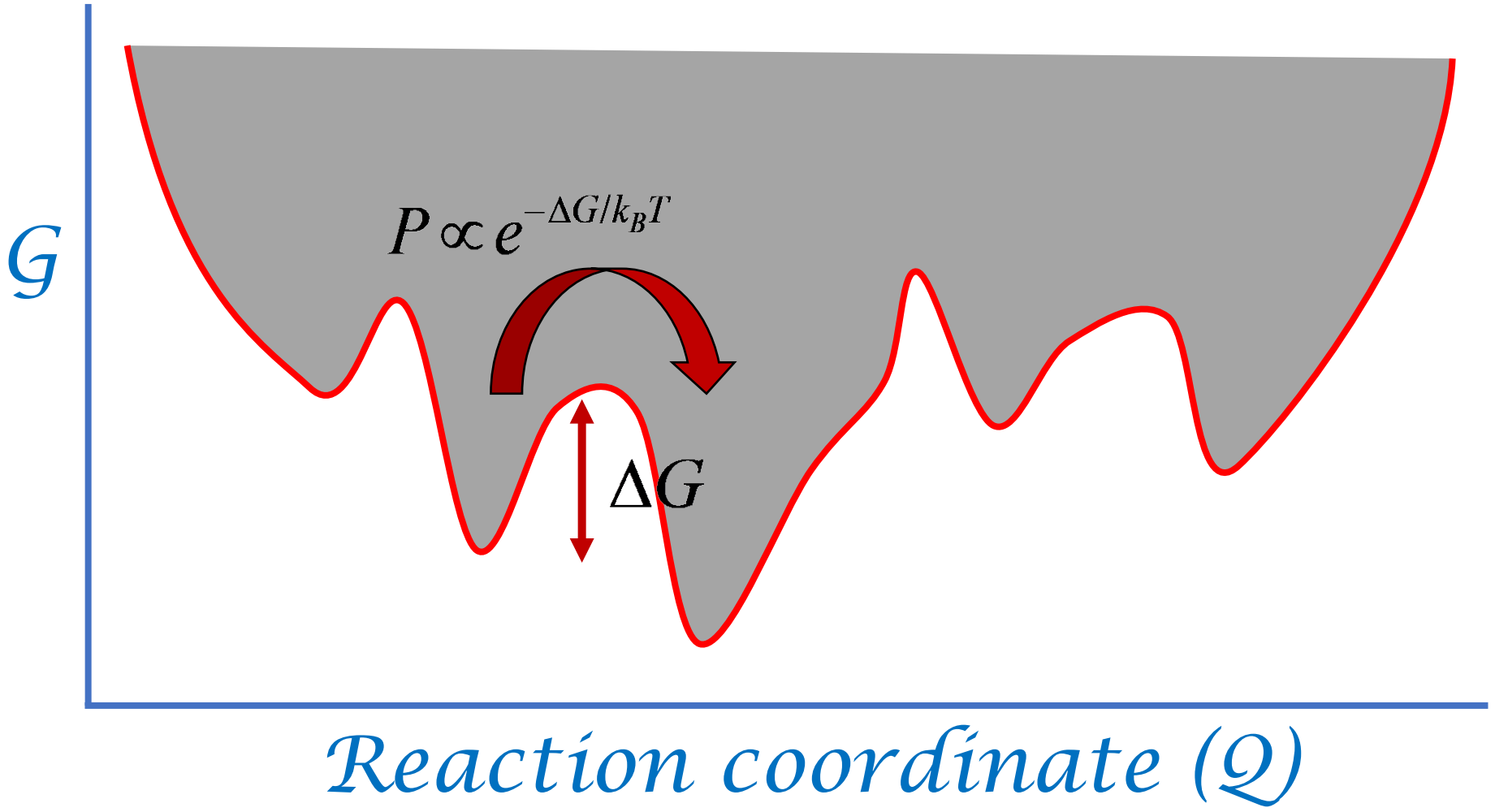



$$P\,(X) \propto e^{-\frac{E(X)}{k_B T}}$$



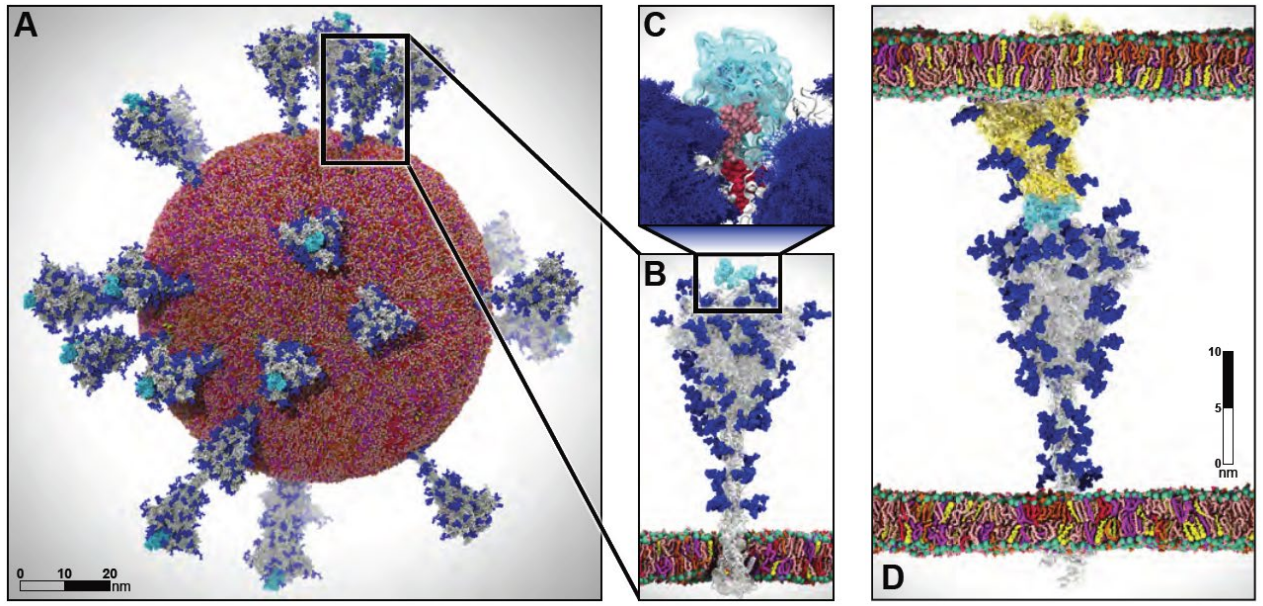

$$P\,(Q) \propto e^{-\frac{G(Q)}{k_B T}}$$

# ACCESSING LARGER CHANGES, LONGER "TIMES"

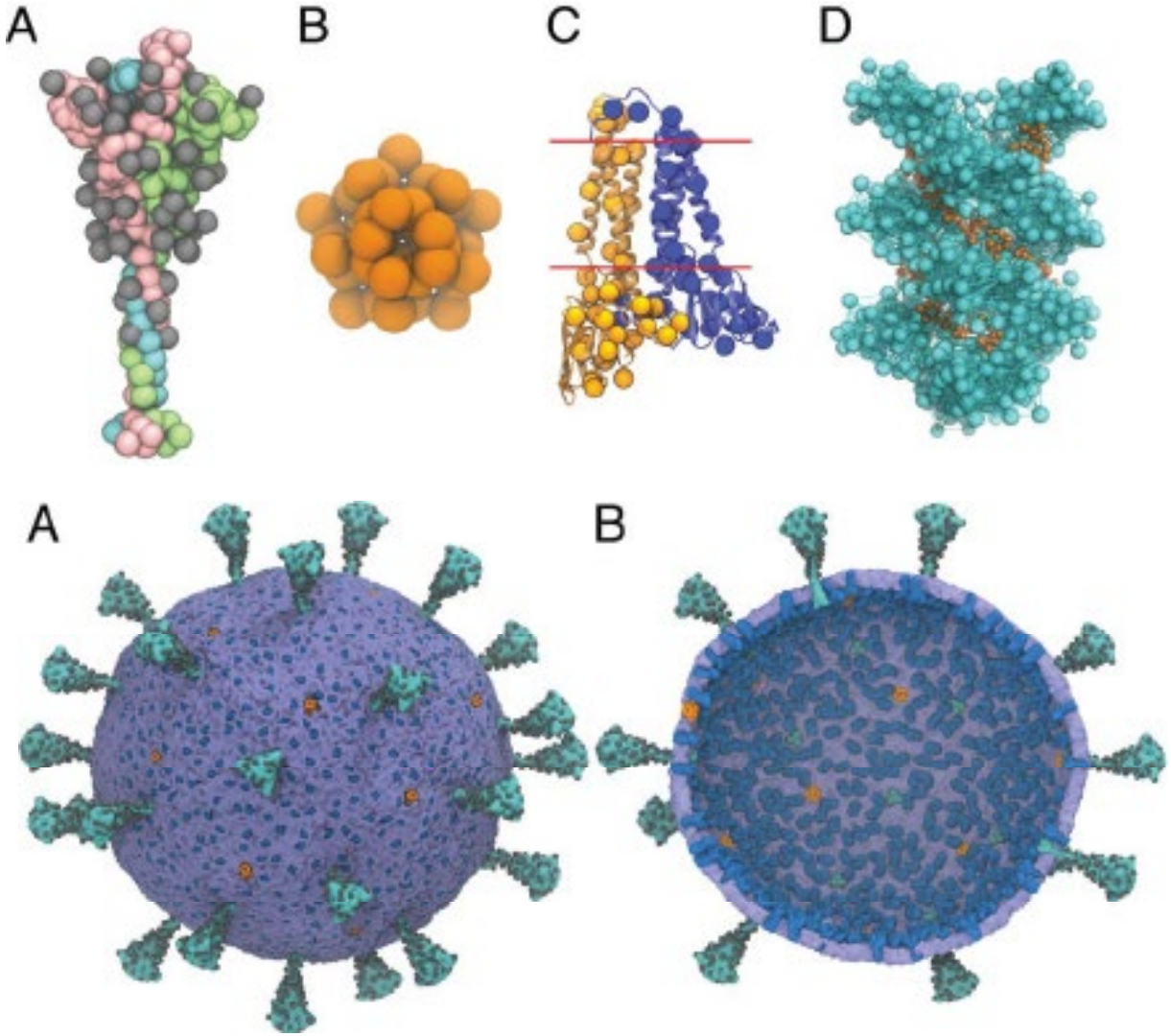



Techniques to compute free energy differences/landscapes necessarily "flatten" these barriers to sampling! This means absolute time accessible to simulation less important.

# GOING TO BIGGER SYSTEMS BY COARSE-GRAINING




305 Million atoms on Summit supercomputer – up to 50 ns/day
Casalino…Amaro, Supercomputing ‘20

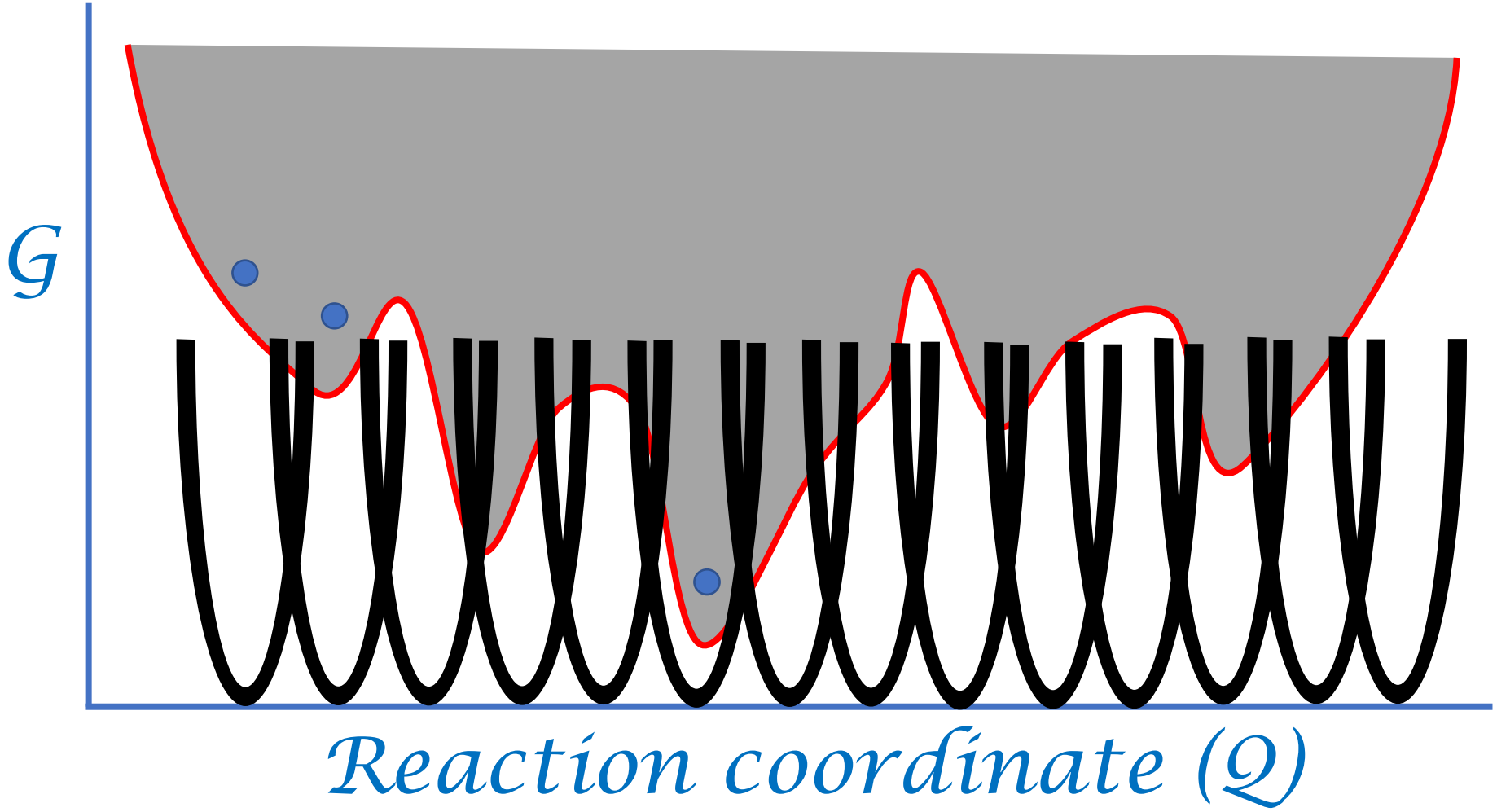



Yu, …, Amaro, Voth. bioRxiv 2020.10.02.323915

# ACCESSING LARGER CHANGES, LONGER "TIMES"

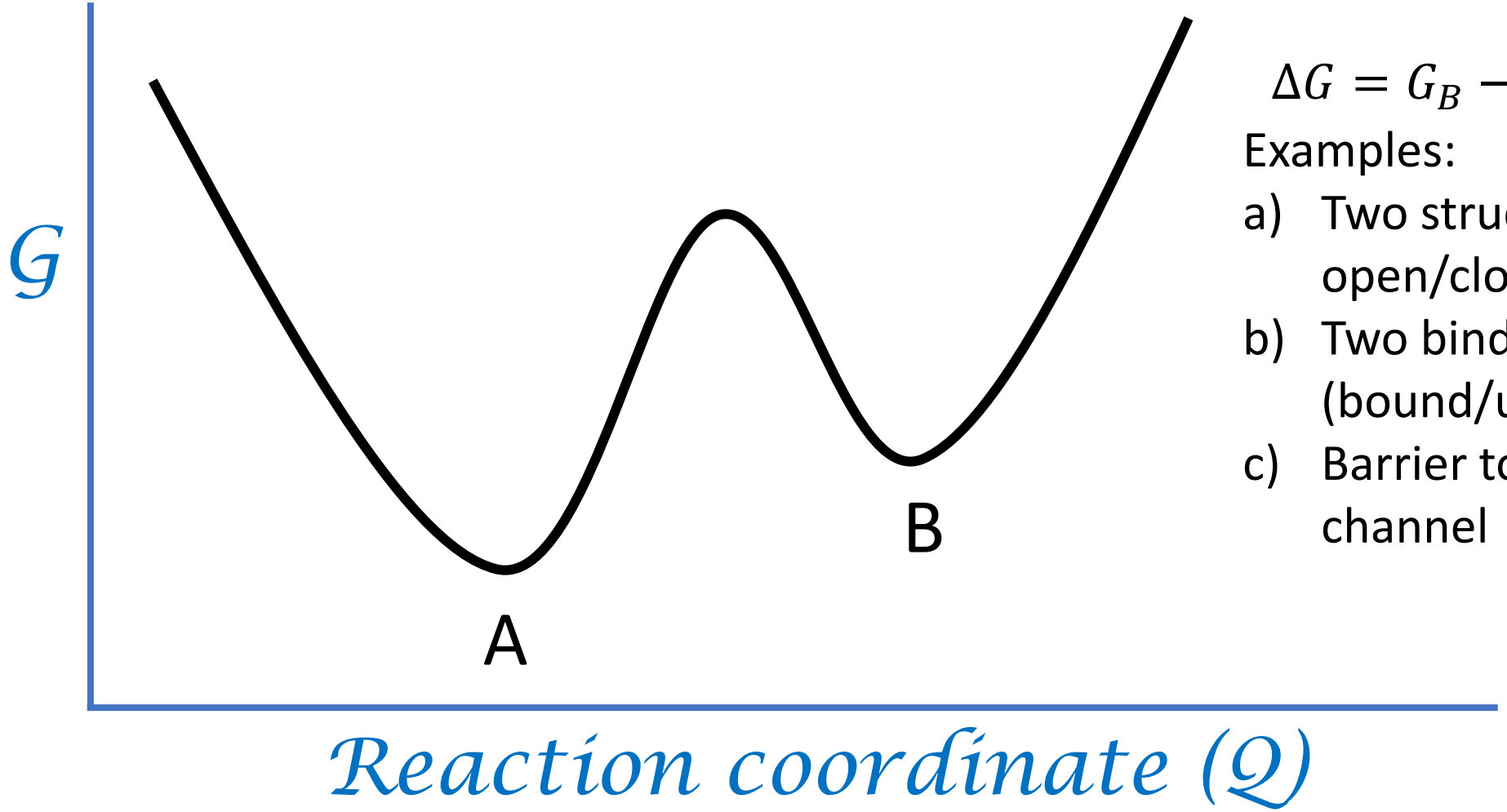
Alternative: force system to explore free energy landscape in a controlled manner

Example: "Umbrella Sampling": $U(X) = U_0(X) + \frac{1}{2}k(Q - Q_i)^2$



# Applying Accelerated Sampling Methods

Accelerated sampling methods can be used to compute free energy differences between two or more states of a system

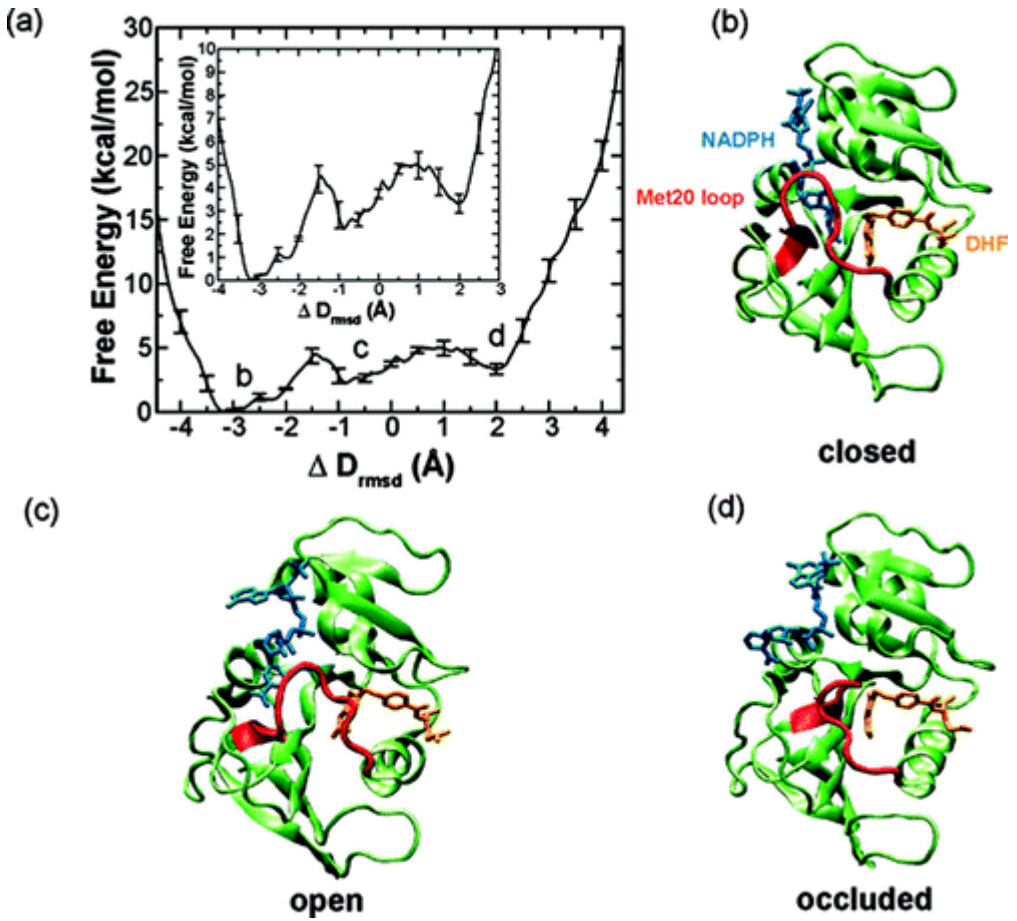


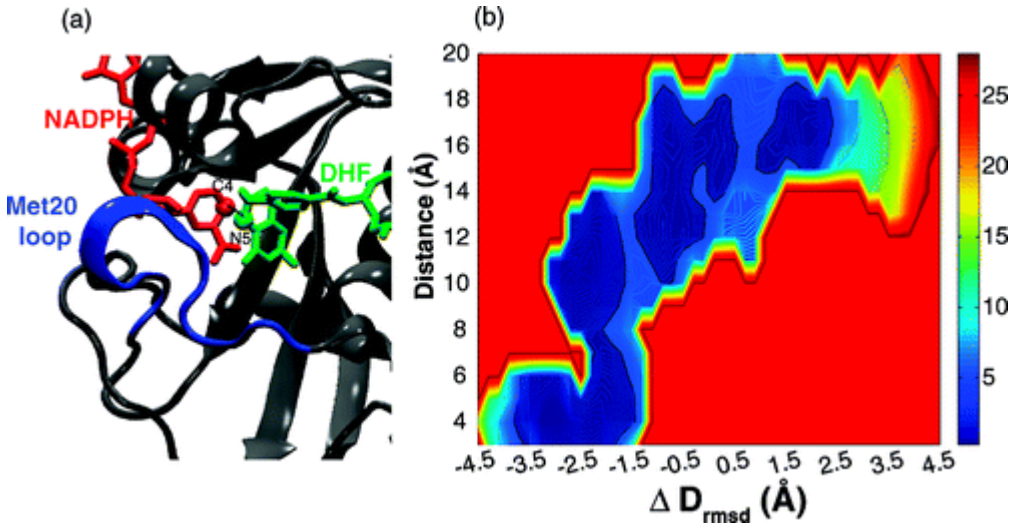

$$\Delta G = G_B - G_A = -RT \ln(\mathrm{K_{eq}})$$

Examples:

a) Two structural states (e.g. open/closed)
b) Two binding states (bound/unbound)
c) Barrier to transport through a channel

# EXAMPLE: DHFR




*J. Am. Chem. Soc.* 2009, 131, 15, 5642-5647

# UNDERSTANDING MOLECULAR COMMUNICATION

Accelerated sampling methods let us study the transition between two different states of a protein (e.g. bound and unbound)
Analyzing regular MD simulations in each state can show us how motions in those states are correlated

Can these correlations give us a hint of how the changes in one part of the protein (mutations, binding a ligand) will affect other parts of the protein?