Structural Bioinformatics

GENOME 541
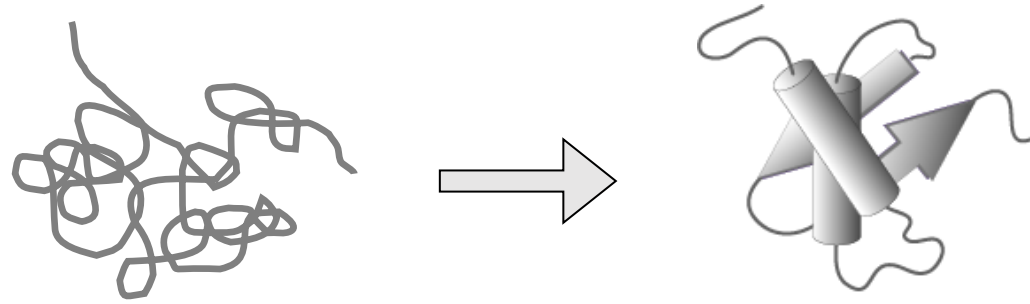Spring 2023

Lecture 2: Biomolecular
Energy Functions
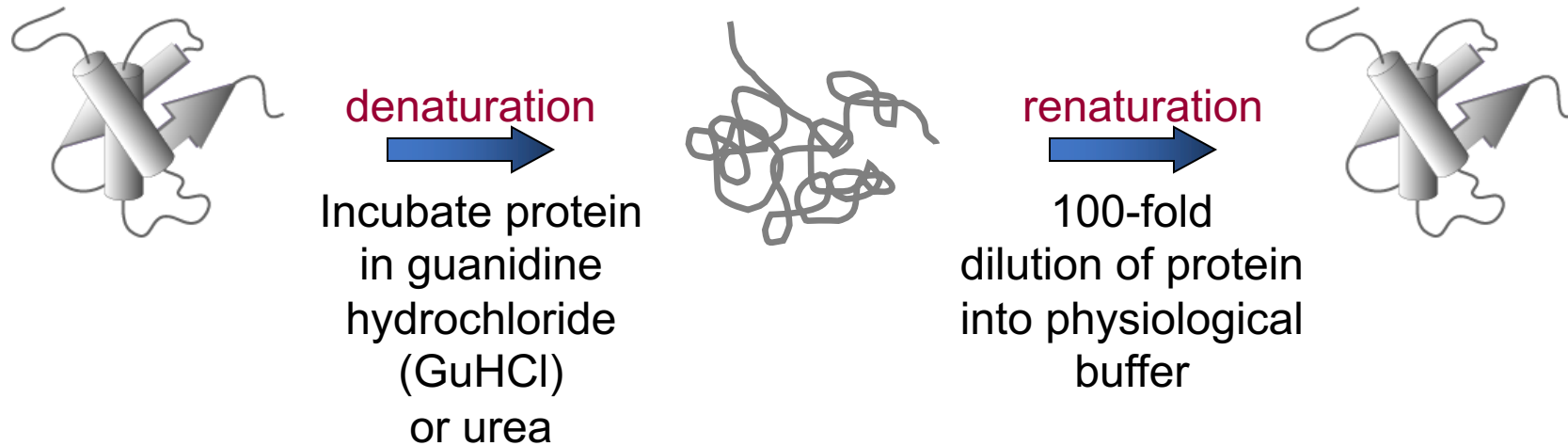
Frank DiMaio (dimaio@uw.edu)

Protein Folding

The process by which a protein goes from being an unfolded polymer with no activity to a uniquely structured and active protein.

Why do we care about protein folding?

- Understanding how protein's folds informs us of *sequence to structure* mapping

- Protein misfolding has been implicated in many human diseases (e.g. Alzheimer's, Parkinson's)

Protein folding *in vitro* is often reversible

denaturation

Incubate protein
in guanidine
hydrochloride
(GuHCl)
or urea

renaturation

100-fold
dilution of protein
into physiological
buffer

- **the amino acid sequence of a polypeptide is sufficient to specify its three-dimensional conformation**

- protein folding is a spontaneous process that does not require the assistance of extraneous factors

Anfinsen, CB (1973) Principles that govern the folding of protein chains. *Science* **181**, 223-230.
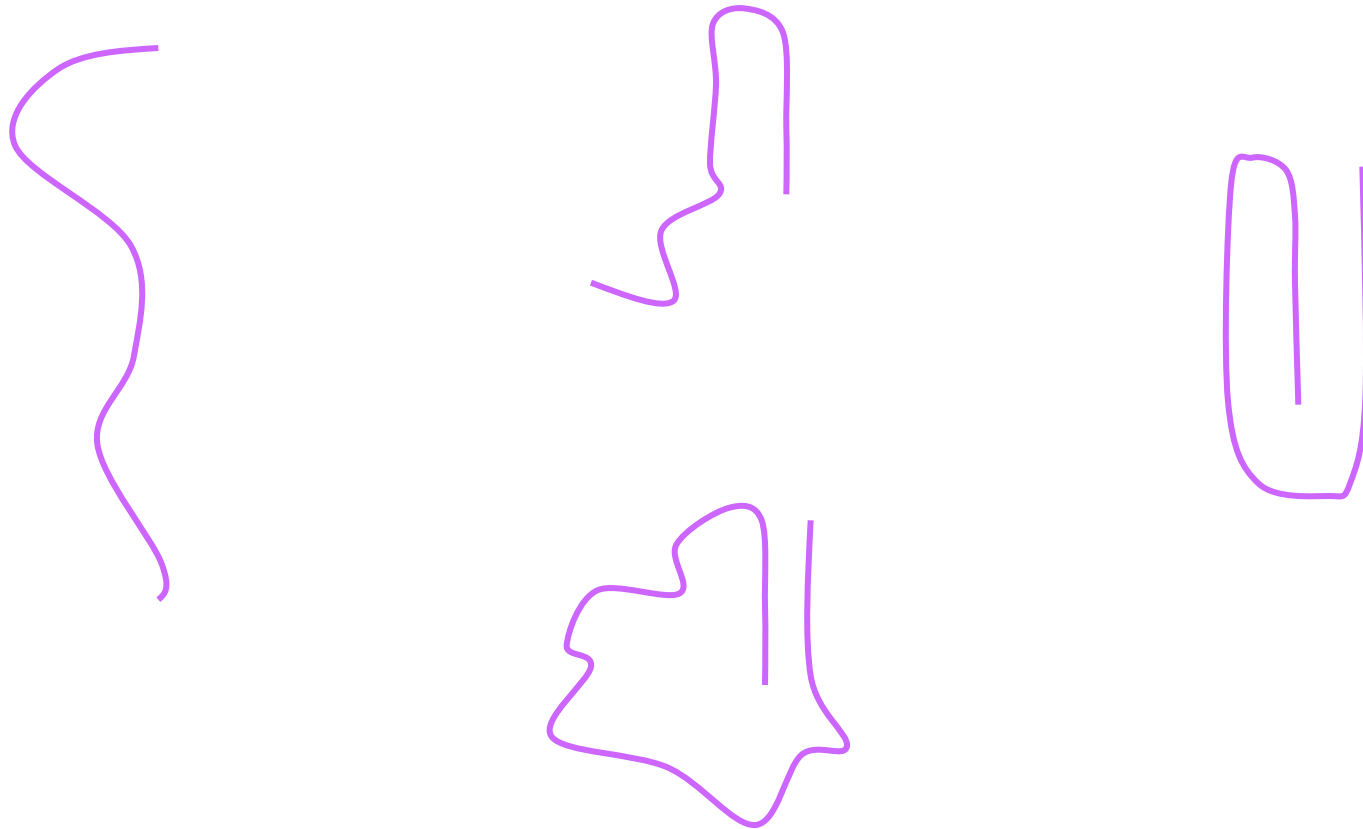
How Do Proteins Fold?

- Cyrus Levinthal tried to estimate how long it would take a protein to do a random search of conformational space for the native fold.

- Imagine a 100-residue protein with three possible conformations per residue. Thus, the number of possible folds = 3^{100} = 5 x 10^{47}.

- Let us assume that protein can explore new conformations at the same rate that bonds can reorient
  (10^{13} structures/second).

- Thus, the time to explore all of conformational space = 5 x 10^{47}/10^{13} = 5 x 10^{34} seconds = 1.6 x 10^{27} years >> age of universe

- This is known as the **Levinthal paradox.**
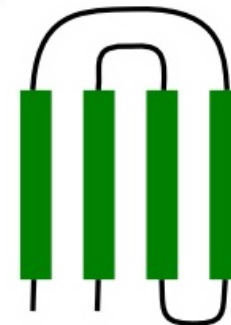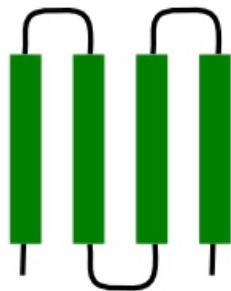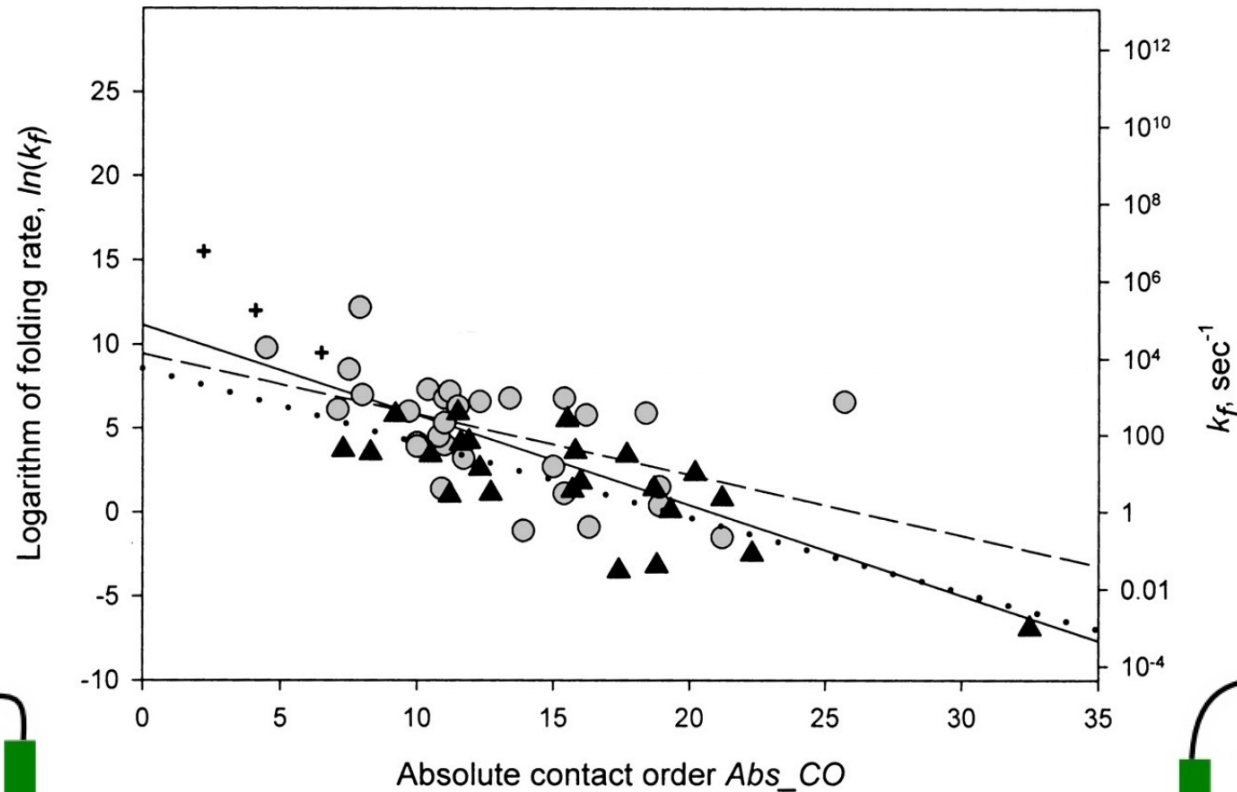
How do proteins fold?

Do proteins fold by a very discrete pathway?

Flat landscape (Levinthal paradox)

Tunnel landscape (discrete pathways)

Realistic landscape ("folding funnel")
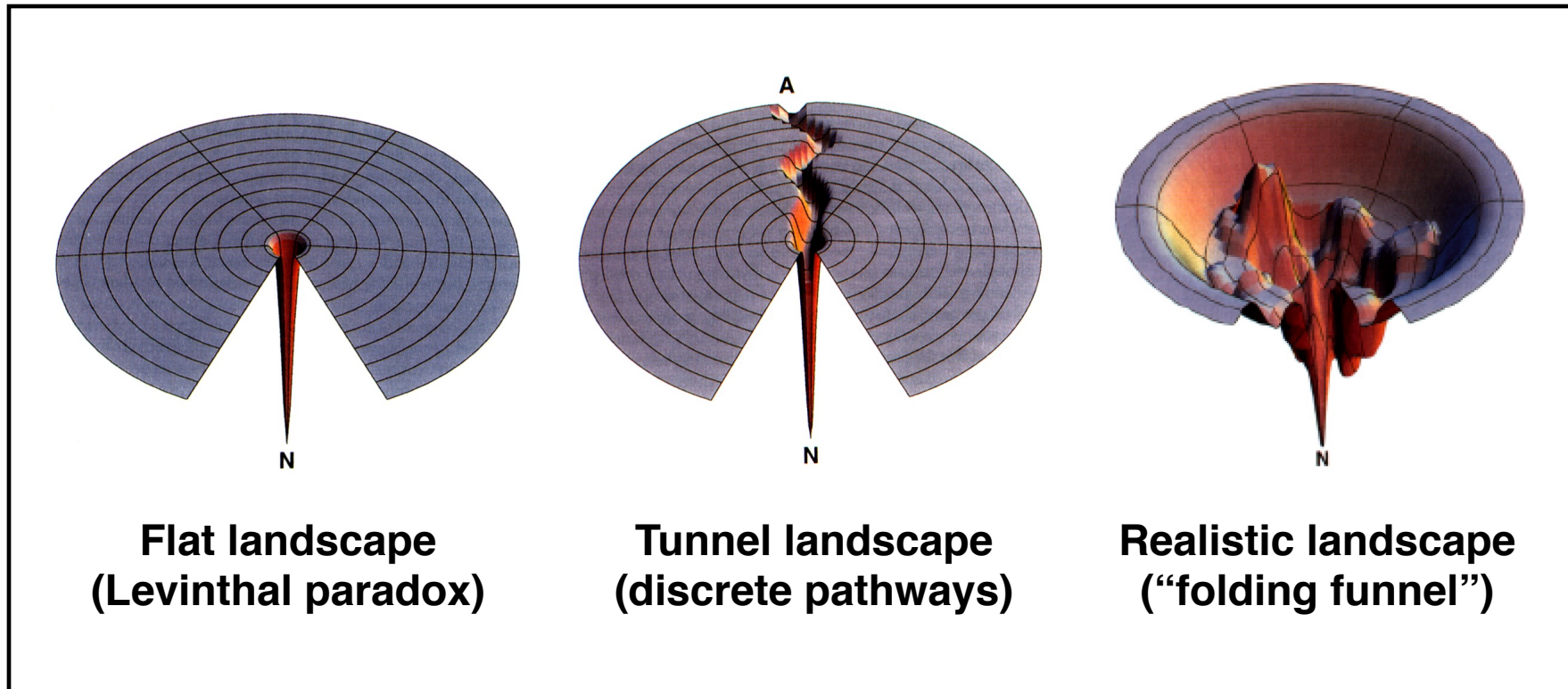
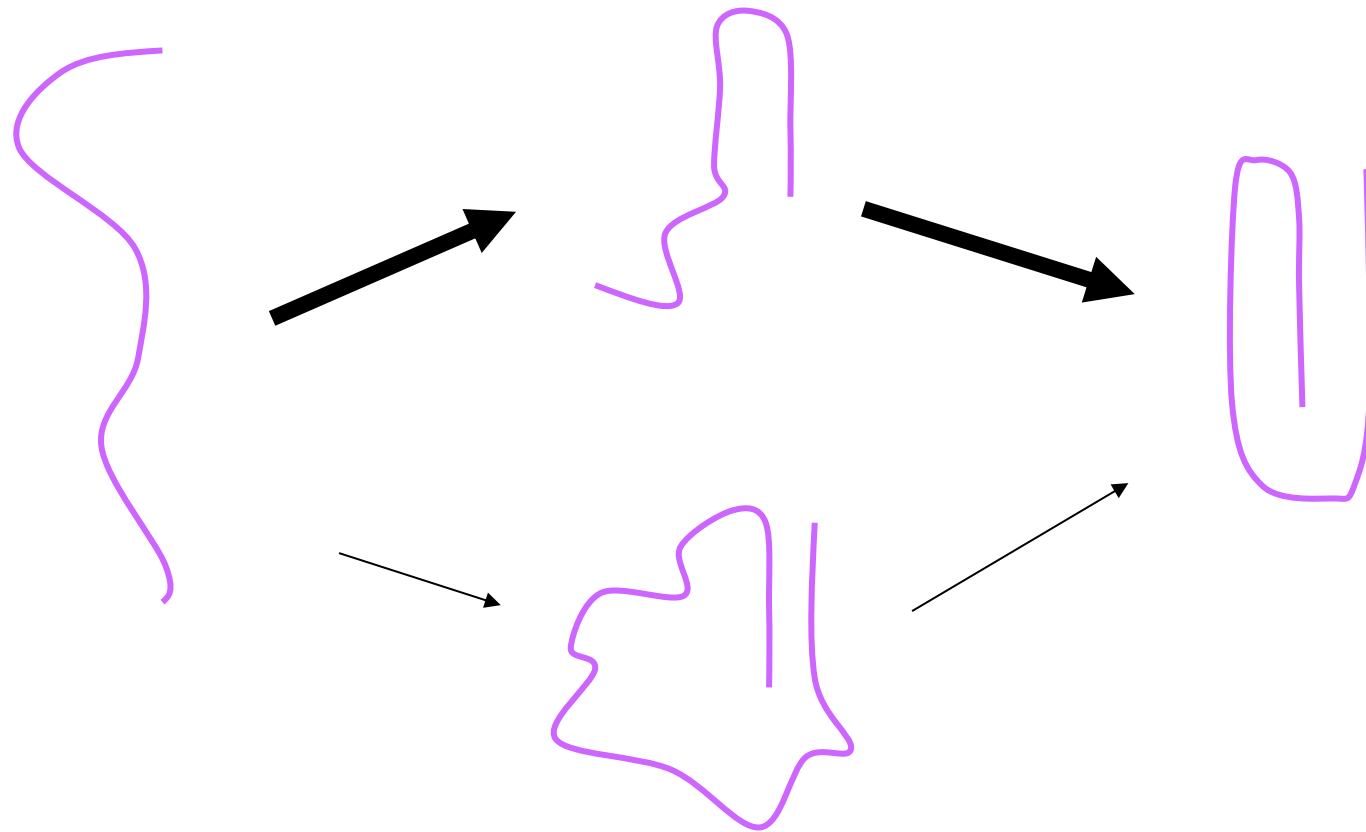# Do certain portions of a protein fold first?

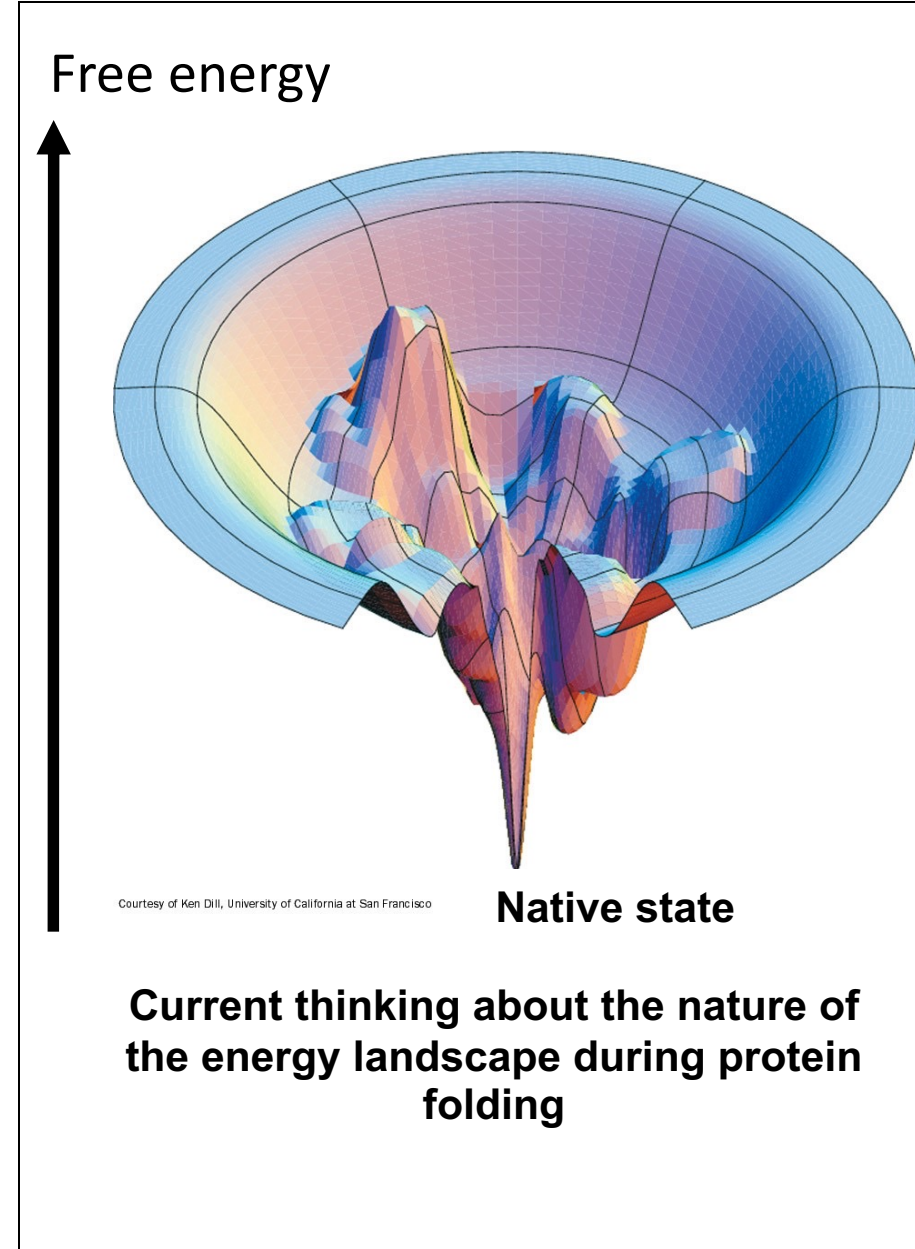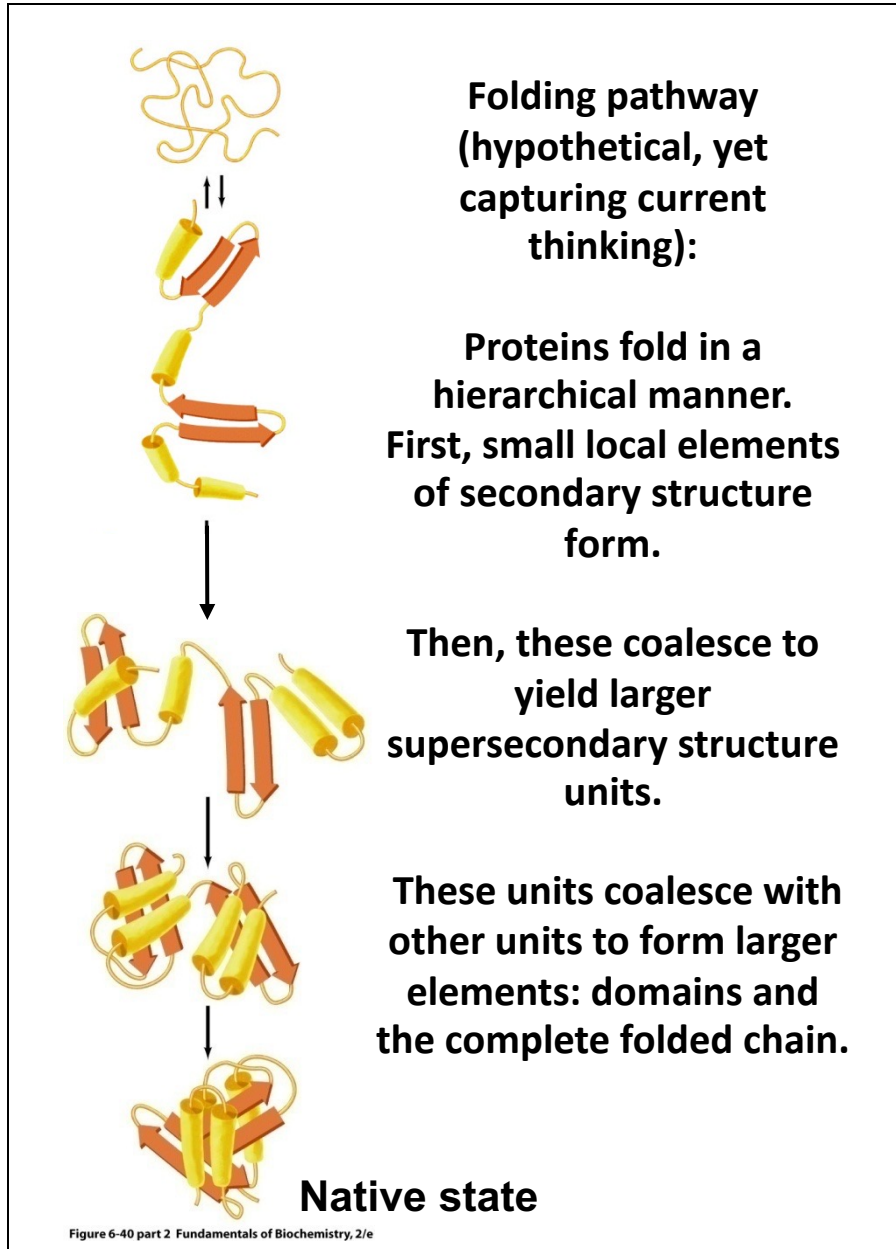# Protein folding rates correspond with contact order

$$CO = \frac{1}{L \cdot N} \sum^{N} \Delta S_{i,j}$$

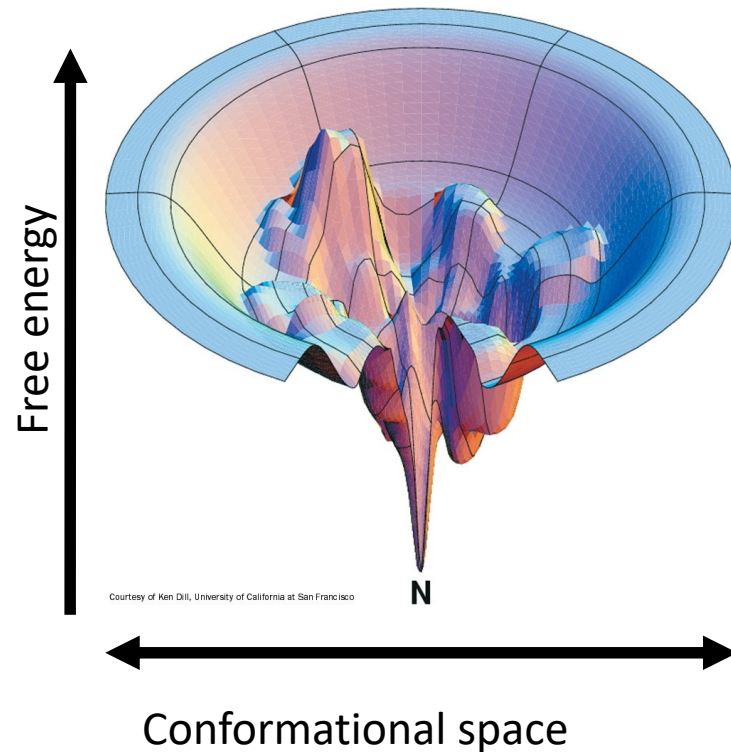# Do certain portions of a protein fold first?

Interactions between residues *close to each other along the polypeptide chain* are more likely to form early in folding.

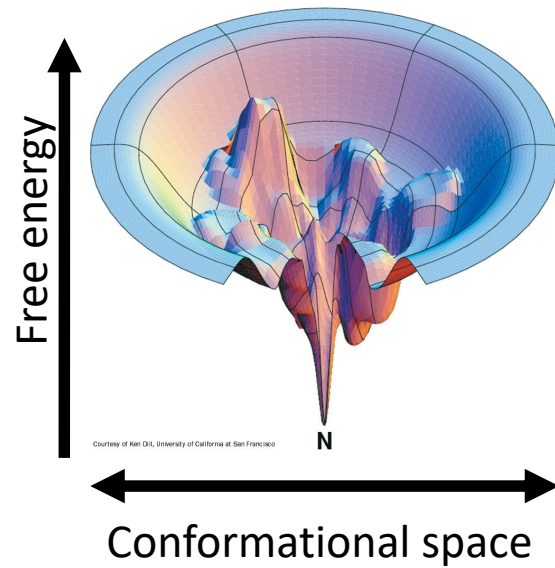# Folding pathways and energy landscapes in protein folding

Folding pathway (hypothetical, yet capturing current thinking):

Proteins fold in a hierarchical manner. First, small local elements of secondary structure form.

Then, these coalesce to yield larger supersecondary structure units.

These units coalesce with other units to form larger elements: domains and the complete folded chain.

Native state

Figure 6-40 part 2 Fundamentals of Biochemistry, 2/e

Free energy

Courtesy of Ken Dill, University of California at San Francisco

Native state

Current thinking about the nature of the energy landscape during protein folding

Modeling the protein free energy landscape

Free energy

Courtesy of Ken Dill, University of California at San Francisco
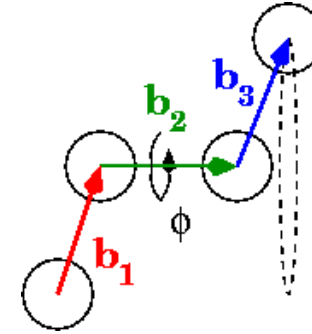
N

Conformational space

- Under *Anfinsen's hypothesis*, the state of lowest free energy is the native state

- Represent the various enthalpic and entropic effects governing folding with *parameterized equations*
    - vdW interactions
    - electrostatic interactions
    - solvent entropy
    - etc.

- **Predicting protein structure** involves identifying the <u>lowest-energy state</u> of the protein
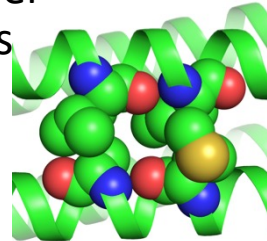
Modeling the protein free energy landscape

Free energy

Conformational space

N

Courtesy of Ken Dill, University of California at San Francisco

Bonded interactions

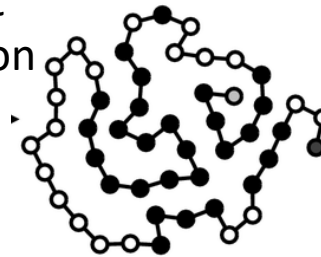
b_3

b_2

b_1

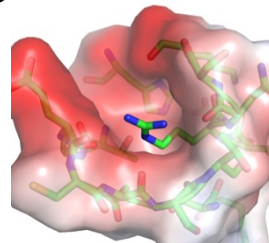ϕ

Non-bonded interactions
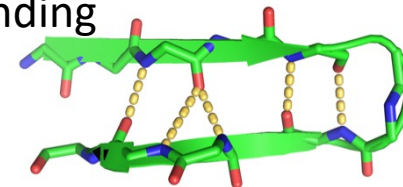
van der Waals

Implicit solvation

Electro-statics

Hydrogen bonding

Modeling covalent forces

Bond lengths

$$V_{bond} = K_b(b - b_0)^2$$

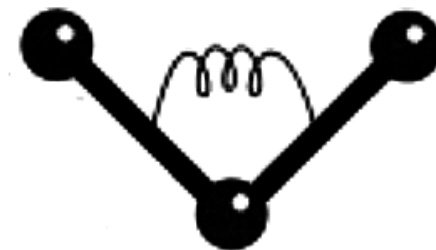K_b = force constant
b_0 = equilibrium length

Chemical type	K_{bond}	b_o
C-C	100 kcal/mole/Å²	1.5 Å
C=C	200 kcal/mole/Å²	1.3 Å
C=-C	400 kcal/mole/Å²	1.2 Å

Bond angle

$$V_{angle} = K_\theta(\theta - \theta_0)^2$$

K_θ = force constant
θ_0 = equilibrium angle

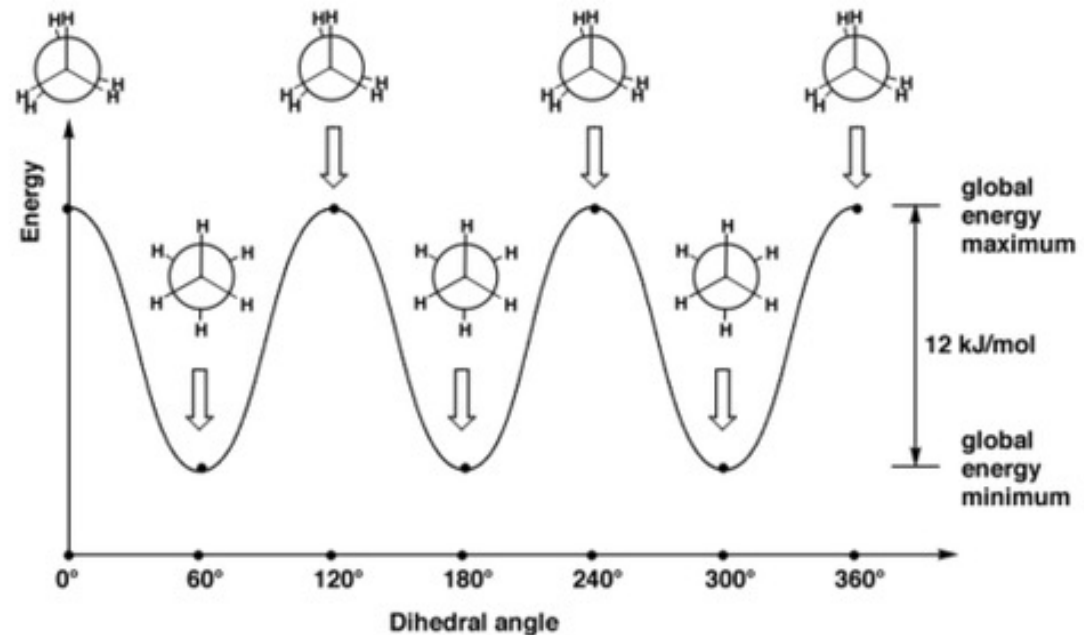# Modeling covalent forces

Torsion angle

- Staggered conformations
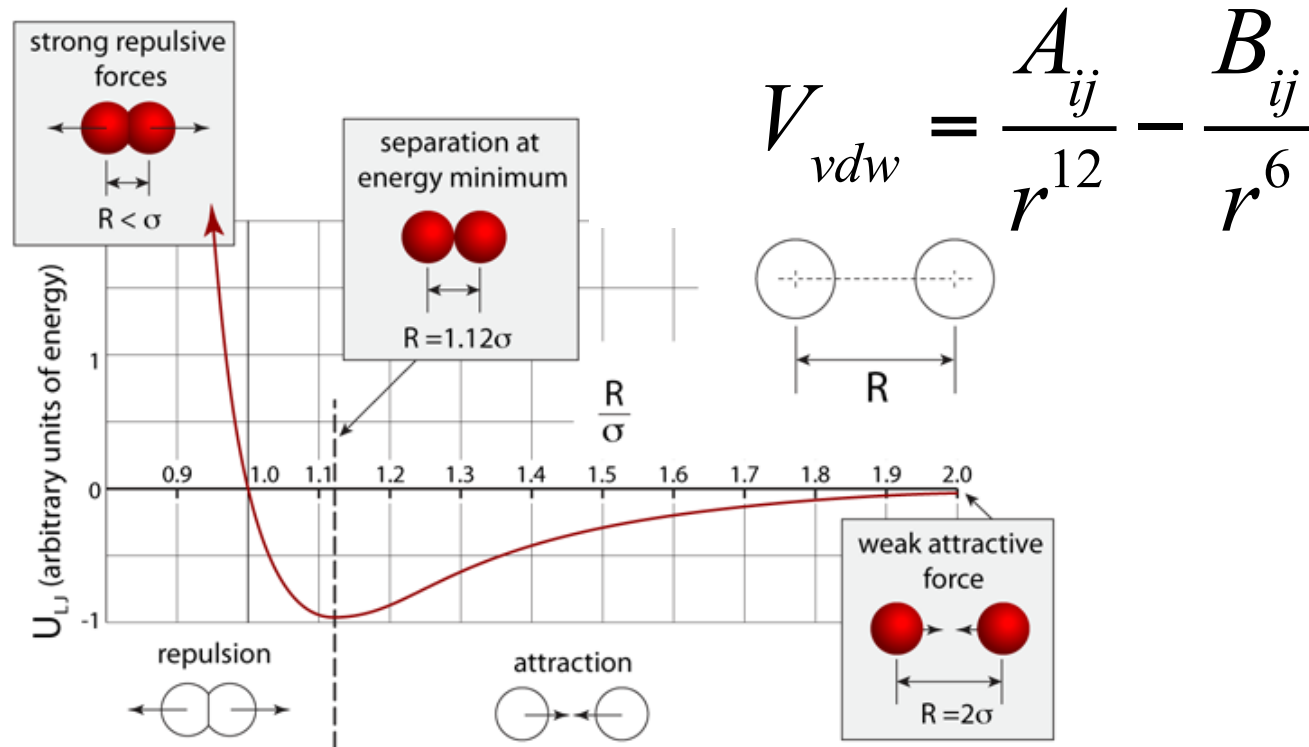 (angle +60, -60 or 180 are preferred).

$$V_{torsion} = \sum_n k_n \cos(n\phi)$$

Nonbonded forces

Van der Waals forces

- Interactions between nonbonded atoms are expressed by the Lennard-Jones potential.

- Very high repulsive force if atoms closer than van der Waals radii; attractive force if distance greater
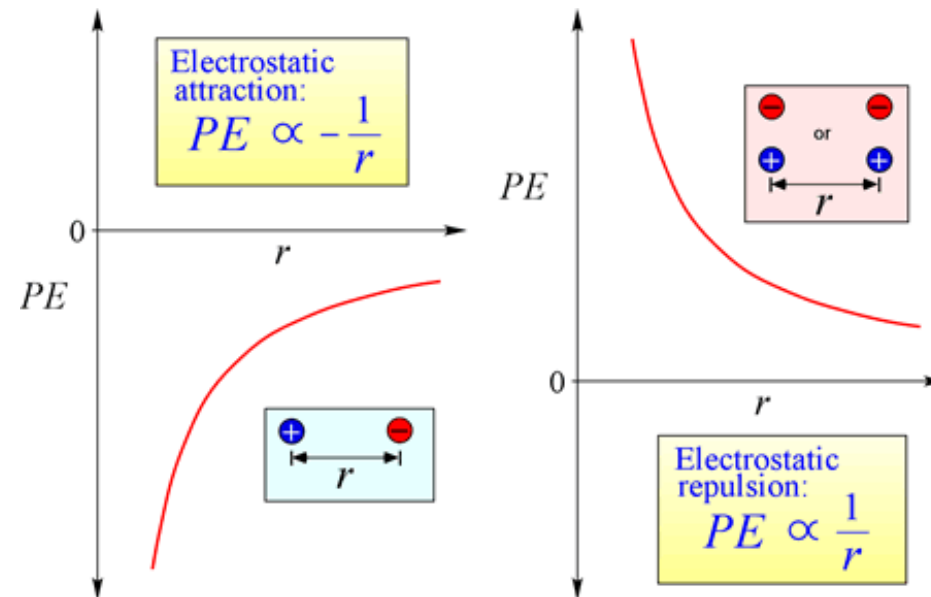
$$V_{vdw} = \frac{A_{ij}}{r^{12}} - \frac{B_{ij}}{r^{6}}$$

Nonbonded forces

Electrostatic interactions

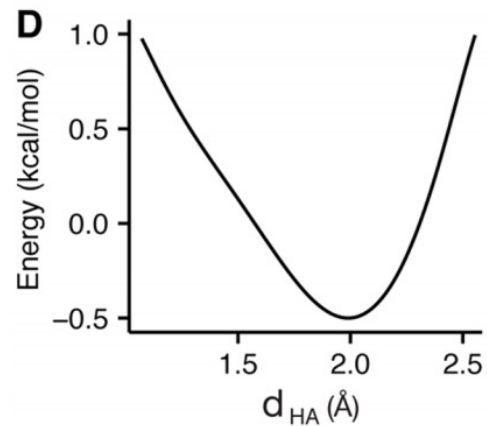- Approximate dipoles by giving atoms a partial charge

- Dielectic constant varies according to media: Ɛ=80 for water, and 4-6?? in the core of protein

- Electrostatic energy falls off much less quickly than for van der Waals interactions (chemically significant at ~15Å)

$$V_{electrostatics} = k_e \frac{q_1 q_2}{\varepsilon r}$$

Electrostatic attraction:
$$PE \propto -\frac{1}{r}$$

Electrostatic repulsion:
$$PE \propto \frac{1}{r}$$

Nonbonded forces

Hydrogen bonding

Modeling the interactions of protein and solvent

A — Isotropic Solvation

B — Anisotropic Solvation

C — In Line with 1HD2

D — Between 1HD1 and 1HD2

E — Out of Plane

fa_sol
lk_ball
lk_ball_wtd
+fa_sol
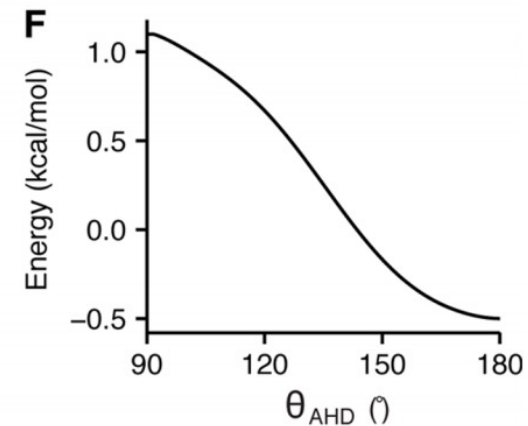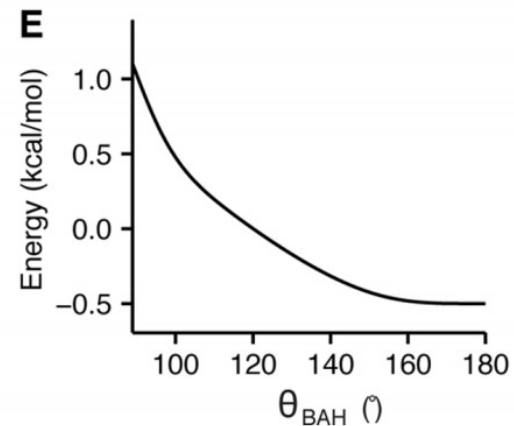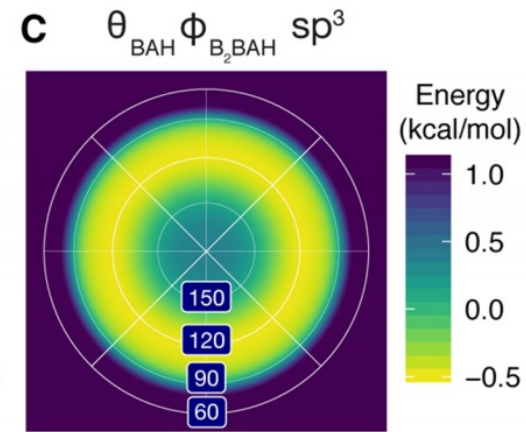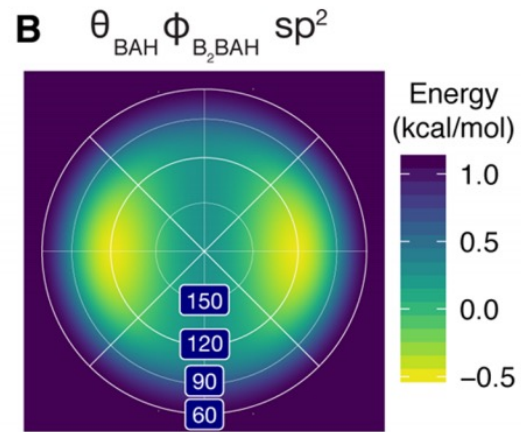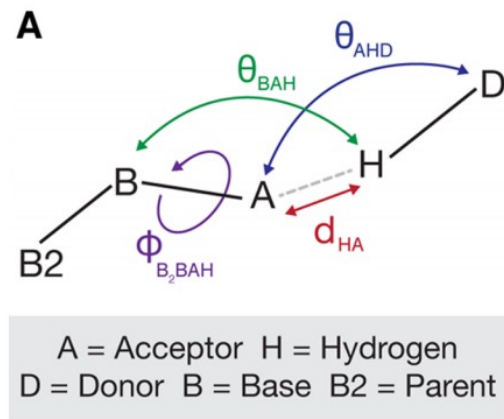
Potential Energy

$$E_{pot} = \underbrace{\sum_b K_2 (b-b_0)^2}_{(1)} + \underbrace{\sum_\theta H_\theta (\theta-\theta_0)^2}_{(2)} + \underbrace{\sum_\phi \frac{V_n}{2} [1 + \cos(n\phi - \phi_0)]}_{(3)}$$

$$+ \underbrace{\sum \varepsilon [(r^*/r)^{12} - 2(r^*/r)^6]}_{(4)} + \underbrace{\sum q_i q_j / \varepsilon_{ij} r_{ij}}_{(5)} + \underbrace{\sum \left[\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}} \right]}_{(6)}$$
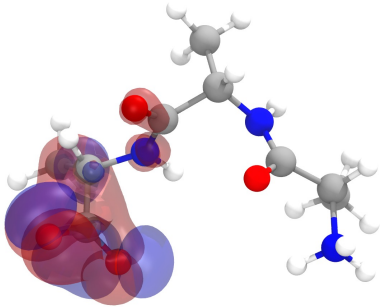
How is this useful?

- Compare **relative energies** of conformers
- Effect of **mutations** on stability
- Refining x-ray structures,
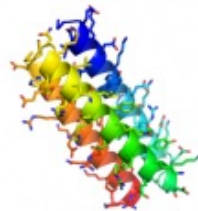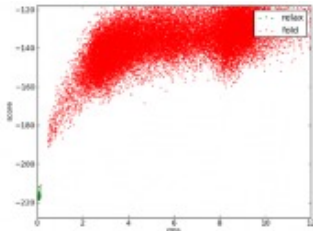 determining structures from NMR data

How are these functions parameterized?
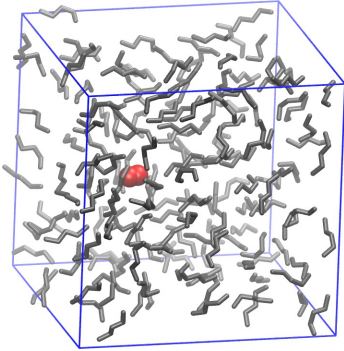
... to match biophysical experiments on small molecules

... to match "higher level theory" simulations on small systems

... to maximize the ability to recapitulate structures/properties from protein crystal structures

Monte Carlo

In molecular simulations, Monte Carlo is an importance sampling technique

    1. Make a random move and produce a new conformation

    2. Calculate the energy change delta *E* for the new conformation

    3. Accept or reject the move based on the *Metropolis criterion*

$$P = \boxed{\exp(-\frac{\Delta E}{kT})} \longrightarrow \quad \textcolor{red}{\text{Boltzmann factor}}$$

If delta *E*<0, then P>1, accept new conformation;
Otherwise:
    if P>rand(0,1), accept,
    else reject.

Simulated Annealing Monte Carlo

In **Simulated Annealing Monte Carlo**, we reduce the temperature as the simulation progresses:
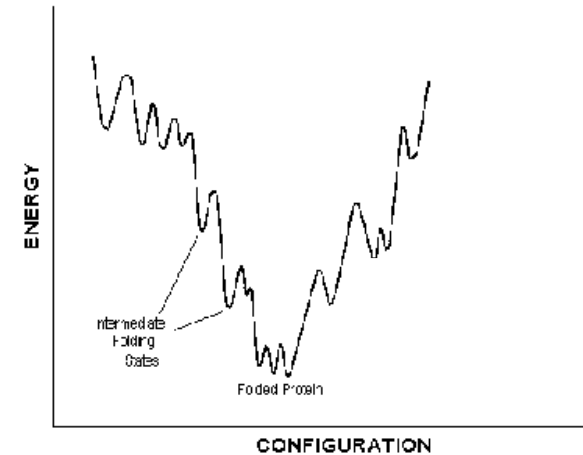
    for i=0:i_{max}

        $T_k=(T_{max}-T_{min}) * (i_{max}-i)/i_{max} + T_{min}$

        Run k steps of Monte Carlo at temperature T_k
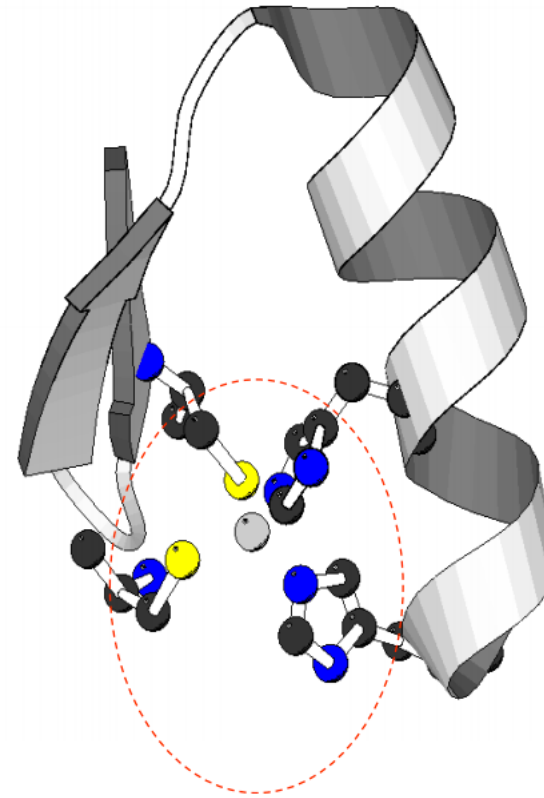
high T: accept almost all structures

low T: accept almost only better structures

Example:
Sidechain rotamer determination

- **Problem**: given the backbone coordinates of a protein, predict the coordinates of the sidechain atoms

- Each sidechain has a discrete number of states ("rotamers")

- Monte Carlo moves:
  - replace sidechain with random rotamer

Molecular Dynamics

Algorithm

- For atom *i,* Newton's equation of motion is given by

$$F_i = m_i a_i \qquad \Longrightarrow \qquad \mathbf{F}_i(t) = m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2}$$

Here, $\mathbf{r}_i$ and m_i represent the position and mass of atom *i* and $\mathbf{F}_i(t)$ is the force on atom *i* at time *t*. $\mathbf{F}_i(t)$ can also be expressed as the gradient of the potential energy

$$\mathbf{F}_i = -\nabla_i V \qquad \Longrightarrow \qquad -\nabla_i V = m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2}$$

V is potential energy. Newton's equation of motion can then relate the derivative of the potential energy to the changes in position as a function of time.

Molecular Dynamics

Numeric integration by using the **Verlet algorithm**

- Given initial velocity 0 and position x_i, numerically integrate to get position at time $t+\delta t$
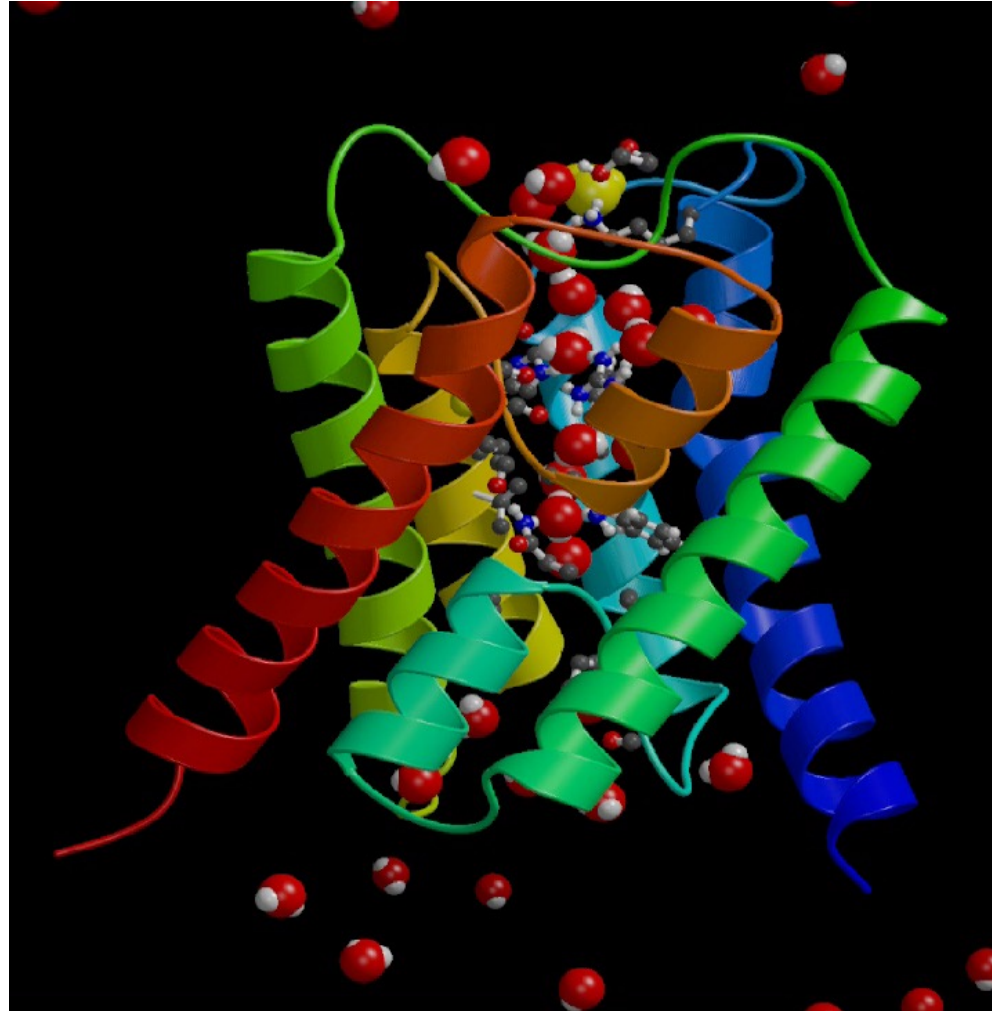- Taylor expansions to 3rd order for i

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + (\delta t)\mathbf{v}(t) + \frac{1}{2}(\delta t)^2 \mathbf{a}(t) + \frac{1}{6}(\delta t)^3 \mathbf{b}(t) + ...$$

$$\mathbf{r}(t - \delta t) = \mathbf{r}(t) - (\delta t)\mathbf{v}(t) + \frac{1}{2}(\delta t)^2 \mathbf{a}(t) - \frac{1}{6}(\delta t)^3 \mathbf{b}(t) + ...$$

- Adding these equations gives [up to order $(\delta t)^4$]:

$$\mathbf{r}(t + \delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + (\delta t)^2 \mathbf{a}(t) + O[(\delta t)^4]$$
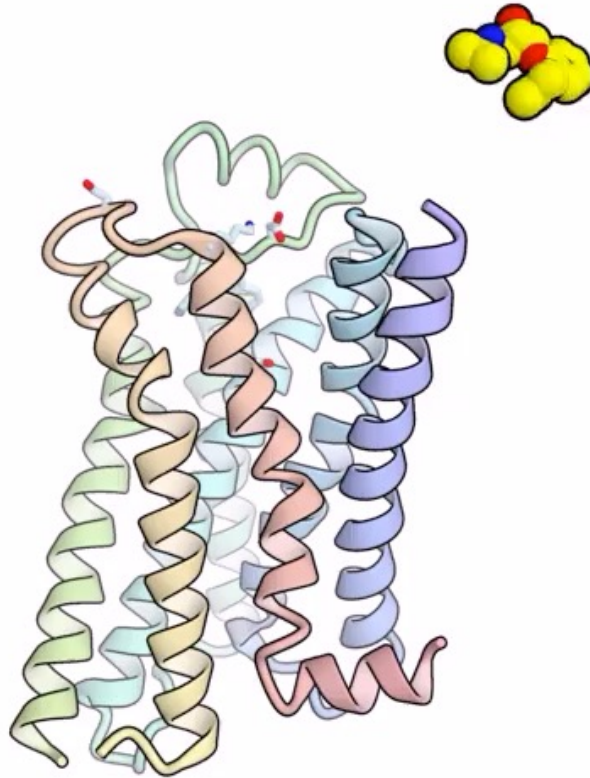
Aquaporin-1

(B.L. de Groot and H. Grubmüller: Science 294, 2353-2357 (2001))

Drug binding to GPCRs

0.00 us

Dror et al., PNAS 2011

Dror *et al. PNAS 2011*