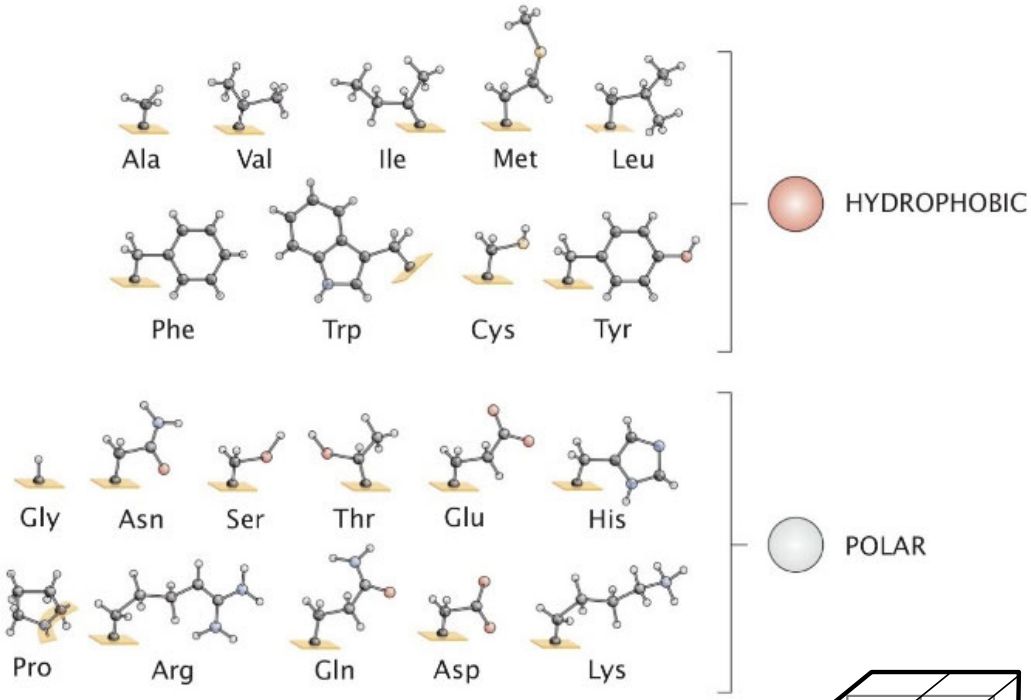
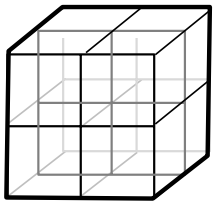
HP-models of protein folding



Adapted from Physical Biology of the Cell (Chapter 8)



categorization of amino acids as **hydrophobic (H)** or **polar (P)**

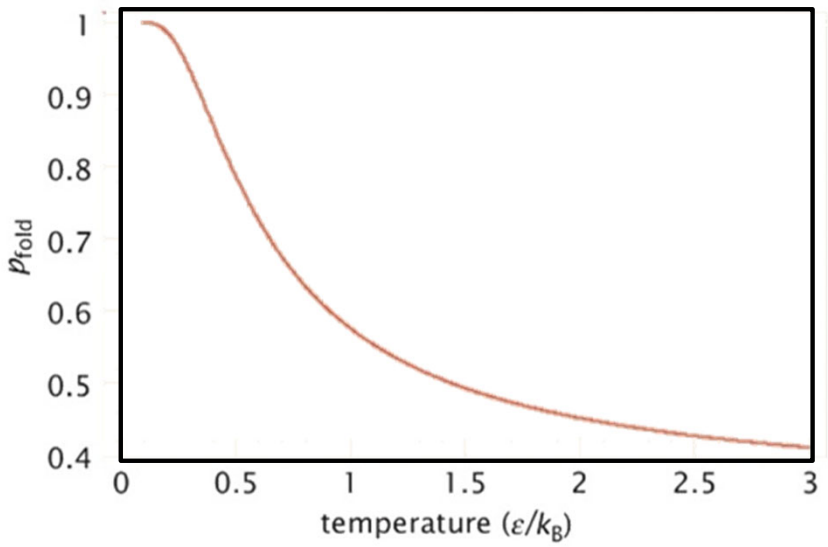
→ **2 letter alphabet allows an abstract model of structure formation**

Of note: finer categorization possible -> charged, acidic, basic, helix forming, etc.

Quiz: How many HP sequences possible on a 3x3x3 lattice?
How many different HP protein configurations are possible?
Note: there possible 103'346 compact structures.

6x2 lattice model

Temperature dependence of folding



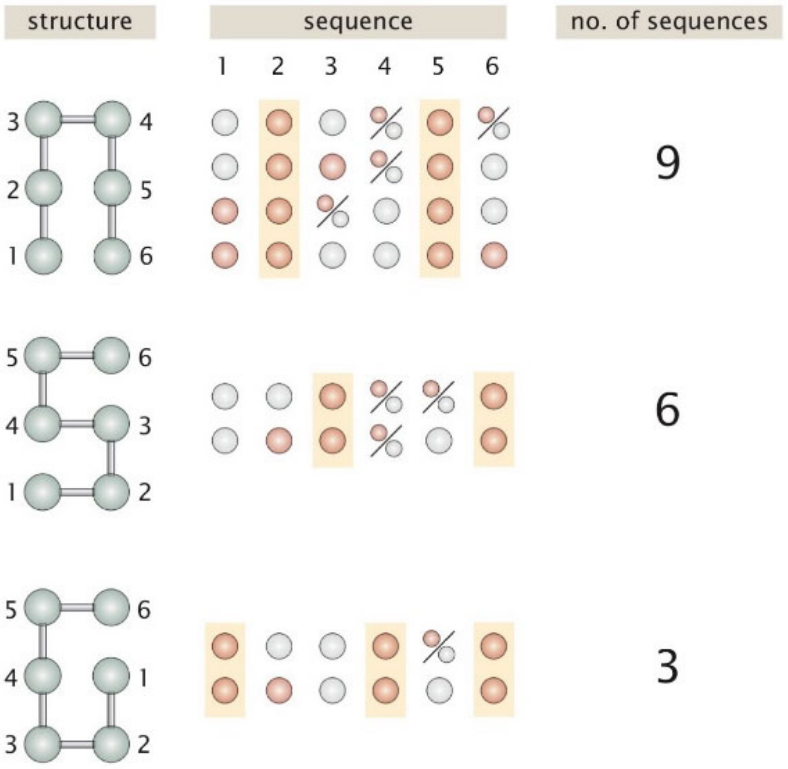
$$p_{fold} = \frac{e^{-2\beta\varepsilon}}{e^{-2\beta\varepsilon} + 2e^{-4\beta\varepsilon}} = \frac{e^{-2\beta\varepsilon}}{Q}$$

sigmoidal transition

reminiscent of real protein denaturation

Adapted from Physical Biology of the Cell (Chapter 8)

Identification of protein-like (foldable) structures



enumerate all 64 possible protein configurations

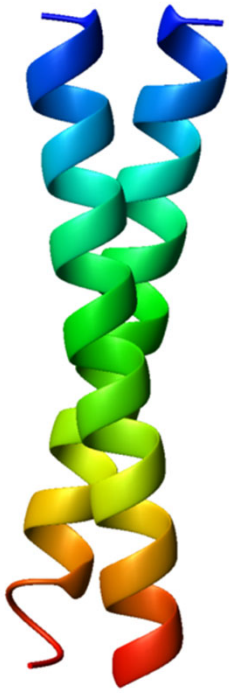
→ determine energy of each

necessary pre-requisite for a stable fold: at least one HH contact

The "hairpin" is the most designable structure

→ simple models like this: **Protein design principles**

Real structures: Simple tertiary structure motifs



GCN4 coiled coil domain
PDB: 1zik

Secondary structure elements form the **elementary motifs** of protein structure.

Quiz:

The sequence of GCN4 is given as follows:
RMKQLEDKVEELLSKNYHLENEVARLKKLVGER

how can it be immediately recognized that this folds into a coiled coil?

Helix bundles: Inherently designable structures

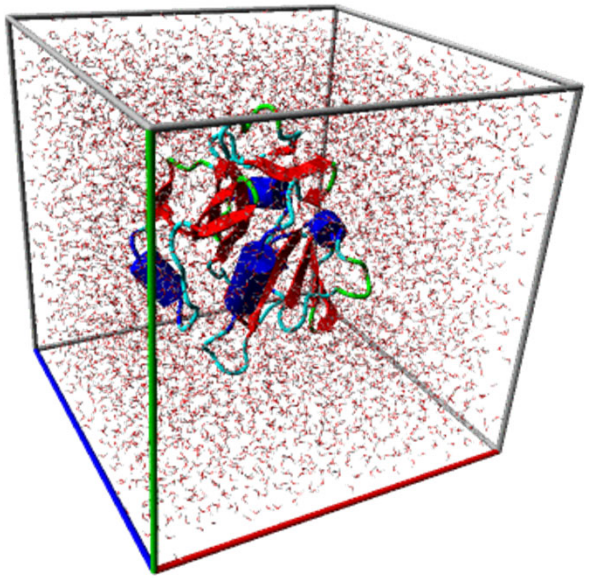
Hecht et al. Protein Sci 2004

Secondary structure
→ fixed, helical

Tertiary structure:
follows HP rules

Provides a platform, on which further functionality can be designed

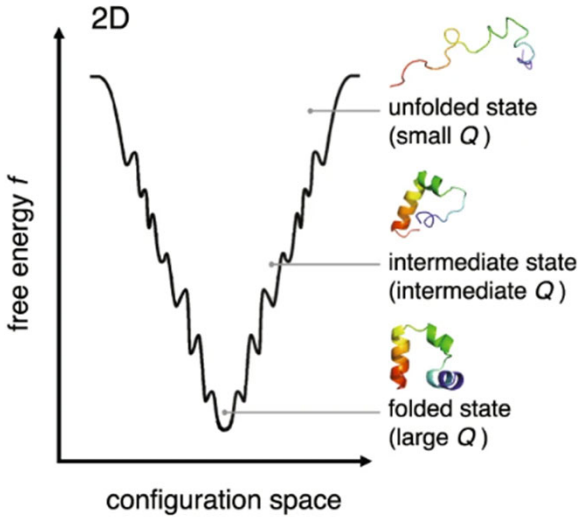
Force field for protein structure



https://in.explara.com/

Force field for protein structure evaluation:

$$V = \sum_{bonds} a_\alpha \left(x_i - x_{i0}\right)^2 + \sum_{bond\ angles} b\beta\left(\theta_i - \theta_{i0}\right)^2 + \sum_{charges} \frac{Z_i Z_j e^2}{D(r) r_{ij}} + \sum_{neutral\ atoms} 4d_\delta \left[\left(\frac{Ai_j}{ri_j}\right)12 - \left(\frac{B_{ij}}{r_{ij}}\right)\right]$$

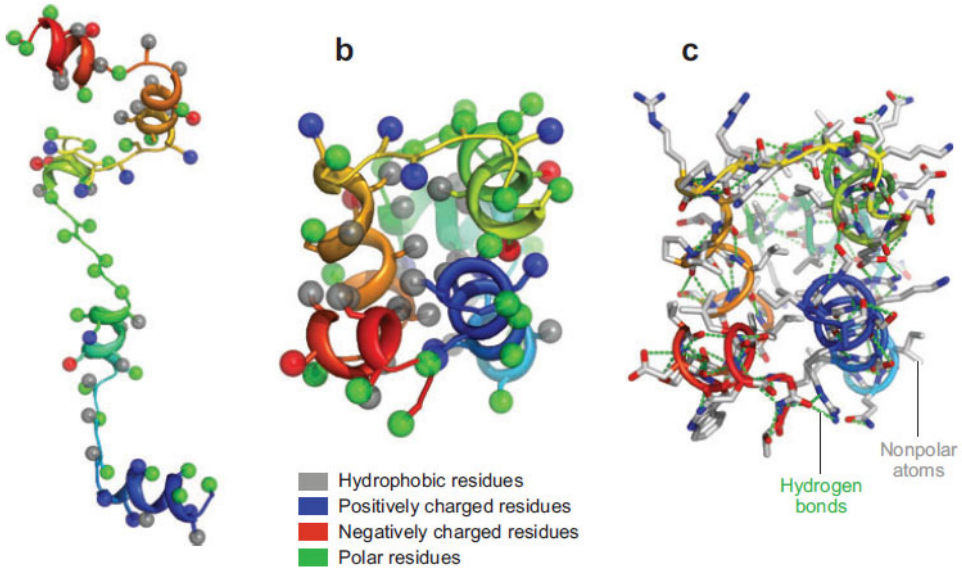


https://www.nature.com/articles/s41598-019-50825-6

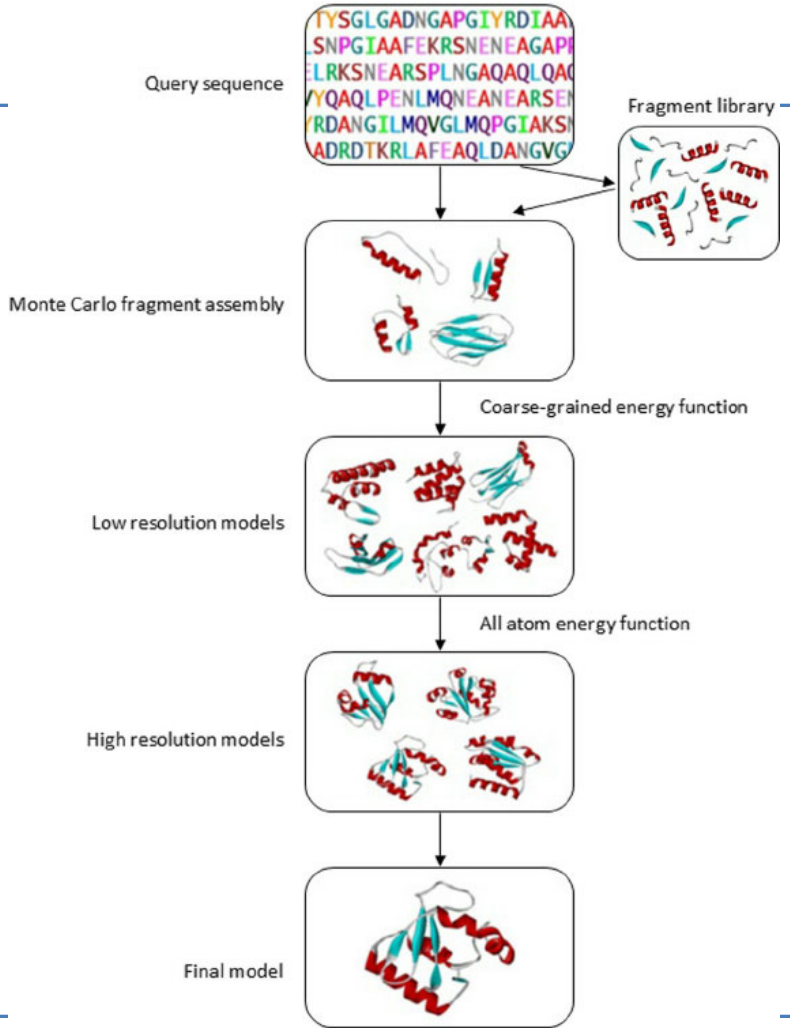
Goal: Minimize the configurational energy, find the global minimum

Protein structure prediction / design

Design from fragment libraries,
followed by refinement (Rosetta)

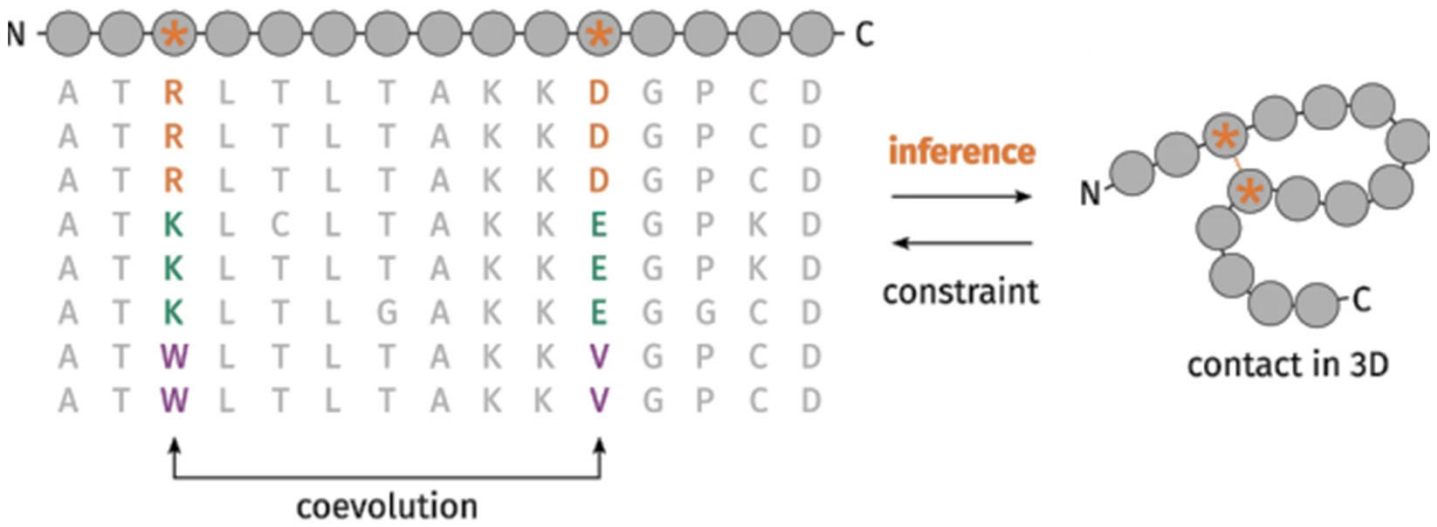


Das & Baker, Annu Rev Biochem 2008



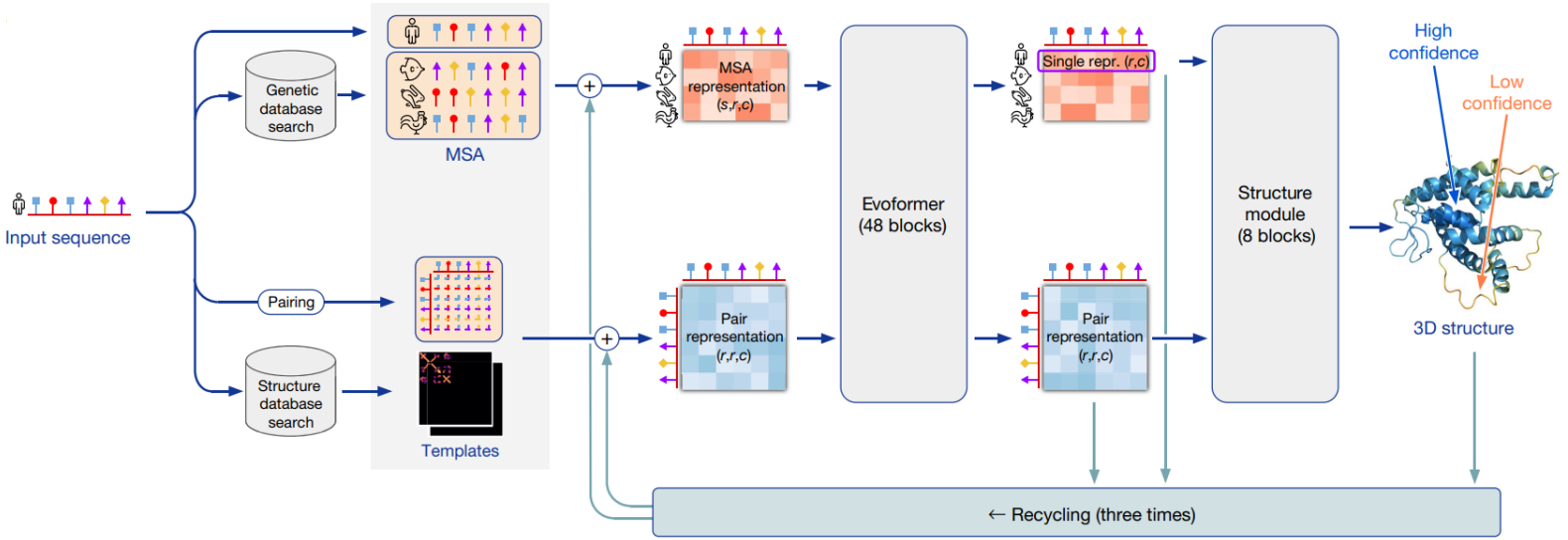
https://tbiomed.biomedcentral.com/articles/10.1186/s12976-015-0014-1

Structural information from sequence: Co-evolution analysis



43146:42mrxuqdd1srqh1335;:99

Protein structure prediction – ML approaches



Input:

- primary sequence of target protein
- multiple sequence alignments of homologues
- structural information of homologues (PDB)

Process 1:

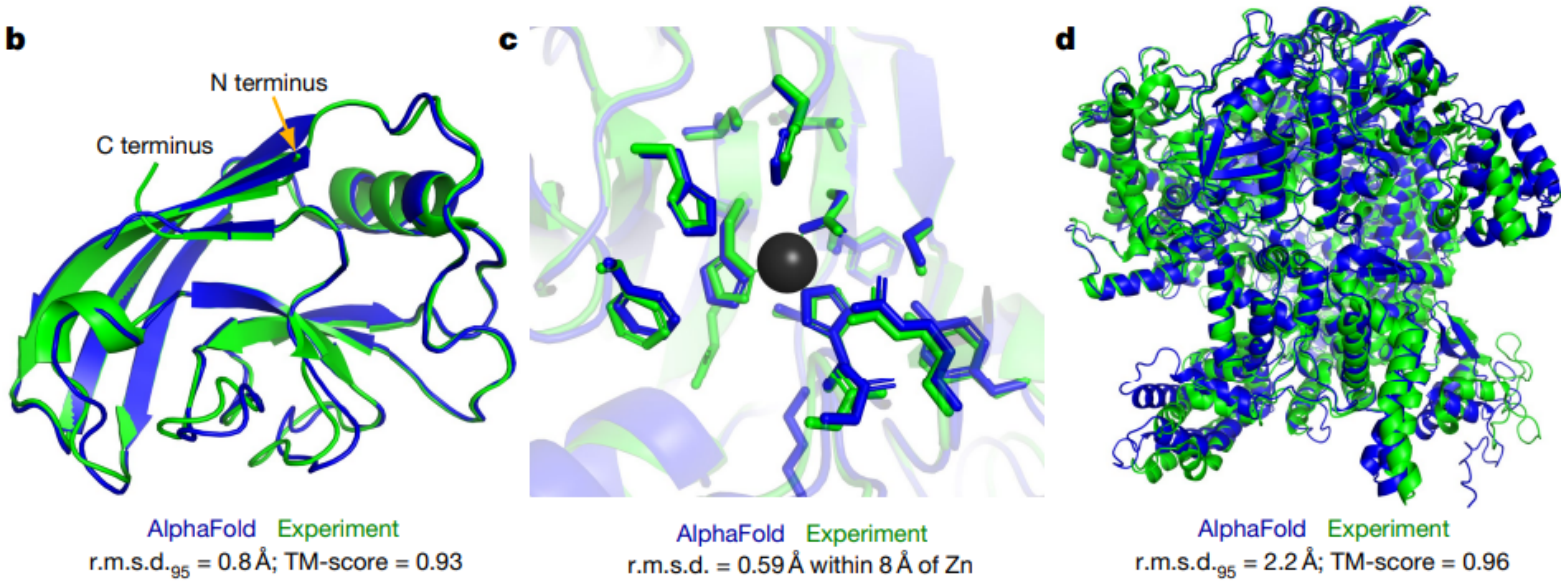
- MSA representation
- Residue pair representation

Process 2:

- 3D structure representation: rotation & translation of each residue
- refinement

AlphaFold2 - Nature 2021

Protein structure prediction – ML approaches

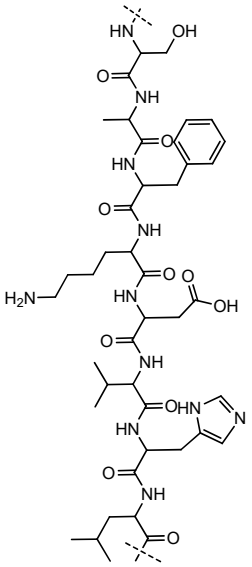


Result:

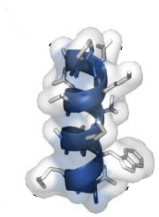
- highly accurate structure prediction
- low backbone RMSD
- correct residue orientation

AlphaFold - Nature 2021

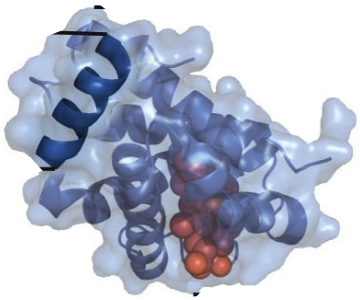
Protein structure – A thermodynamic view



Amino acid chain
primary structure



α-helix
secondary
structure



Protein domain
Tertiary structure

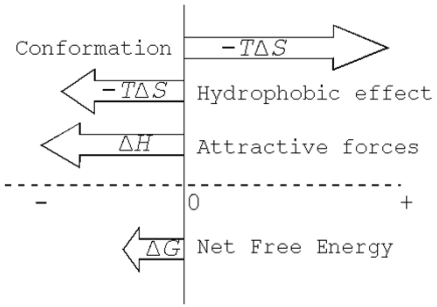
Protein complex
Quarternary structrue

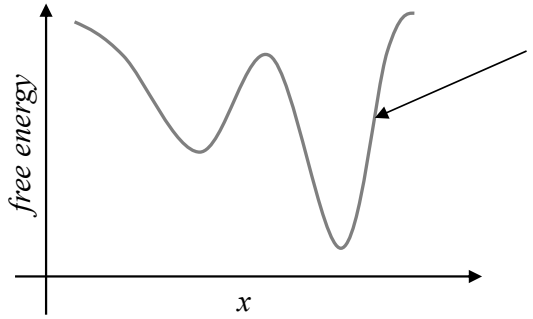
transitions

A free energy surface to understand protein structure

using a basic chemistry principles, **the free energy surface**, to describe state transitions in proteins

Free energy:





each (populated) protein conformation falls somewhere on this line

some reaction coordinate:

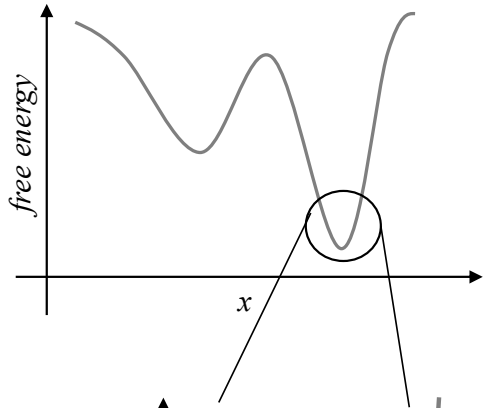
solvent accessibility
compactness
% hydrogen bonds saturated

Global states in proteins

For a biophysical understanding → simplifications are required.

Global state
Coarse grained

can be
characterized

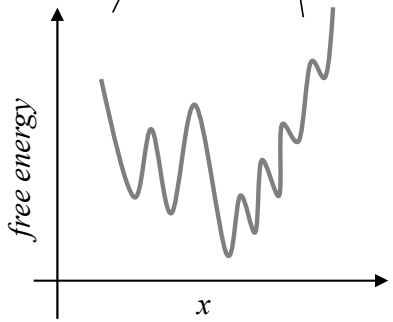


ensemble

***i* Microstates**
Atom positions
Dihedral angles
Bond vibrations
...

interchanging
rapidly

characterization
difficult



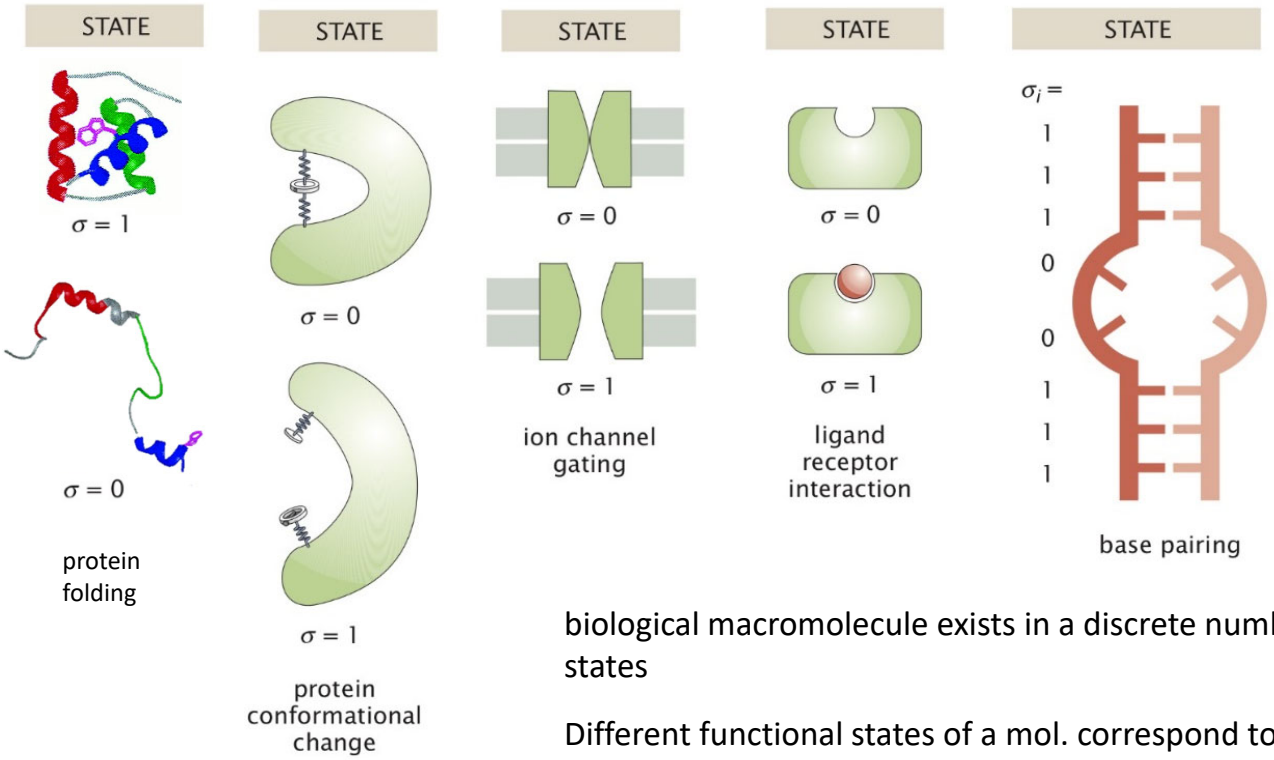
Global states in proteins

Biology: Global states are useful concept

- We can assume that any biological macromolecule exists in a discrete number of global states, dependent e.g. on the enviroment
- Different functional states of a protein correspond to different global states
- Changes in global states could be related to function

As discussed earlier, complex systems such as proteins can be treated with **statistical thermodynamics**: This allows to extract useful parameters.

Motivation – biomolecules exist in multiple states

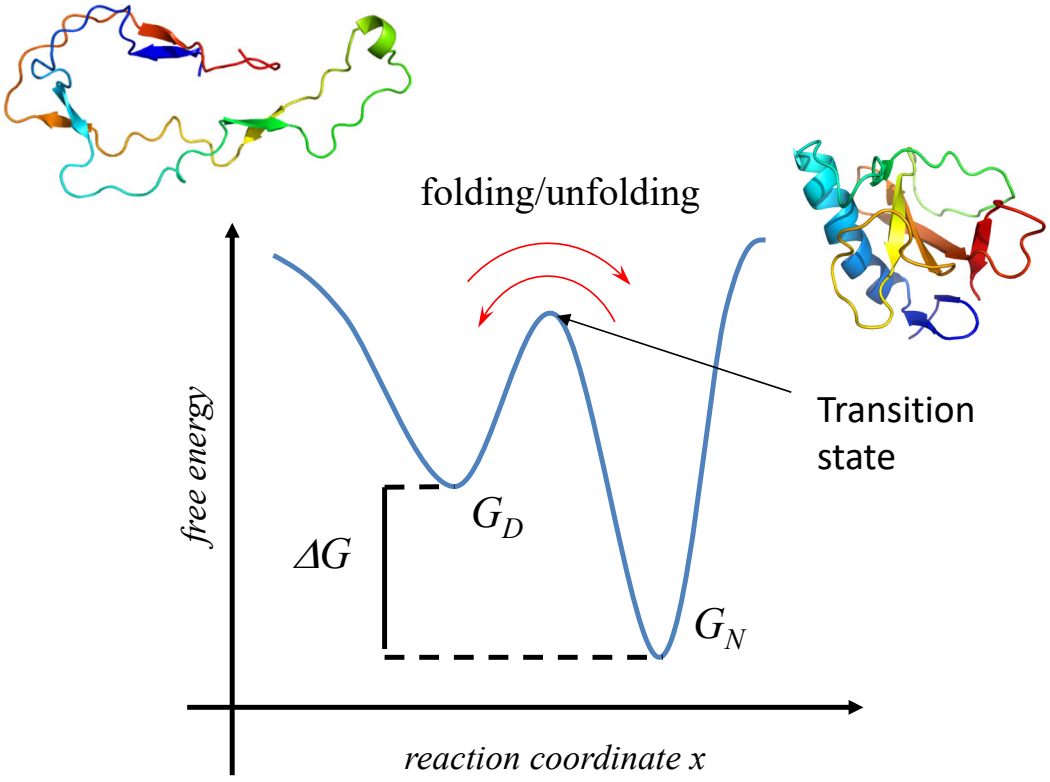


biological macromolecule exists in a discrete number of global states

Different functional states of a mol. correspond to different global states

Changes in global states could be related to function

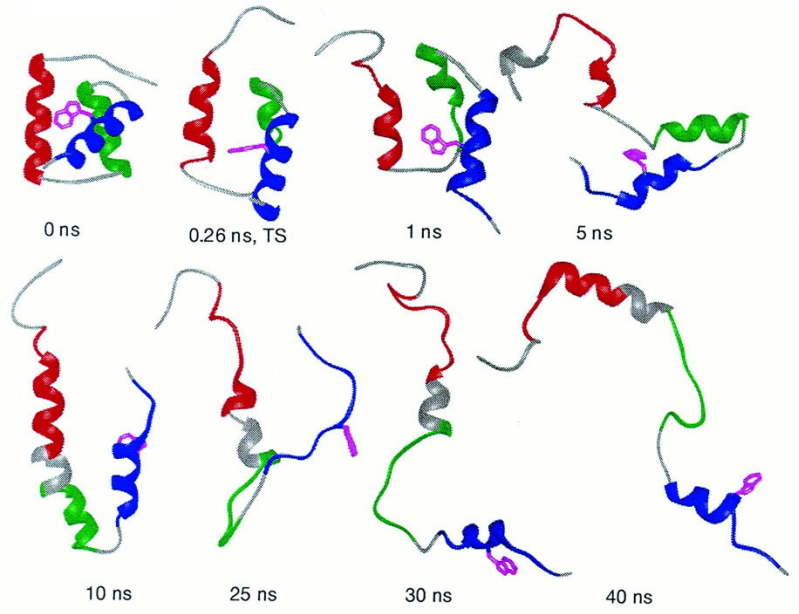
Global states in a protein: N and D



Quiz:

What could be a measure for the reaction coordinate x?
i.e. what could x represent?

Protein denaturation process



Mayor & Fersht, PNAS 2000

can be reversible:

- Loss of defined tertiary structure
- Partial loss / change in secondary structure
- exposure of buried hydrophobic amino acids

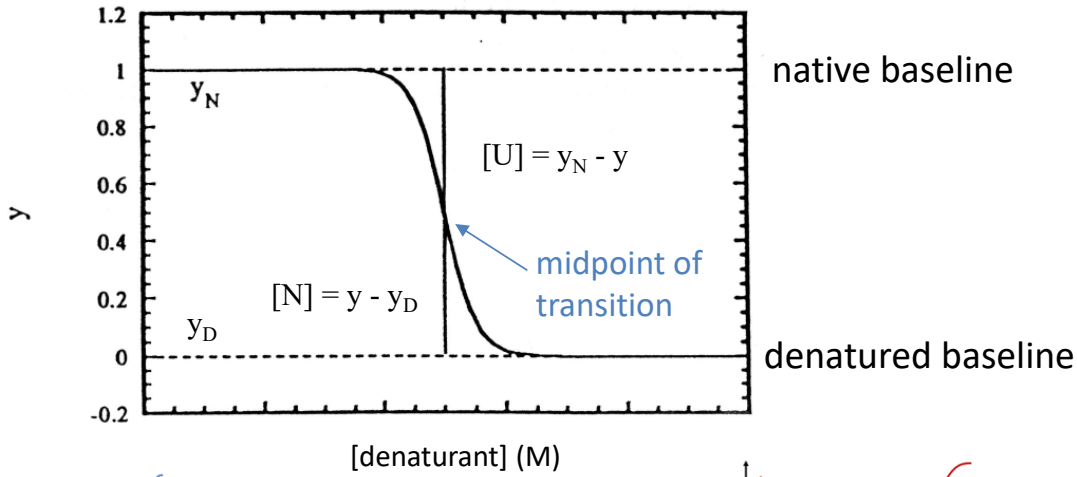
or irreversible

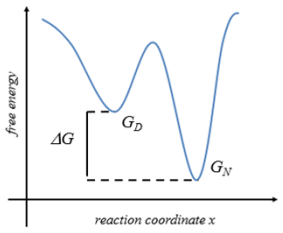
- at high protein concentration, aggregation
- can make sick (prions)

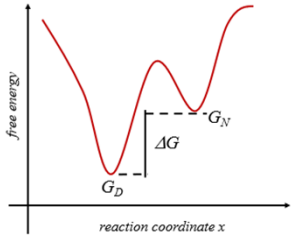
Denaturation of globular, monomeric protein (2-states)

Isolated protein in solution

We monitor one variable (e.g. **tyrosine fluorescence**) and vary another variable (e.g. **temperature** or a **chemical denaturant**)







Protein denaturation equilibrium:

$$N \overset{K}{\rightleftarrows} D$$

Measurement variable y:

$$y = [\mathrm{N}]\,y_N + [D]\,y_D$$

Determination of K throughout the transition:

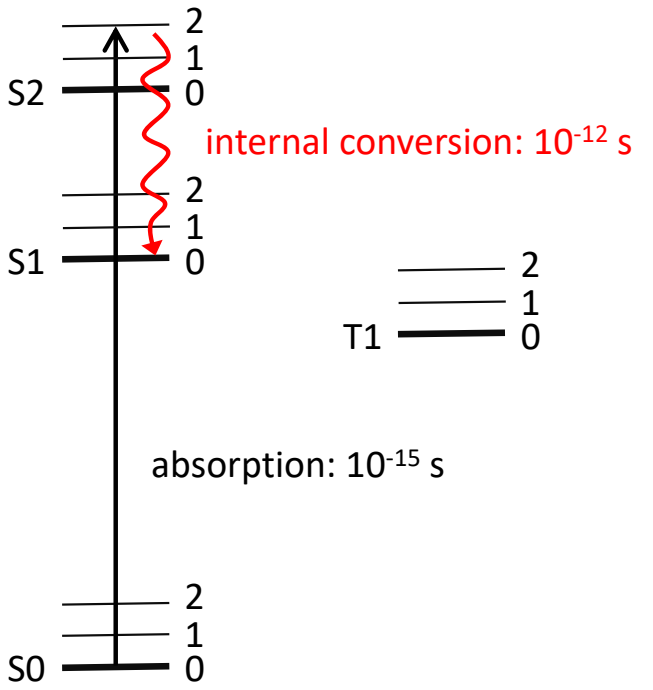
$$K = \frac{[\mathrm{D}]}{[N]} = \frac{y_N - y}{y - y_D}$$

How can we observe two-state transitions?

Any methods that can distinguish between U and N

- **Absorbance** (e.g. Trp, Tyr) due to change in the micro-environment
- **Fluorescence** (Trp)-difference in emission spectrum & intensity, due to change in microenvironment
- **Circular dichroism** (far or near UV), due to change in asymmetric environment of fluorophores
- **Calorimetry** (DSC), due to change in heat capacity and heat absorption
- **NMR spectroscopy**
- **Gel electrophoresis** or **size exclusion chromatography**
- **Catalytic activity**, functional assays
- **External probes** (chromophores, fluorophores)

Lifetimes of excited states and emission



Absorption

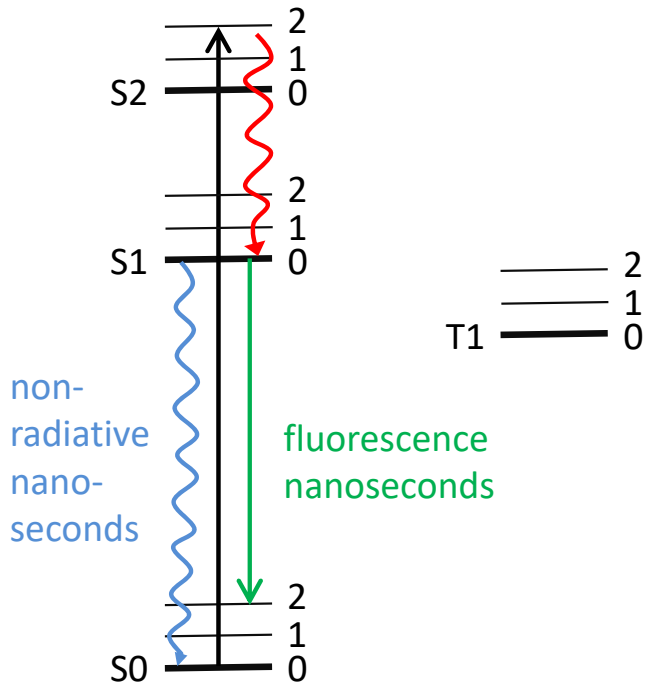
time: 10^{-15} s, corresponds to the time it takes for light wave to pass molecule

Internal conversion

Molecule returns to lowest singlet state, lowest vibrational niveau

time: 10^{-12} s

Lifetimes of excited states and emission



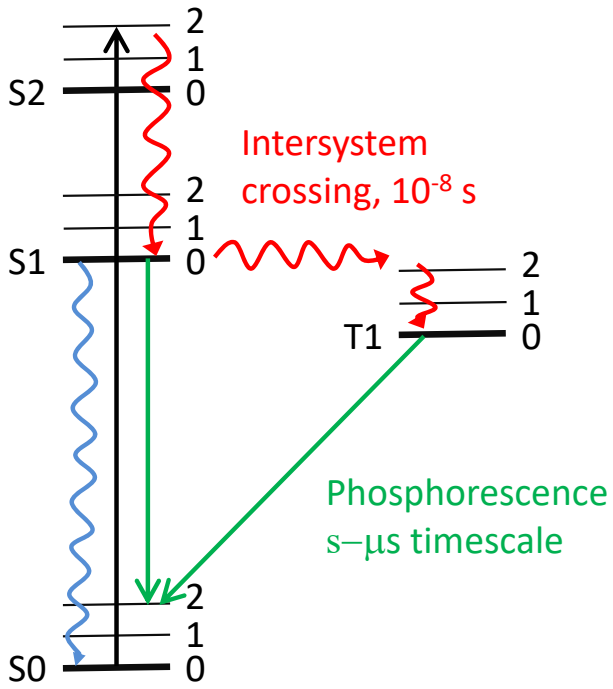
Fluorescence

time: 10^{-8} - 10^{-10} s

Nonradiative, internal conversion

time: 10^{-8} - 10^{-10} s

Lifetimes of excited states and emission



Intersystem crossing

time: 10^{-8} s

inversion of the electron spin

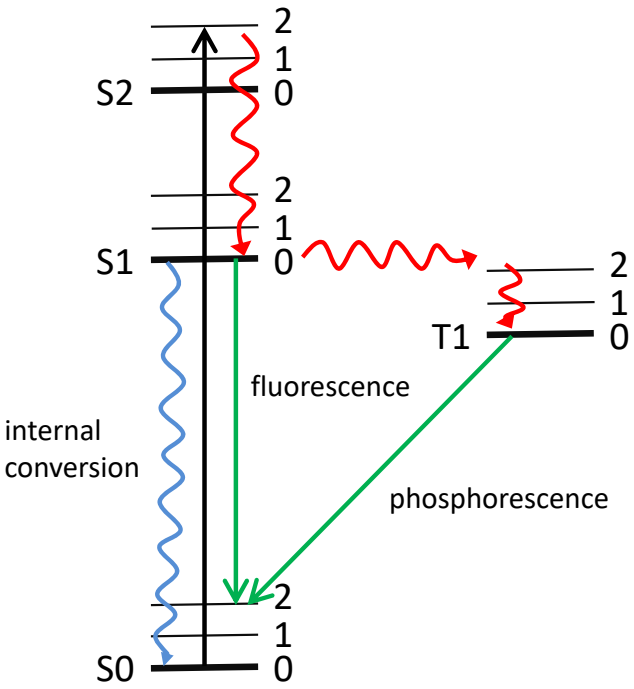
Phosphorescence

time: 10^{2} - 10^{-4} s

requires again inversion of e- spin

forbidden transition, thus long timescale

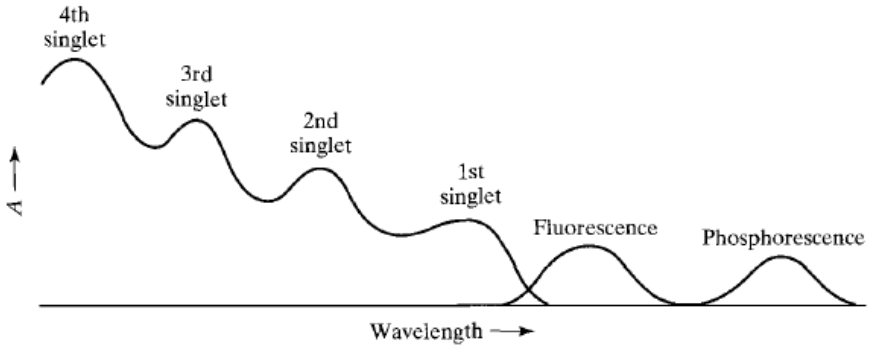
Absorption and emission spectra



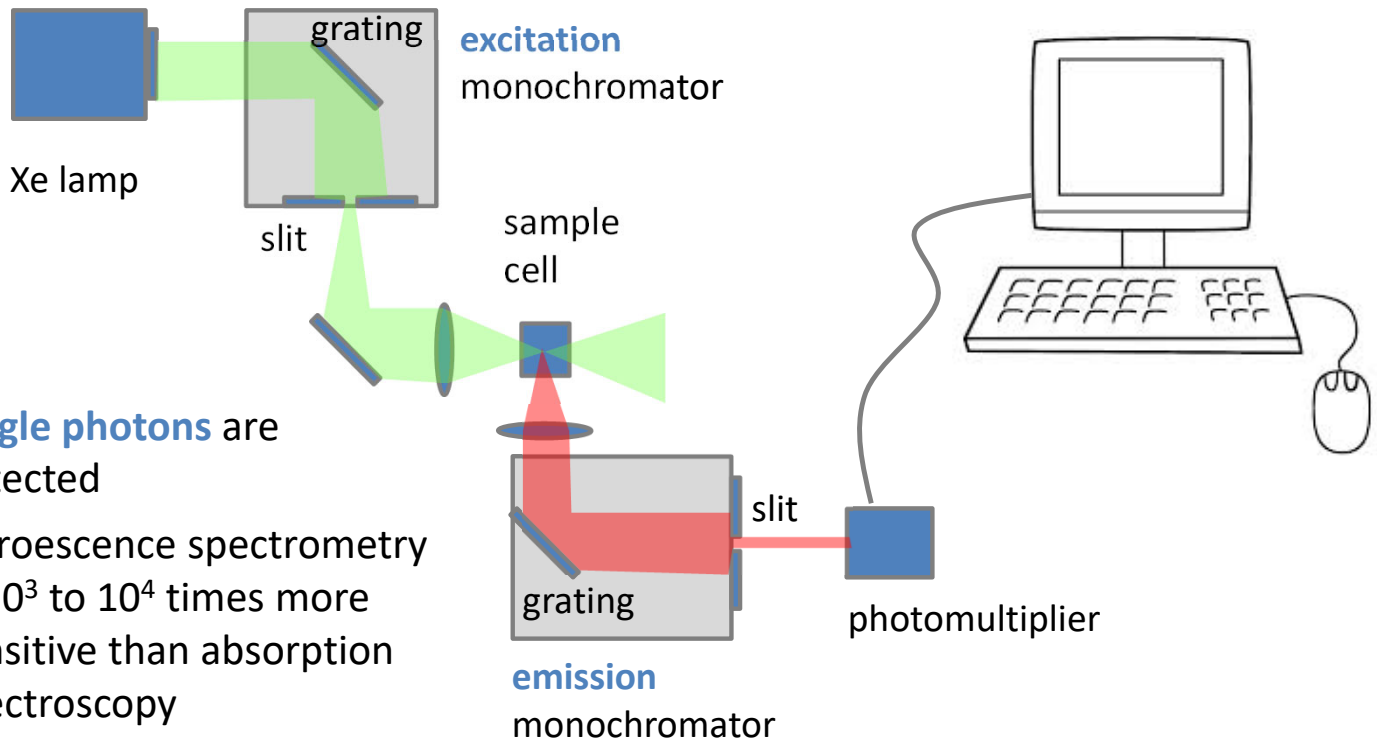
Spectra:

Fluorescence is red-shifted to the absorption spectrum

phophorescence is at even higher wavelengths



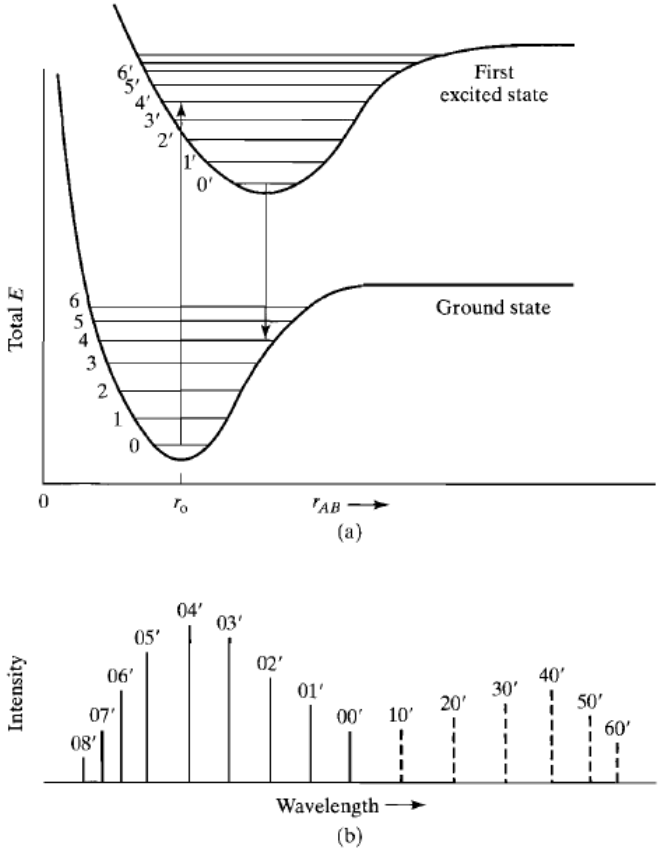
Measuring fluorescence



Single photons are detected

Fluroescence spectrometry is 10^3 to 10^4 times more sensitive than absorption spectroscopy

Fluorescence spectrum



Shape of **fluroescence spectra**:

The arrows denote the most probable transitions from

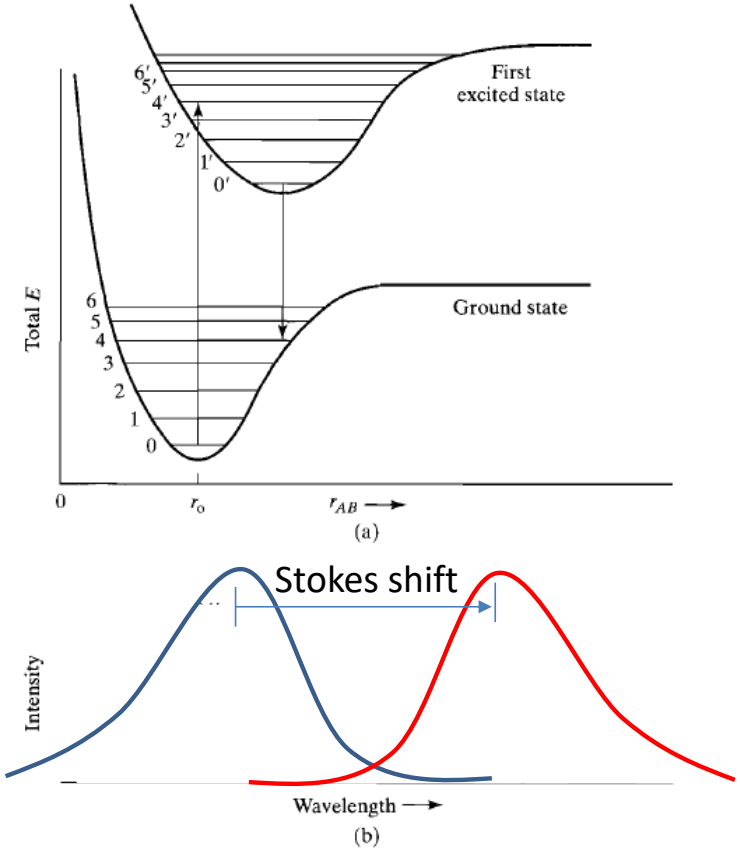
ground state -> s1

s1 -> ground state

The fluorescence spectrum is a **mirror image** of the absorption spectrum

the different in absorbance and fluorescence maximum is the **Stokes shift**

Fluorescence spectrum



Shape of **fluroescence spectra**:

The arrows denote the most probable transitions from

ground state -> s1

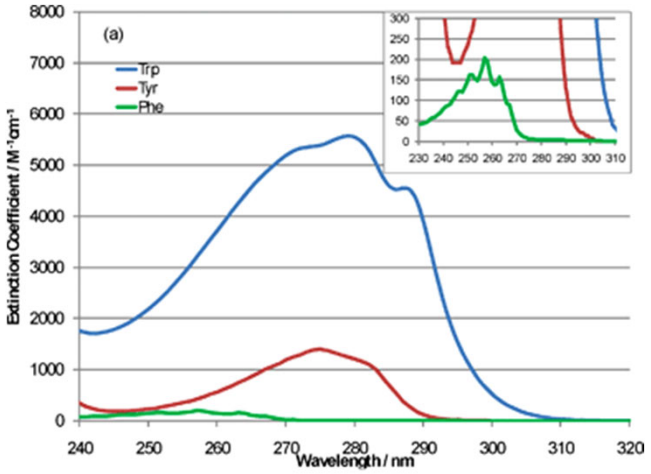
s1 -> ground state

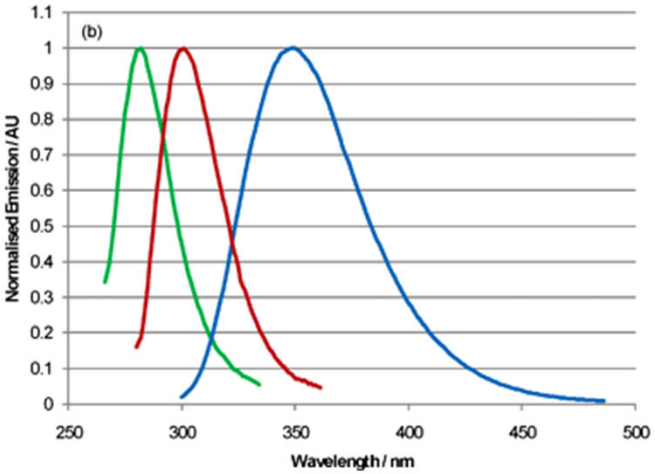
The fluorescence spectrum is a **mirror image** of the absorption spectrum

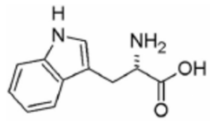
the different in absorbance and fluorescence maximum is the **Stokes shift**

Amino acids: Fluroescence spectra

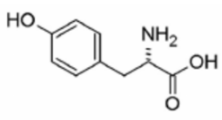
Absorbance and fluroescence spectra of aromatic amino acids



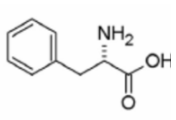




Trp



Tyr

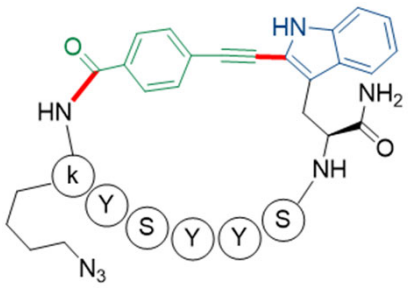


Phe

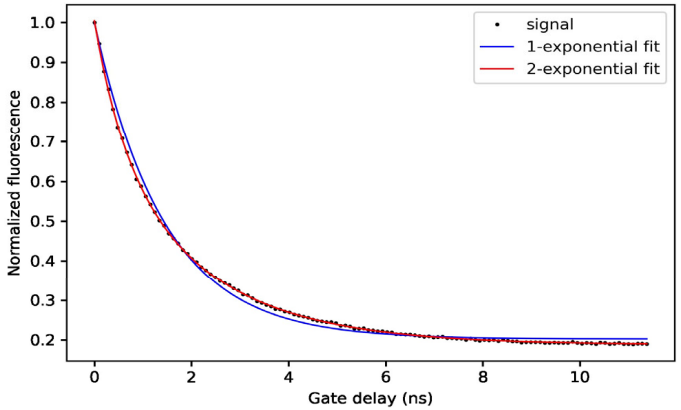
Quantum yield for tryptophan: $Q_F=0.2$

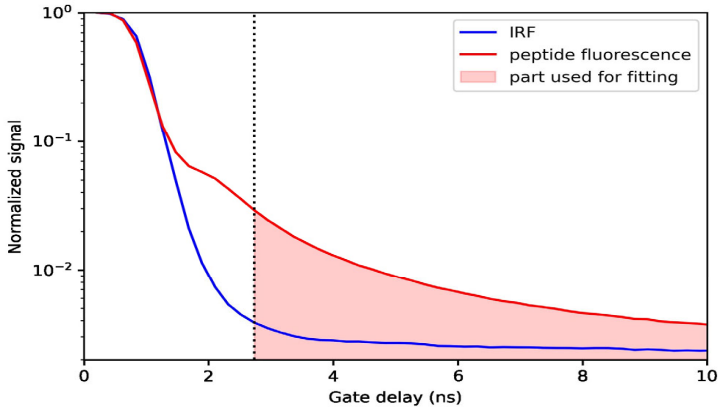
→ Quantum yield

Fluorescence lifetime



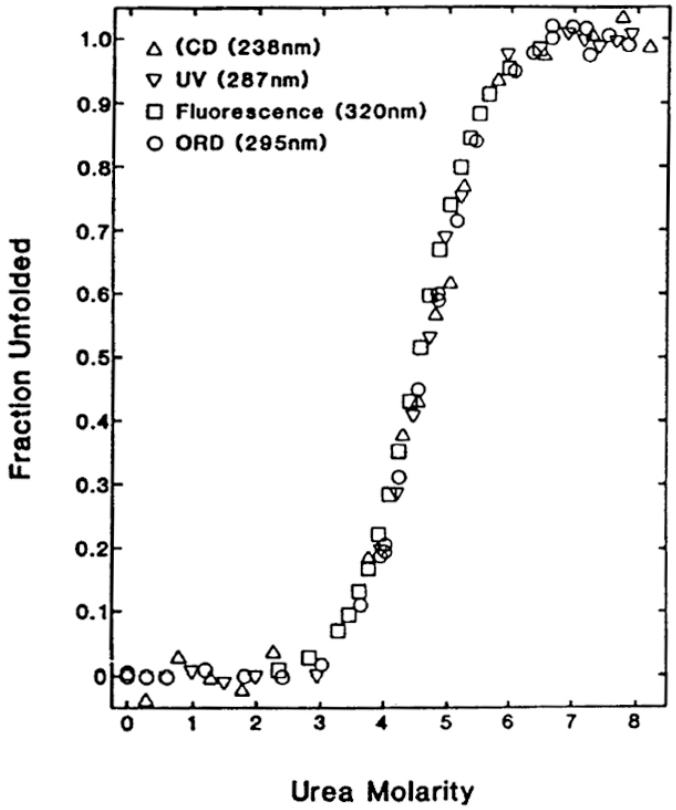
kYSYYSW-NH$_2$
2aa, 53%, 95:5





Xing-Yu Liu, Nathan Ronceray

Denaturation of globular, monomeric protein : RNase T1



$$N \rightleftarrows D$$

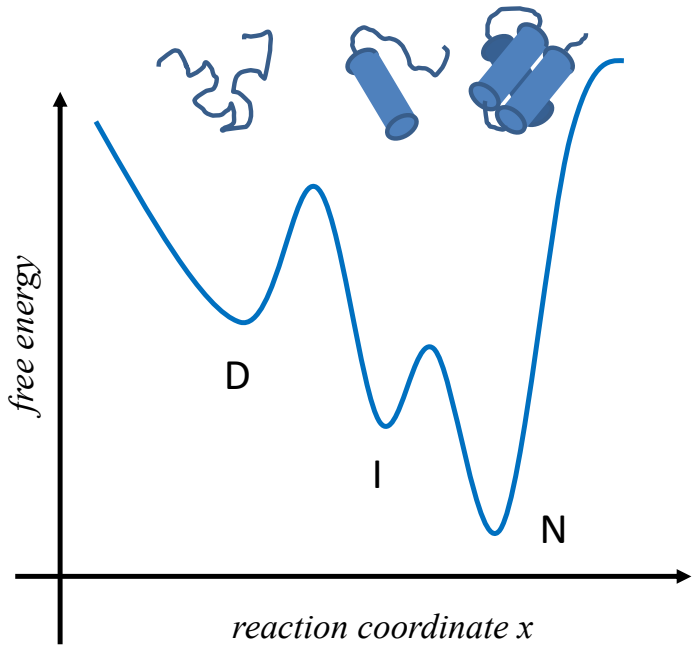
$$y = [\mathrm{N}]\,\mathrm{y_N} + [D]\,\mathrm{y}_D$$

$$K_D = \frac{[\mathrm{D}]}{[N]} = \frac{\mathrm{y}_N - y}{y - y_D}$$

different parameters completely overlap

normalized transitions ($\theta_N = [N] / ([N] + [D])$) overlap

Three (or more) state equilibrium transitions

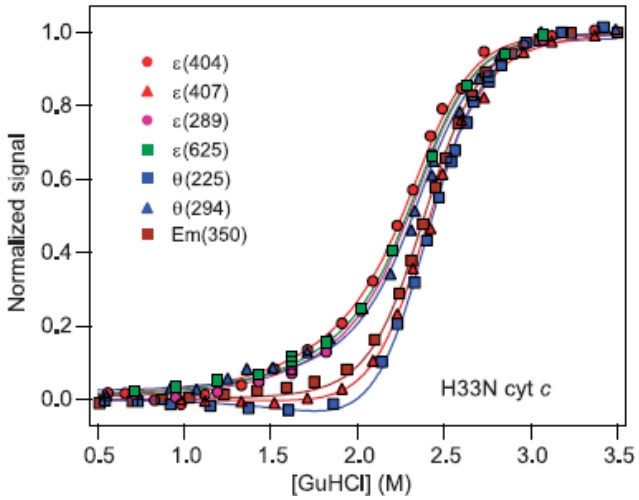


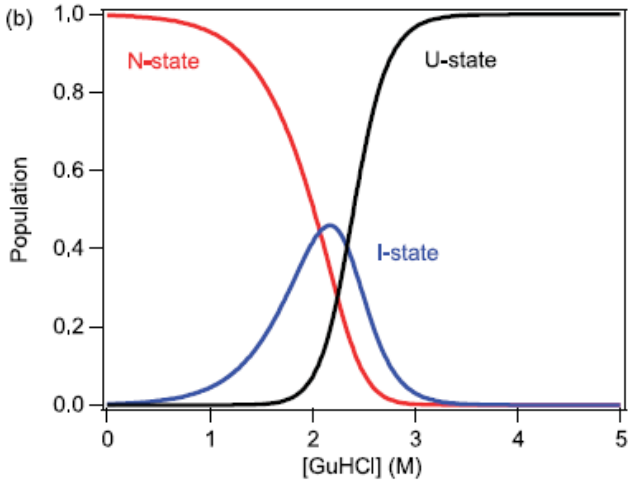
In many proteins, intermediates are populated.

Intermediates include proteins, where only partial structure has formed.

This complicates the analysis.

Multistate transitions in cytochrome c unfolding



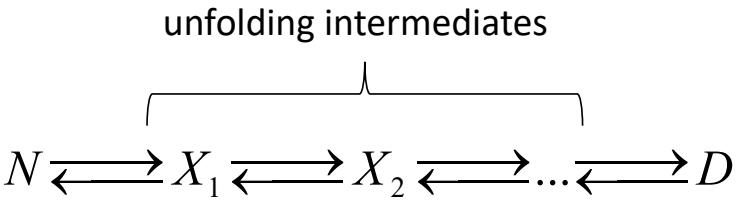


Latypov et al.
JMB 2006

If different parameters are observed, unfolding traces do no longer coincide.

Intermediates accumulate during the unfolding transition

Multi-state transition



apparent stability, eludicated by summing the individual contributions

$$K_{app} = K_D \frac{1 + \sum d_i \frac{K_i}{K_D}}{1 + \sum (1 - d_i) K_i} = \frac{K_D + \sum d_i K_i}{1 + \sum (1 - d_i) K_i}$$

$$d_i = \frac{y_i - y_N}{y_D - y_N}, \quad 0 < d_i < 1$$

$$K_i = \frac{[\mathrm{X_i}]}{[\mathrm{N}]} \qquad K_D = \frac{[D]}{[\mathrm{N}]}$$

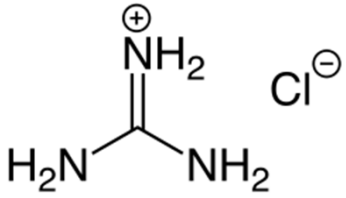
Denaturation with chemical agents

Molecular mechanism of denaturant action:

1. Direct effect:

   H-bonding to polar groups, mostly the protein backbone, thereby competing with internal H-bonds

   If charged: Interaction with ionic groups

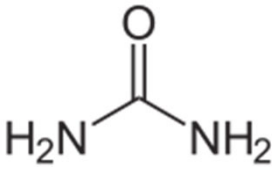


guanidinium chloride

2. Indirect effect:

   Alteration of water structure and thus diminishment of the hydrophobic effect

   Facilitation of the exposure of hydrophobic groups.



urea

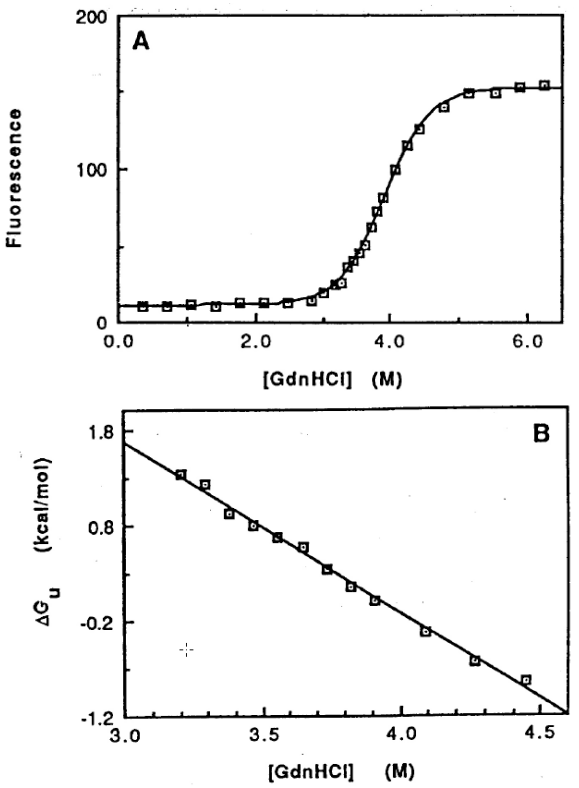
Protein denaturation with denaturants

The **effect of denaturants** on the free energy is linear (empirical finding)

$$\Delta G^0 = \Delta G^0_{H_2O} + m\,[\text{denaturant}]$$

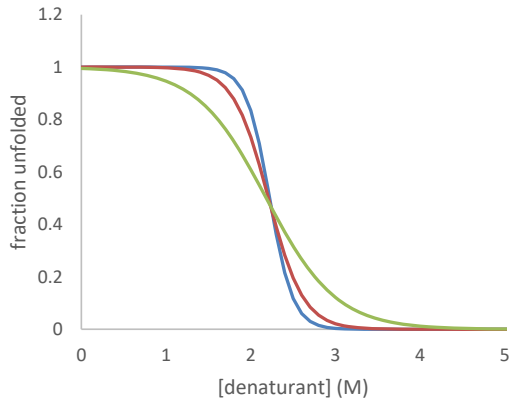
The **free energy of unfolding** can thus be determined by an extrapolation to **0 M denaturant**

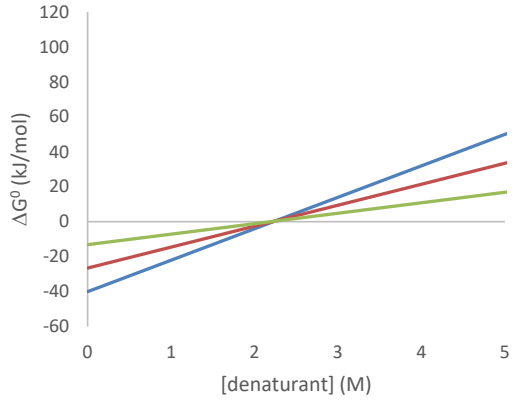
Example: Chymotrypsin inhibitor 2 (CI2)



Sensitivity to denaturant

simulated unfolding curves





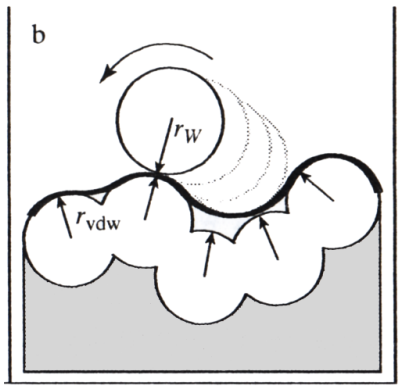
Parameters

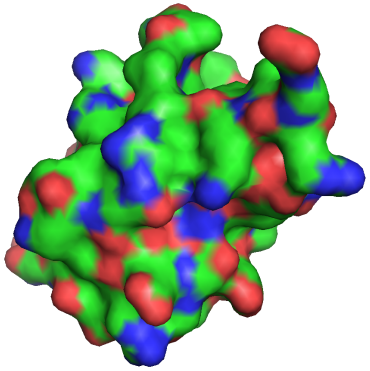
- ΔG_0 = -40 *kJ/mol*, m = 18 *kJ/mol/M*
- ΔG_0 = -26.5 *kJ/mol*, m = 12 *kJ/mol/M*
- ΔG_0 = -13.1 *kJ/mol*, m = 6 *kJ/mol/M*

molecular meaning of m-value:

- proportional of buried ASA
- proteins with large hydrophobic core exhibit high m-value
- the higher the m-value the stronger the dependence of a folding transition to denaturant (steepness)

M-values are proportional to change in ASA





Calculating ASA

- a water-sized sphere is rolled across the chemical structure keeping VdW radii
- the accessible surface corresponds to the ASA

molecular meaning of m-value:

- proportional to change in ASA
- proteins with large hydrophobic core exhibit high m-value
- the higher the m-value the stronger the dependence of a folding transition to denaturant (steepness)

Small quiz:

- We have a small protein, whose standard free energy of folding (stability) is $\Delta G^{o}_{f} = 20\ kJ/mol$
- Upon addition of guanidinium hydrochloride (GdmHCl), the protein denatures reversibly
- Fluorescence measurements determined a mid-point of the transition at 2 M GdmHCl
- What is the *m-value* for this protein?
- If we compare this to a different protein with $m = 5\ kJ/mol/M$, what can we say about its structure?