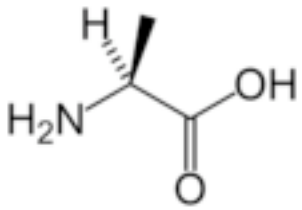
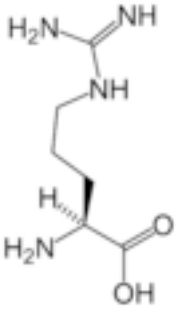
Protein folding… linear molecule to functional object

- DNA transcribes a sequence of amino acids in a protein
- Generally a protein is comprised of 20 amino acids
- In solution, at finite temperature, protein folds
- Resulting structure determines biological function
- Complex, fundamental problem still unresolved!

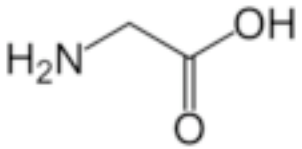
Examples… admittedly from Wikipedia…



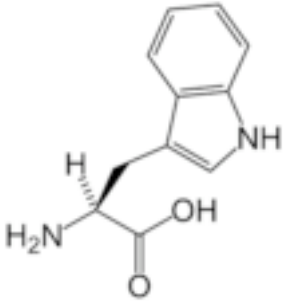
alanine



arginine



glycine



tryptophan

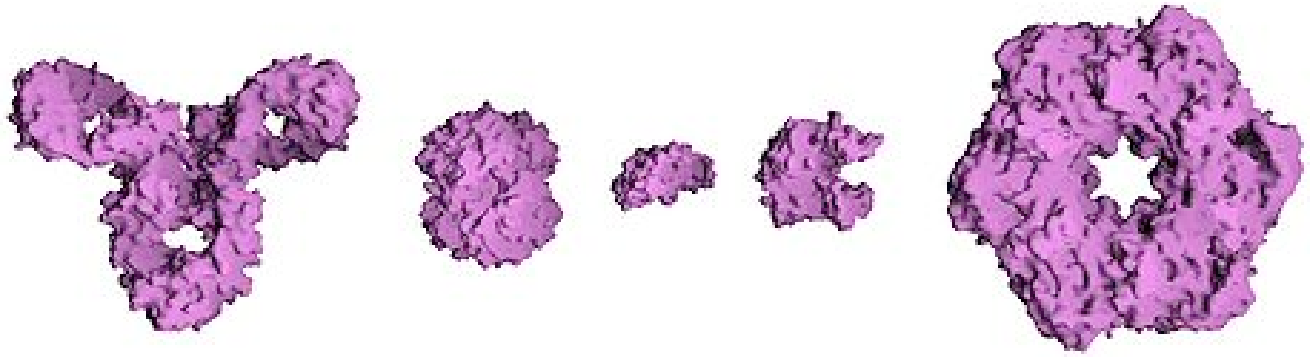
Etc.

Structural characteristics

- Primary structure - amino acid sequence
- Secondary structure-

  Repeating local structure stabilized by hydrogen bonds

  e.g. alpha helix (characterized by L. Pauling)
- Tertiary structure- overall shape



Immunoglobulin hemoglobin insulin adenylate kinase glutamine synthetase

- Makes up our enzymes, antibodies, and hormones

What determines tertiary structure?

- Set by amino acid sequence
- Depends on interactions and temperature
- Hydrophobic residues like to sit in center of folded protein
- Other interactions including hydrogen bonds important

Extremely simple approach using a toy model, and Hamiltonian…

$$H = \sum_{<m,n>} J_{A(m),A(n)}$$

Summation over amino acids that are nearest neighbors and not already bonded with a covalent interaction

$J_{A,B}$ is a 20x20 matrix of interactions

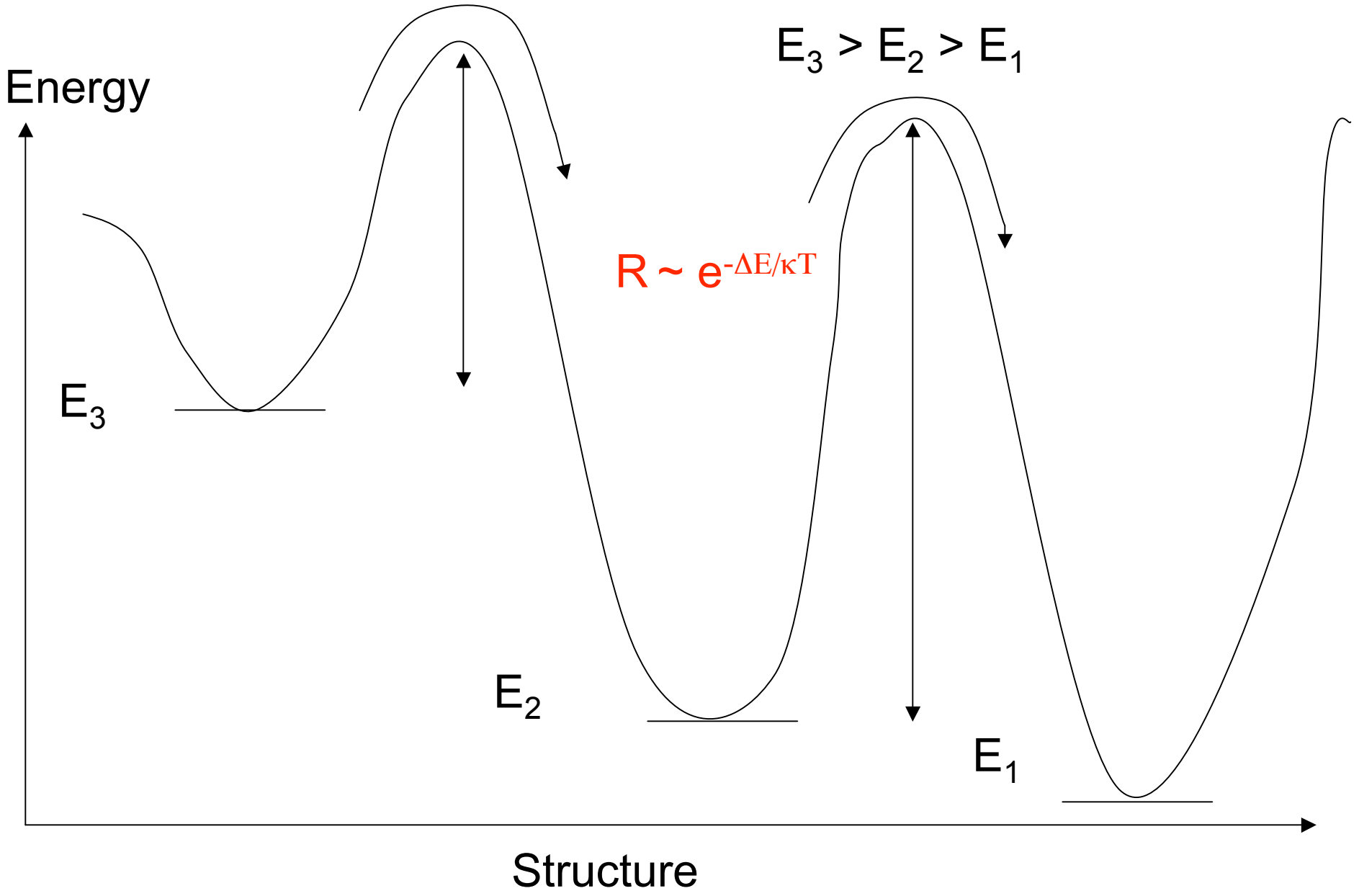
How does protein fold rapidly and reproducibly?

- Extremely large number of possible structures (~4^N)
- Many not greatly different in energy compared to kT
- Estimates for time ~ age of the universe!
- Experimentally, takes a few seconds…
- Reproducibility key to life… if our hemoglobin frequently folded incorrectly… we'd probably suffocate
- Temperature, salt concentration, etc. can cause errors
- Basic physics clear, process complex!

$$P \sim e^{-E/kT}$$

- This suggests we try Metropolis MC procedure on our model protein (rather a toy model…)

The origin of the problem… metastable states



Transitions between states

- Transition states key to the evolution
- Probability of being in a transition state $P \sim e^{-\Delta E/kT}$
- If $\Delta E >> kT$, transitions happen slowly
- If $\Delta E \sim kT$, transitions fast, system explores wide range of space
- Many local minima may exist to trap system
- Simulated annealing is an appropriate approach

Simulated annealing

- Probability of accepting uphill step ~ $e^{-\Delta E/kT}$
- High T (compared to typical ΔE) accepts many uphill moves
- Low T, uphill moves usually rejected

Only way to approach ground state is by starting at high T and gradually lowering the temperature

Even with simulated annealing, and long simulation times, it is possible to not find the ground state

Other complications with protein folding… Too big, slow, and complex!

- Ideally would like a quantum model of interactions
- Costly, ~100-10,000 atoms in protein
- In principle need water molecules, solvated ions, etc.
- Dynamics not on a grid (like our project)

A typical approach is to use empirical interaction models…
Still extremely difficult but tractable to at least simulate

Biggest problem is ~ns simulation time scale compared to seconds or minutes in real system

Monte Carlo can partly alleviate timescale challenges…

Project in protein folding

- Use a chain 40 amino acids in length
- Randomly choose each amino acid from 20 possible kinds
- Determine a random matrix $J_{i,j}$ for interactions, between -1, +1
- Perform a Monte Carlo simulation with simulated annealing
- Determine the energy and end-to-end length as a function time
- Try for different annealing schedules, simulation "times" to explore metastable, local minima states
- Consider how large your temperature should be… what is the appropriate energy scale?
- Determine average energy/end to end length as a function of T
- How do we know equilibrium is achieved? (not an easy question
- How might we get the heat capacity ?(it must release some ener as T is lowered)