

Energy functions and their relationship to molecular conformation

CS/BioE/Biophys/BMI/CME 279

Oct. 3 and 5, 2023

Ron Dror

Outline

- Overview
 - High-level questions
 - Demo
- Energy functions for biomolecular systems
 - Definition and properties
 - Molecular mechanics force fields
- What does the energy function tell us about biomolecular structure/conformation?
 - The Boltzmann distribution
 - Conformations and conformational states
 - Free energy

Overview

A biomolecule adopts many shapes

- The atoms in biomolecules are constantly jiggling around
 - “Everything that living things do can be understood in terms of the jiggings and wiggings of atoms.” — Richard Feynman, 1963 (Nobel Prize, 1965)
- A biomolecule adopts many geometries/shapes!
- We refer to each geometry of a molecule (i.e., precise arrangement of atoms, specified by 3D coordinates) as a **conformation**
 - “Conformation” is similar to “structure,” except that “structure” is often used to describe an average structure, which is what one typically gets when determining a structure experimentally

The big questions

- **Given a biomolecule (e.g., protein), which conformations will it adopt? How frequently will it adopt each conformation?**
 - Note that this depends on the other molecules surrounding it, so we typically consider a “molecular system” consisting of multiple molecules
 - It also depends on temperature
 - We can ask these questions either for individual conformations or for sets of similar conformations (referred to as “conformational states”)

Demo

- Take-aways:
 - The system adopts many conformations
 - It adopts low-energy conformations more frequently than high-energy conformations
 - If we can define the energy associated with each conformation, we can determine how often the system will adopt each conformation
- We'll thus discuss how to calculate energies for conformations of biomolecules (and biomolecular systems)

Key difference between demo and molecular systems

- To specify the “conformation” (horizontal position) of the cheerio or ball, I need only two numbers
- To specify the conformation of a molecular system (or a single biomolecule), I need to specify the x , y , and z coordinates of each atom.
 - For N atoms, that’s $3N$ coordinates.
 - Energy depends on all of these coordinates!
- All the take-aways still apply to molecular systems!

Energy functions for biomolecular systems

Energy functions for biomolecular systems

Definition and properties

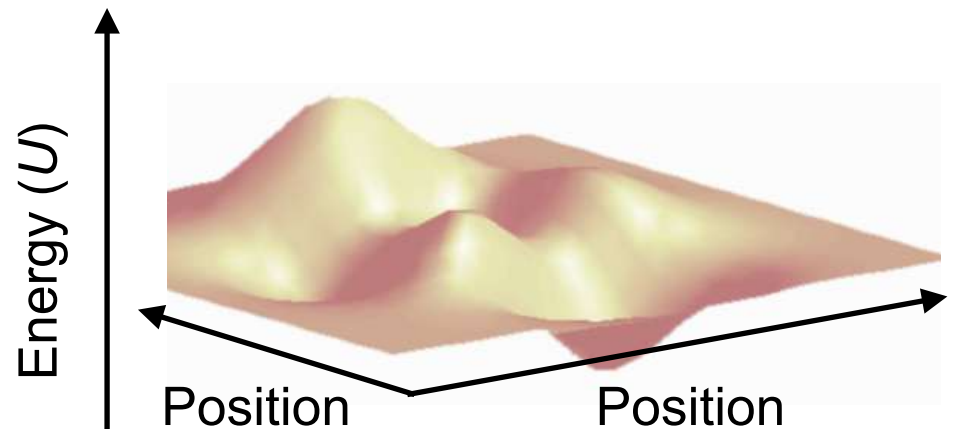
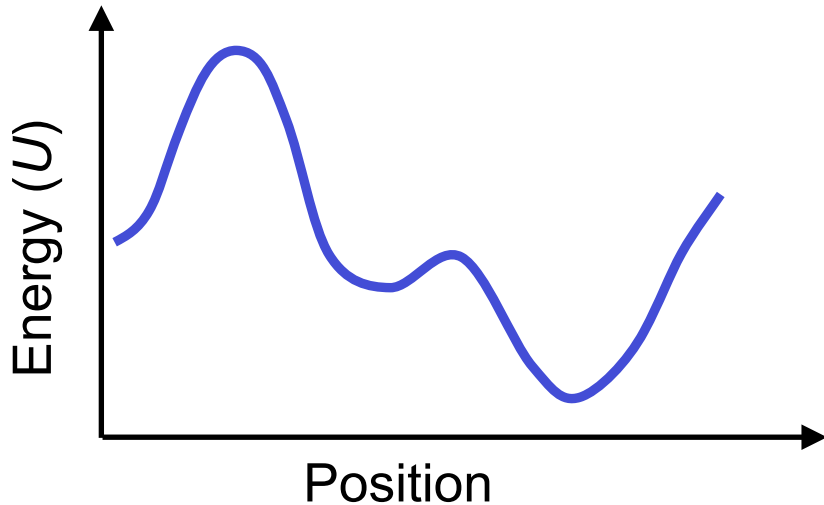
Specifying atom positions

- For a molecular system with N atoms, we can specify the position of all atoms by a single vector \mathbf{x} of length $3N$
 - This vector contains the x , y , and z coordinates of every atom

$$\mathbf{x} = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \\ x_2 \\ y_2 \\ z_2 \\ \vdots \\ x_N \\ y_N \\ z_N \end{pmatrix}$$

Energy function

- A potential energy function $U(\mathbf{x})$ specifies the total potential energy of a system of atoms as a function of all their positions (\mathbf{x})
 - In the general case, include not only atoms in the protein but also surrounding atoms (e.g., water)
- The potential energy function U is also called a *force field*, because one can use it to compute forces on atoms



Types of force fields (energy functions)

- A wide variety of force fields are used in atomic-level modeling of macromolecules
- Physics-based vs. knowledge-based
 - Physics-based force fields attempt to model actual physical forces
 - Knowledge-based force fields are based on statistics about, for example, known protein structures
 - Most real force fields are somewhere in between
- Atoms represented
 - Most realistic choice is to model all atoms
 - Some force fields omit waters and other surrounding molecules. Some omit certain atoms within the protein.

Energy functions for biomolecular
systems

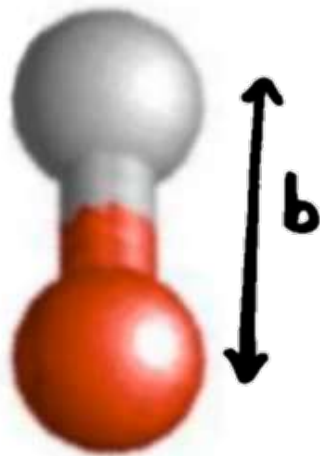
Molecular mechanics force fields

Molecular mechanics force fields

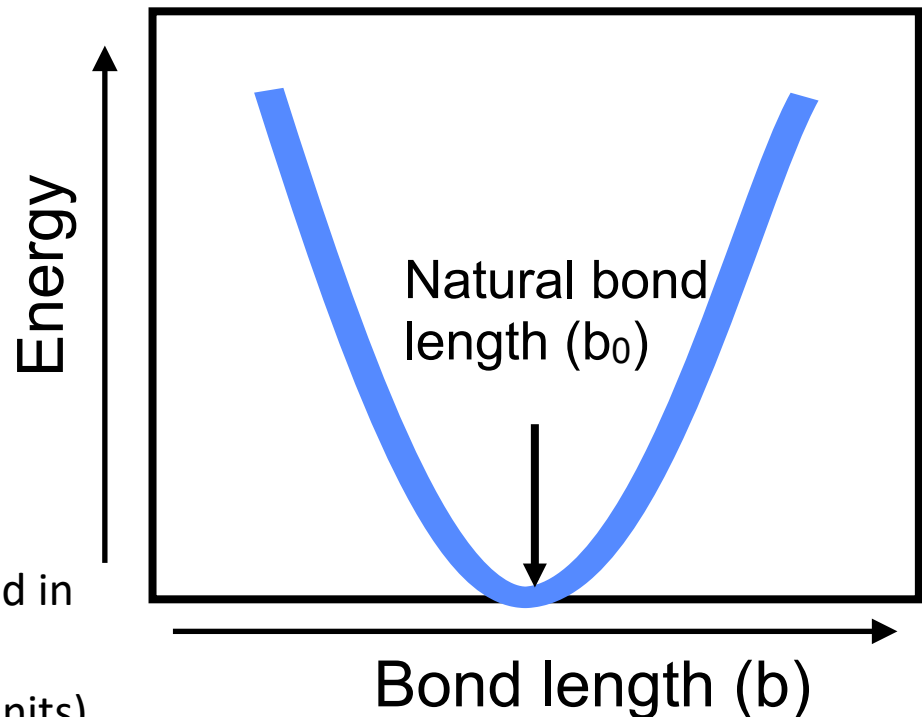
- Today, we'll focus on *molecular mechanics force fields*, which are often used for molecular simulations
- These are more toward the physics-based, all-atom end (i.e., the more “realistic” force fields)
 - Represent physical forces explicitly
 - Typically represent solvent molecules (e.g., water) explicitly
- We'll revisit the forces acting between atoms and write down the functional forms typically used to approximate them

Bond length stretching

- A bonded pair of atoms is effectively connected by a spring with some preferred (natural) length. Stretching or compressing it requires energy.



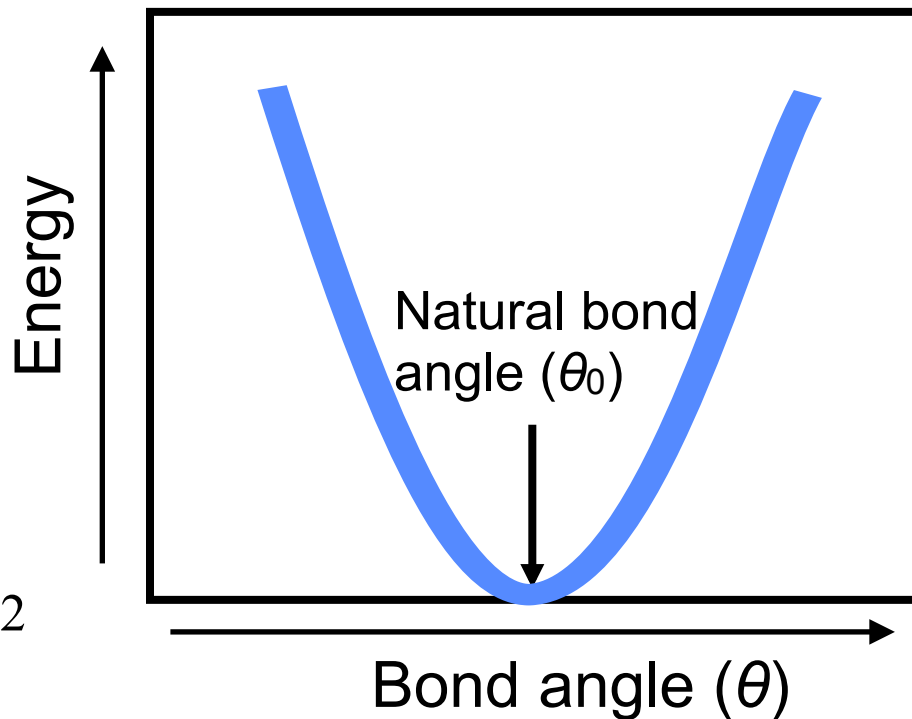
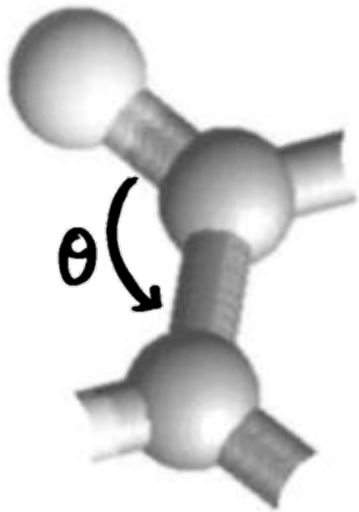
$$U(b) = k_b (b - b_0)^2$$



Note: A factor of 1/2 is sometimes included in this equation. I'm ignoring such constant factors (they can be folded into k_b or the units).

Bond angle bending

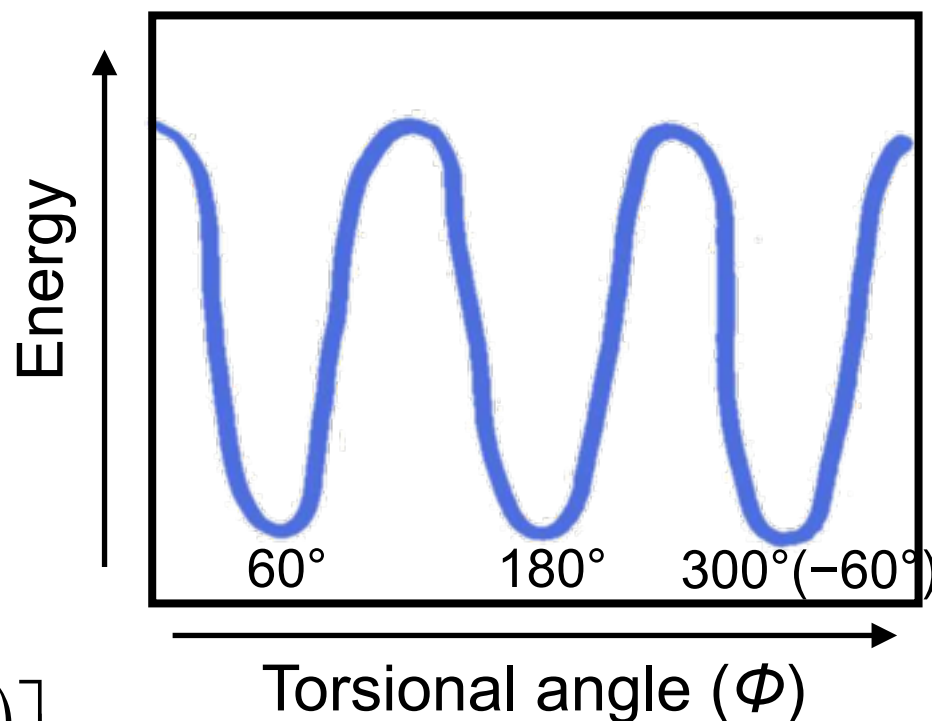
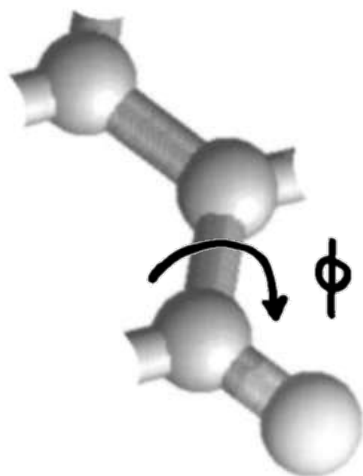
- Likewise, each bond angle has some natural value. Increasing or decreasing it requires energy.



$$U(\theta) = k_{\theta} (\theta - \theta_0)^2$$

Torsional angle twisting

- Certain values of each torsional angle are preferred over others.

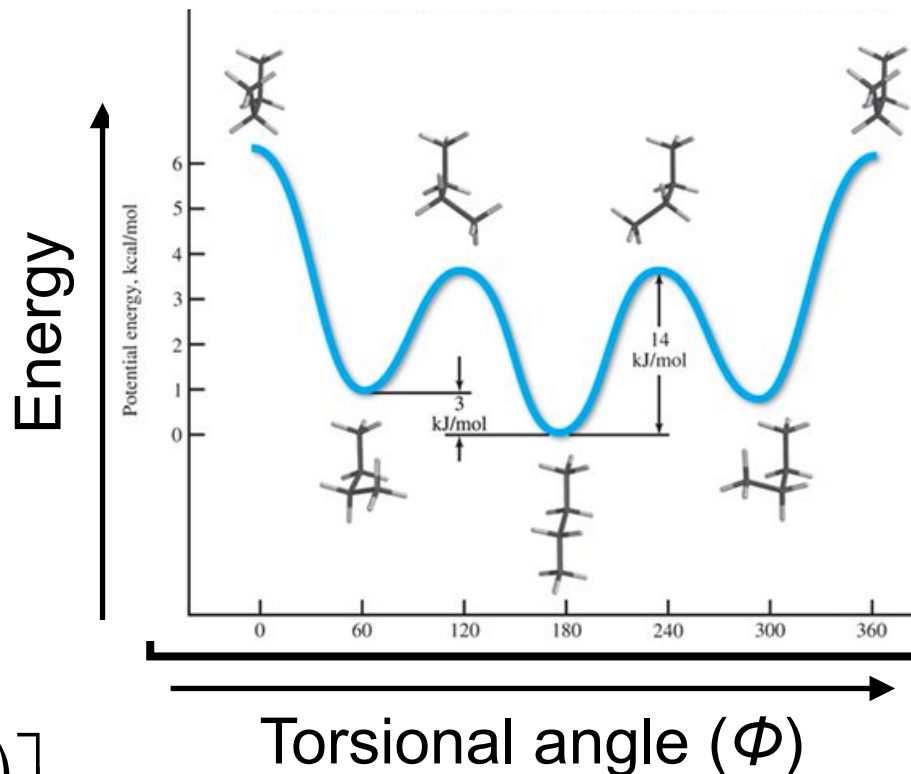
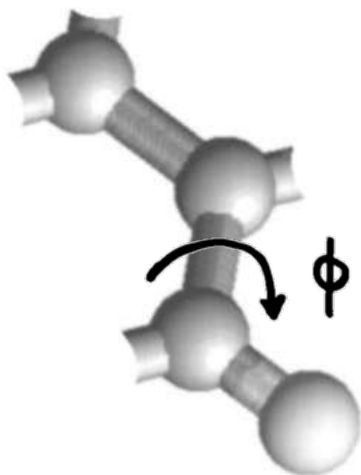


$$U(\phi) = \sum_n k_{\phi,n} \left[1 + \cos(n\phi - \phi_n) \right]$$

Typically n takes on one or a few values between 1 and 6

Torsional angle twisting

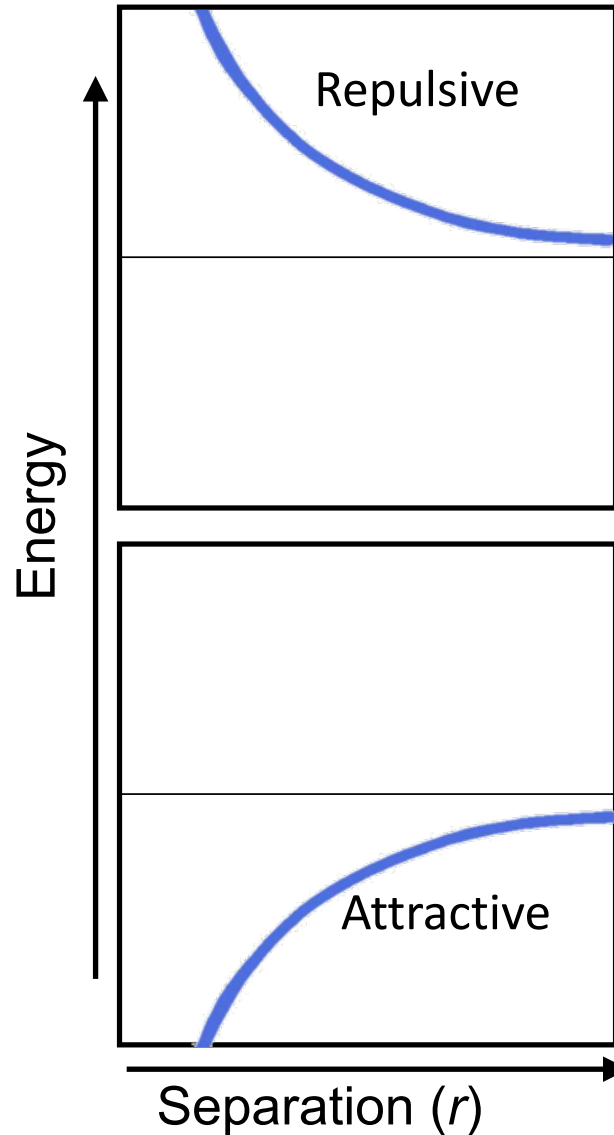
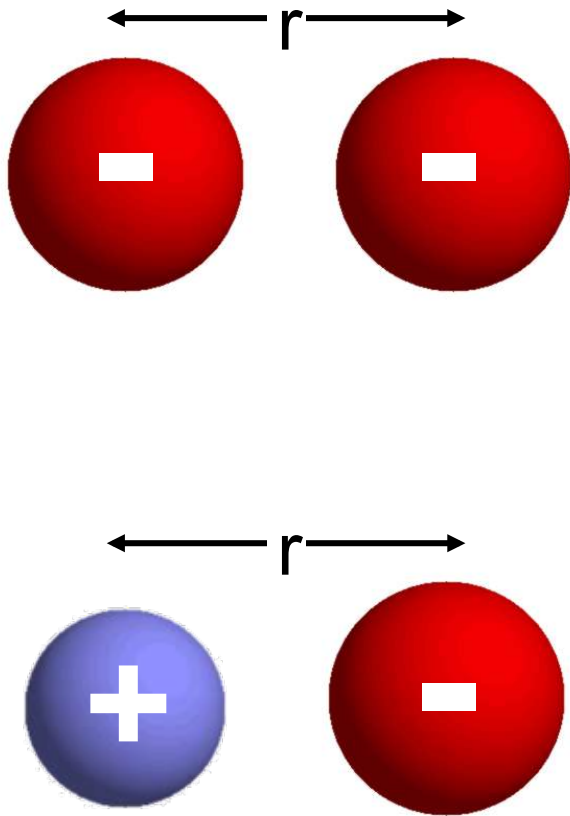
- Certain values of each torsional angle are preferred over others.



$$U(\phi) = \sum_n k_{\phi,n} \left[1 + \cos(n\phi - \phi_n) \right]$$

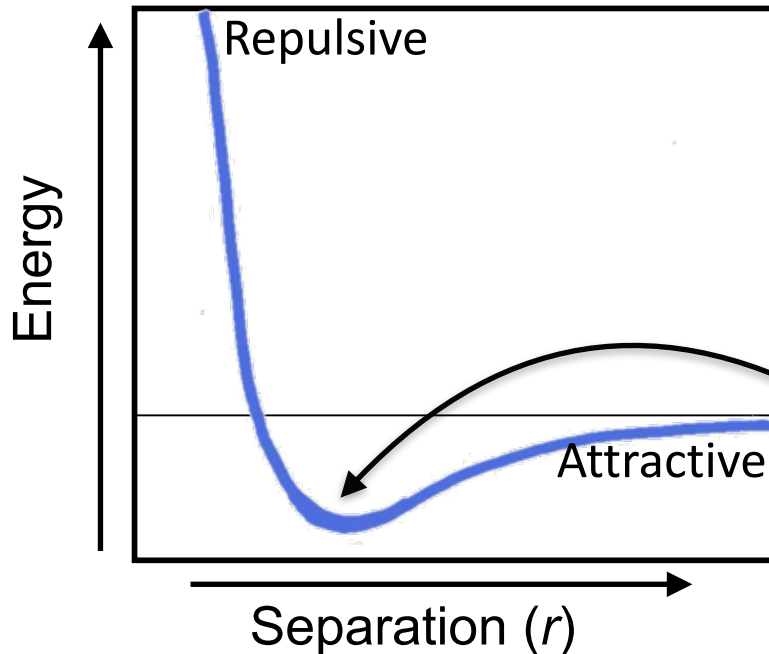
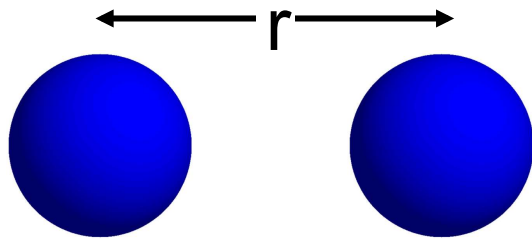
Typically n takes on one or a few values between 1 and 6

Electrostatics interaction



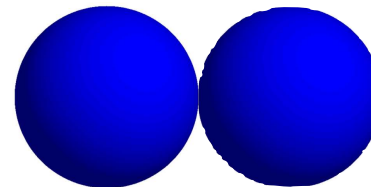
- Like charges repel. Opposite charges attract.
 - Acts between all pairs of atoms, including those in different molecules.
 - Each atom carries some “partial charge” (may be a fraction of an elementary charge), which depends on which atoms it's connected to
- $$U(r) = \frac{q_i q_j}{r}$$
- where q_i and q_j are partial charges on atoms i and j

van der Waals interaction

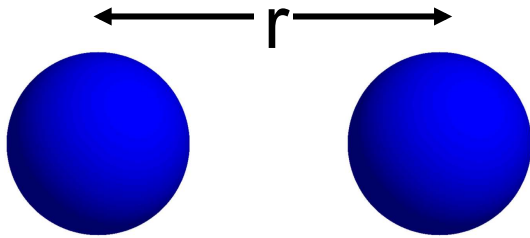


- van der Waals forces act between all pairs of atoms and do not depend on charge.
- When two atoms are too close together, they repel strongly.
- When two atoms are a bit further apart, they attract one another weakly.

Energy is minimal when atoms are “just touching” one another



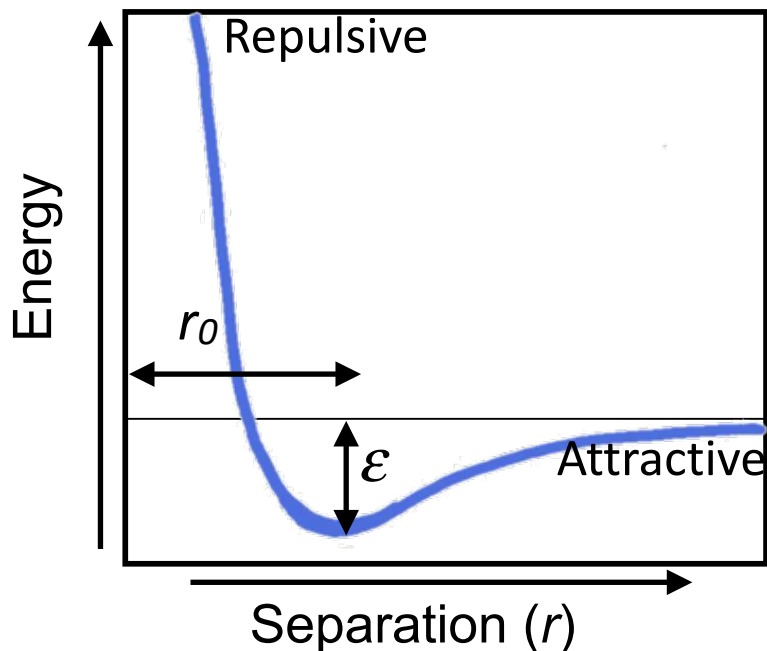
van der Waals interaction



$$U(r) = \frac{A_{ij}}{r^{12}} - \frac{B_{ij}}{r^6}$$

We can also write this as:

$$U(r) = \epsilon \left[\left(\frac{r_0}{r} \right)^{12} - 2 \left(\frac{r_0}{r} \right)^6 \right]$$



Note: Historically, r^{12} term was chosen for computational convenience; other forms are sometimes used

A typical molecular mechanics force field

$$U = \sum_{\text{bonds}} k_b (b - b_0)^2$$

$$+ \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2$$

$$+ \sum_{\text{torsions}} \sum_n k_{\phi,n} \left[1 + \cos(n\phi - \phi_n) \right]$$

Bond lengths (“Stretch”)

Bond angles (“Bend”)

**Bonded
terms**

Torsional/dihedral angles

$$+ \sum_i \sum_{j>i} \frac{q_i q_j}{r_{ij}}$$

$$+ \sum_i \sum_{j>i} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6}$$

Electrostatics

Van der Waals

**Non-
bonded
terms**

How are the parameters fit?

- Combination of:
 - Quantum mechanical calculations
 - Experimental data
 - For example: b_0 can be estimated from x-ray crystallography, and K_b from spectroscopy (infrared absorption)

$$U(b) = K_b (b - b_0)^2$$

- The torsional parameters are usually fit last. They absorb the “slop.” Fidelity to physics is debatable.
- These force fields are approximations!

Neural network–based force fields

- Researchers have recently begun developing force fields by training neural networks to predict results of quantum mechanical calculations
 - See optional reading on course website

What does the energy function tell us about biomolecular structure/conformation?

What does the energy function tell us about biomolecular conformation?

The Boltzmann distribution

Relating energy to probability

- Given the potential energy associated with a particular conformation (i.e., arrangement of atoms, or set of atomic coordinates), what is the probability that the molecular system will adopt that conformation at a given point in time?
- Assumptions:
 - System is at constant temperature (so atoms are constantly jiggling around).
 - We watch the system for a really long time (allowing it to fully equilibrate).

The Boltzmann Distribution

- The Boltzmann distribution relates the potential energy of a particular arrangement of atoms to the probability of observing that arrangement of atoms (at equilibrium):

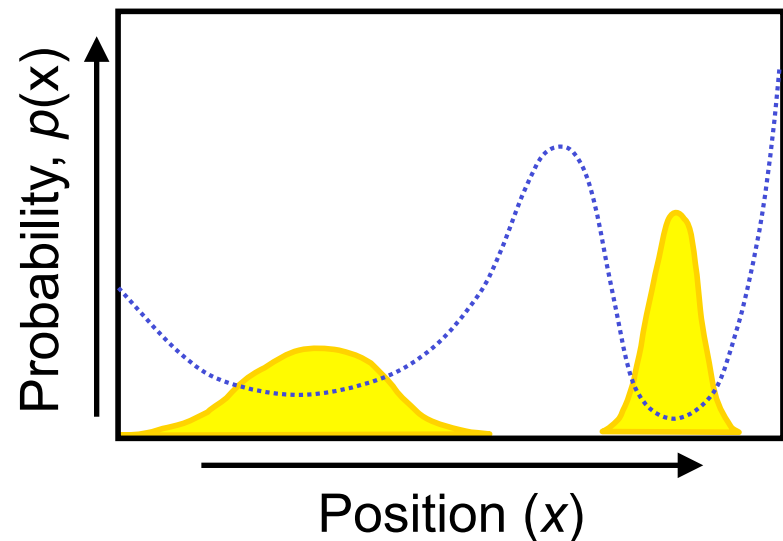
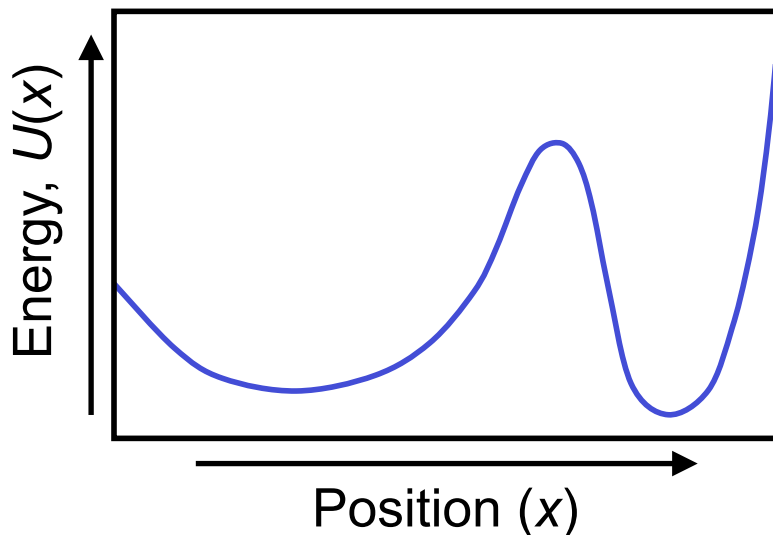
$$p(\mathbf{x}) \propto \exp\left(\frac{-U(\mathbf{x})}{k_B T}\right)$$

Equivalently,

$$p(\mathbf{x}) = \frac{1}{Z} \exp\left(\frac{-U(\mathbf{x})}{k_B T}\right)$$

where T is temperature and k_B is the Boltzmann constant

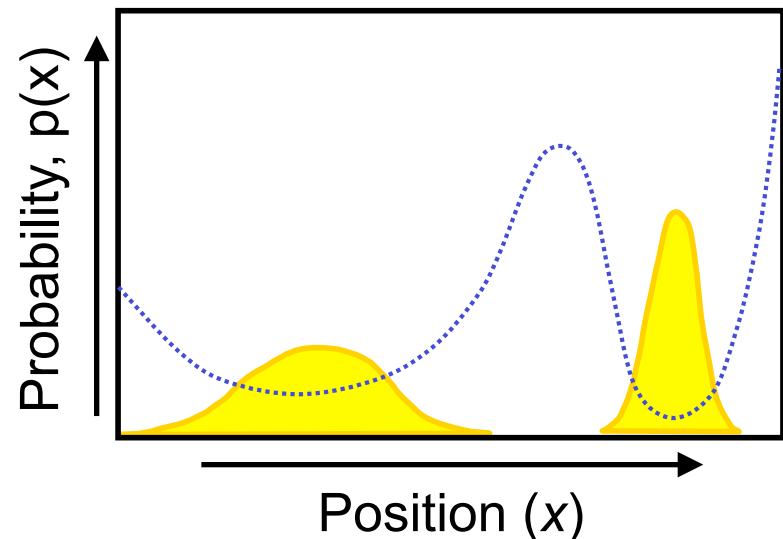
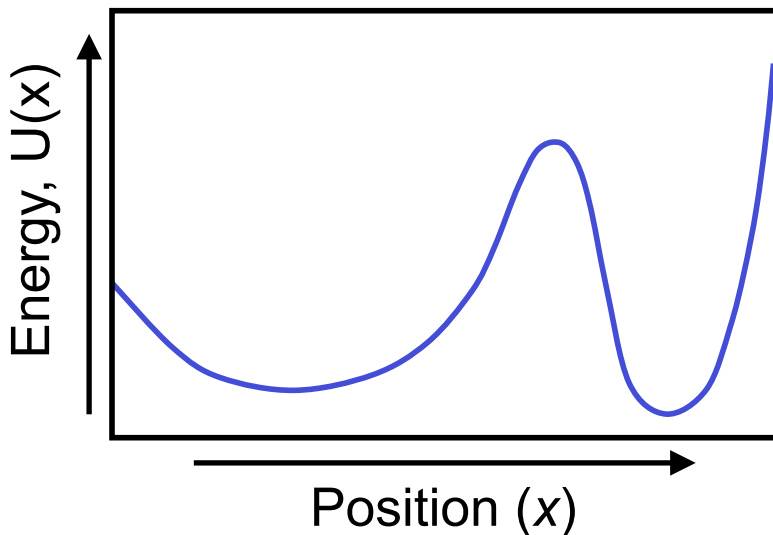
- Note: Z is chosen such that the probabilities sum to 1 across all arrangements of atoms. It depends on U and T but not on \mathbf{x} .



The Boltzmann Distribution

- Key properties:
 - Higher energy gives lower probability
 - Exponential relationship: as energy increases, probability goes down quickly
 - Temperature dependence: increasing temperature decreases differences in probability between high-energy and low-energy conformations

$$p(\mathbf{x}) \propto \exp\left(\frac{-U(\mathbf{x})}{k_B T}\right)$$



What does the energy function tell us about biomolecular conformation?

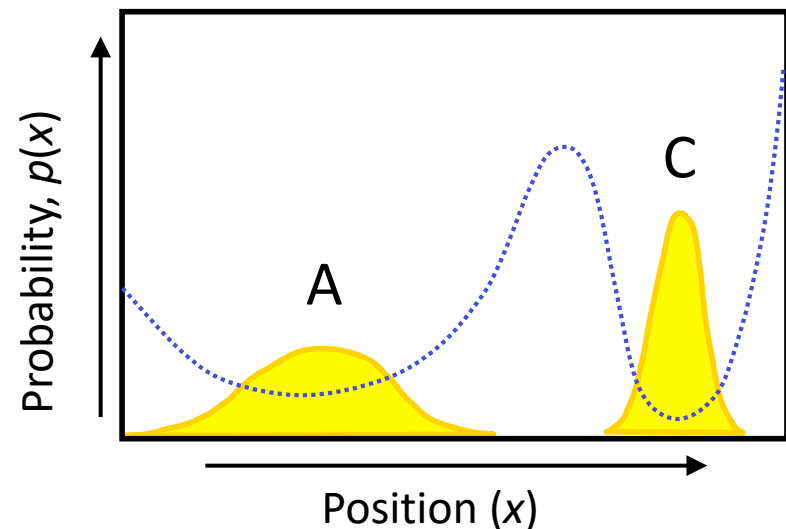
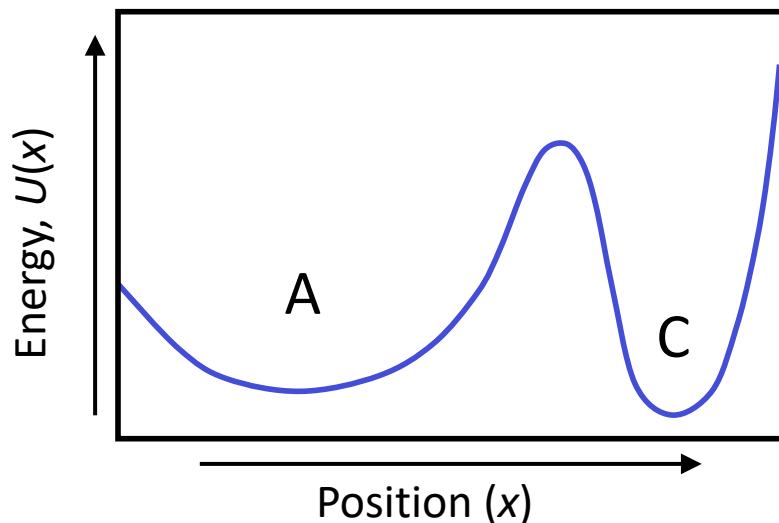
Conformations and conformational states

Protein (or other biomolecular) structure: what we care about

- We **don't** really care about the probability that all the atoms of the protein and all the surrounding water atoms will be in one precise arrangement
- Instead, we care about the probability that protein atoms will be in some *approximate* arrangement, with *any* arrangement of surrounding water

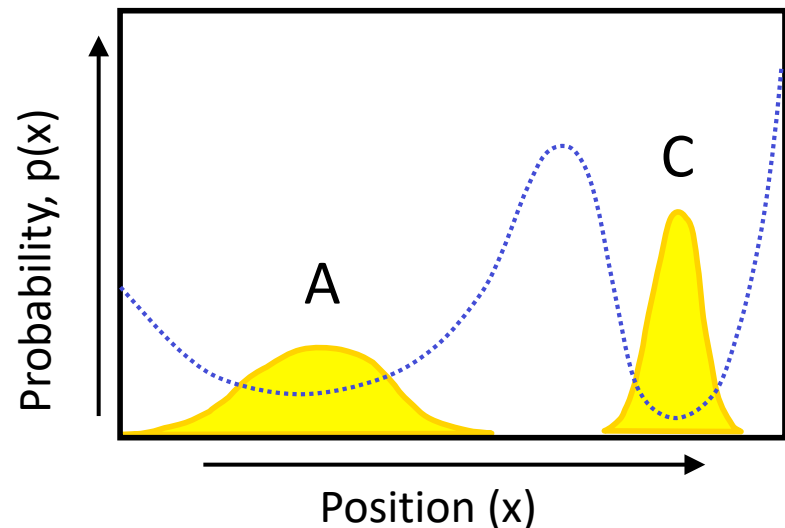
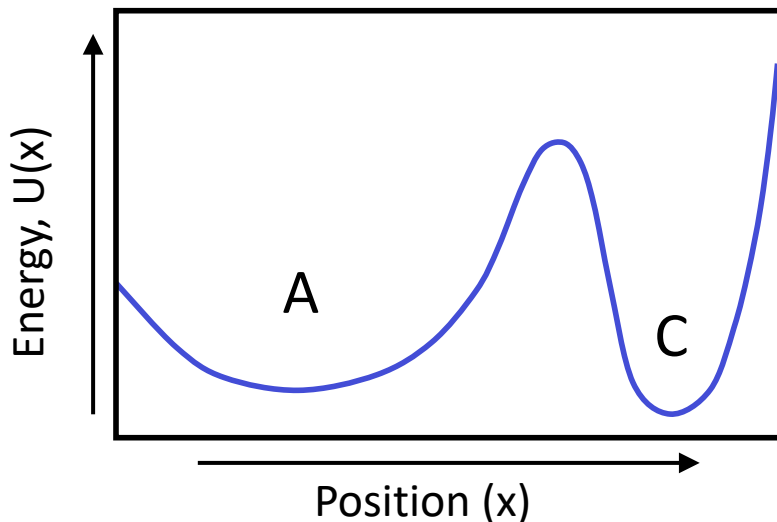
Protein (or other biomolecular) structure: what we care about

- In other words, we wish to compare probabilities of different sets (neighborhoods) of atomic arrangements
- We define each of these sets as a *conformational state* (A, C). Each *conformational state* includes many *conformations*, or specific atom arrangements \mathbf{x} .
 - In this example, conformational states correspond to wells in the energy landscape
 - A more general term for “conformational state” is “macrostate,” and a more general term for “conformation” is “microstate”



Probabilities of conformational states

- Which has greater probability, A or C?
 - C is a deeper well, so the individual atomic arrangements within it are more likely
 - A is a broader well, so it includes more distinct individual arrangements

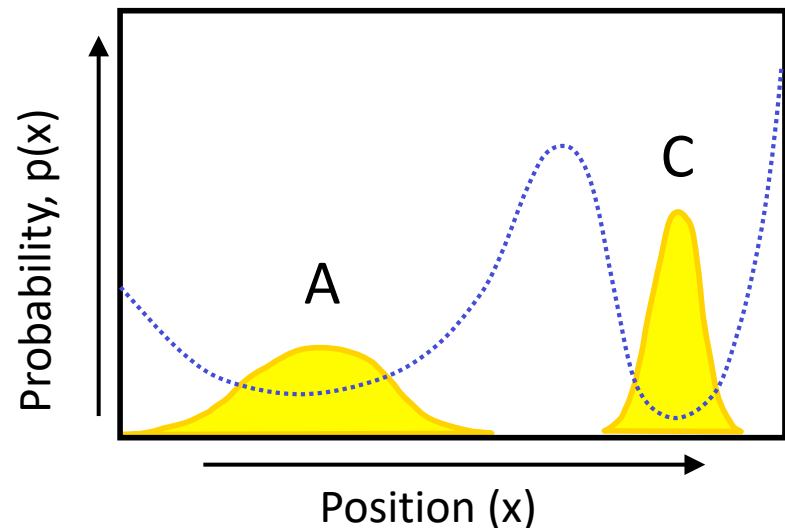
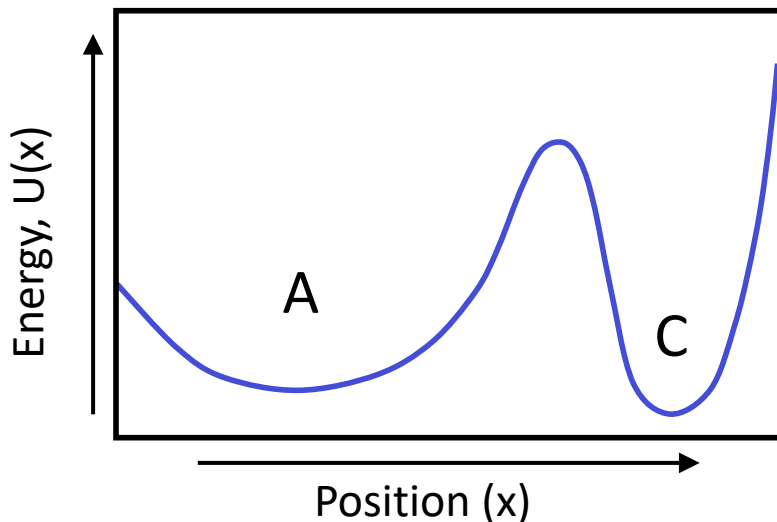


Probabilities of conformational states

- Which has greater probability, A or C?
- To get probability of a conformational state, sum/integrate over all conformations within it

$$P(A) = \int_{x \in A} P(\mathbf{x}) \propto \int_{x \in A} \exp\left(\frac{-U(\mathbf{x})}{k_B T}\right) d\mathbf{x}$$

- At low temperature, $P(C) > P(A)$
- At high temperature, $P(A) > P(C)$



What does the energy function tell us about biomolecular conformation?

Free energy

Free energy of a conformational state

- So far we have assigned energies only to individual conformations, but it's useful to assign them to conformational states as well.
- Define the *free energy* G_A of a conformational state A such that:

$$P(A) = \exp\left(\frac{-G_A}{k_B T}\right)$$

- This is analogous to Boltzmann distribution formula:

$$p(\mathbf{x}) \propto \exp\left(\frac{-U(\mathbf{x})}{k_B T}\right)$$

- **Key takeaway: Free energy is for a conformational state (i.e., set of conformations) what potential energy is for an individual conformation**

So which conformational state will a biomolecule (e.g., protein) adopt?

- The one with the *minimum free energy*
 - Wide, shallow wells often win out over narrow, deep ones
- This depends on temperature
- At room or body temperature, the conformational state (macrostate) of minimum free energy is usually very different from the conformation with minimum potential energy

Comparing structures (conformations) of a biomolecule

- The most common measure of the similarity/difference between two structures of the same molecule is *root mean squared deviation (RMSD)*, defined as

$$\sqrt{\frac{1}{N} \sum_{i=1}^{3N} (\mathbf{x}_i - \mathbf{w}_i)^2}$$

where N is the number of atoms, \mathbf{x} gives the coordinates for one structure, and \mathbf{w} gives the coordinates for the other structure.

- We generally want to align the structures, which can be done by finding the rigid-body rotation and translation of one structure that will minimize its RMSD from the other
 - **The relevant measure of similarity is RMSD *after* alignment**