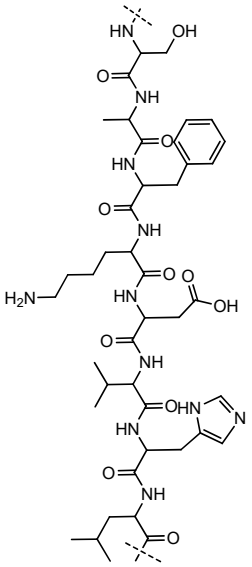
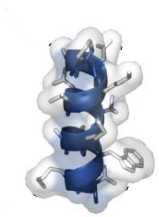
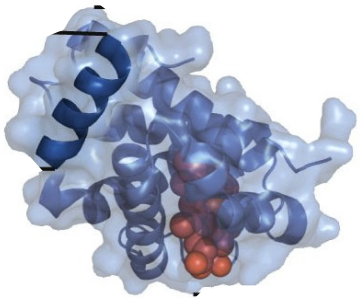
Proteins – Structural hierarchy



Amino acid chain
primary structure



α-helix
secondary
structure



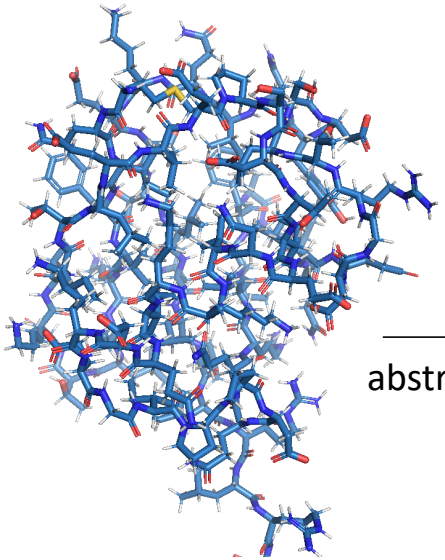
Protein domain
Tertiary structure

Protein complex
Quarternary structrue

transitions

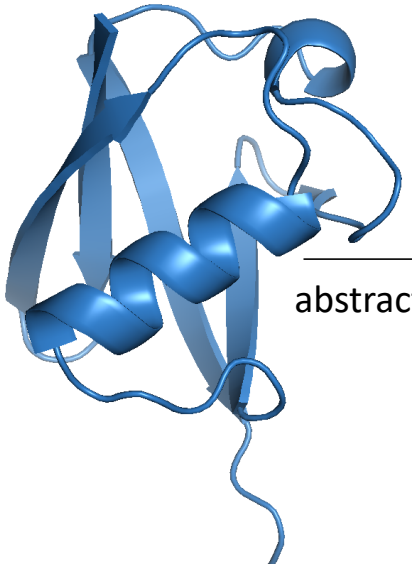
Protein organization in 3D

molecular representation



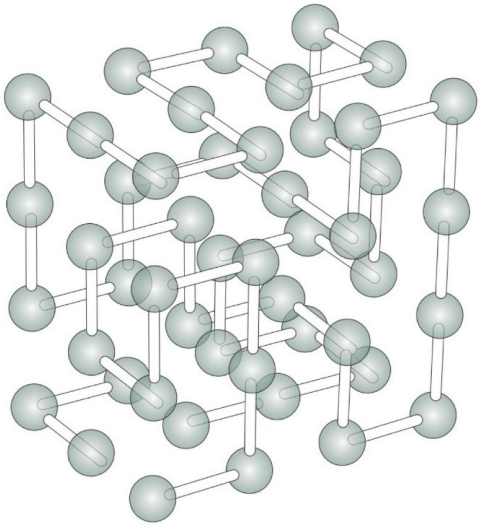
abstraction

cartoon
(secondary structure)



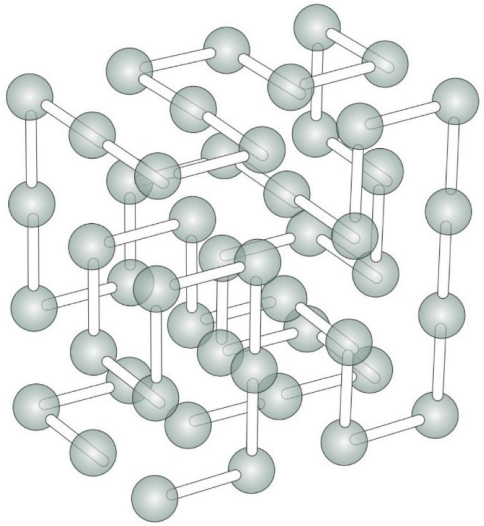
abstraction

lattice model

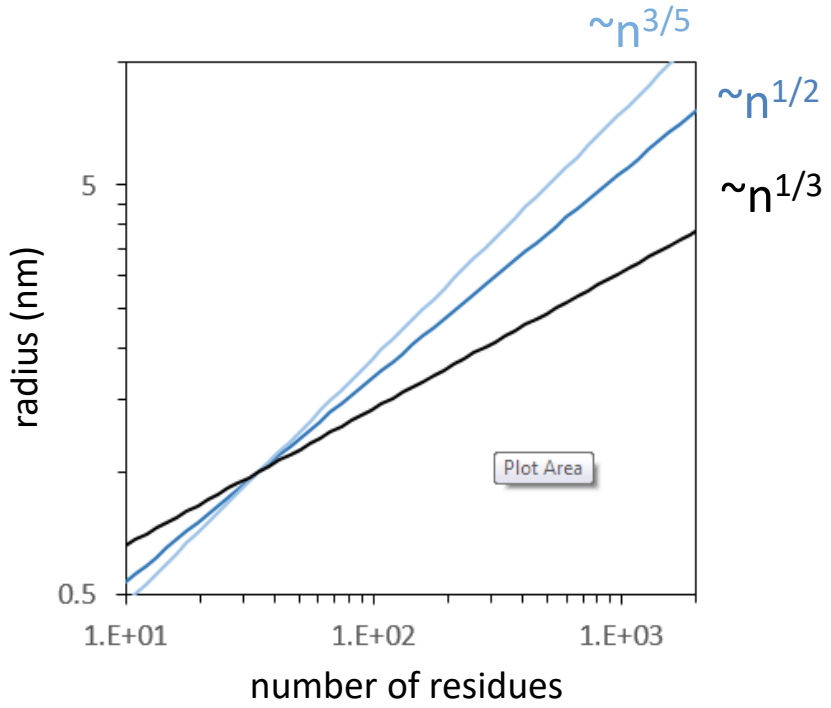


random walk on a 3D lattice

Scaling laws for protein size

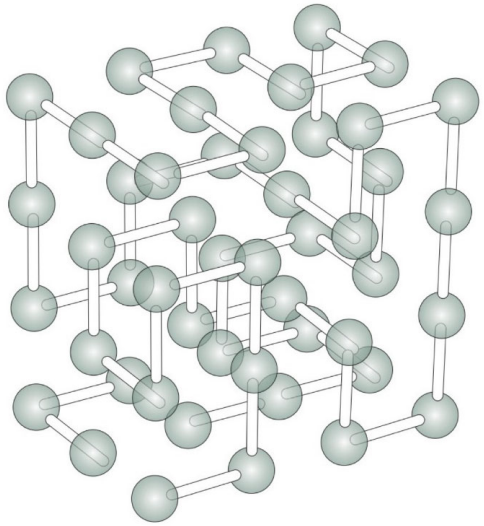


Scaling laws for a random walk
self-avoiding chain: $r \sim n^{3/5}$
random chain: $r \sim n^{1/2}$
compact chain: $r \sim n^{1/3}$

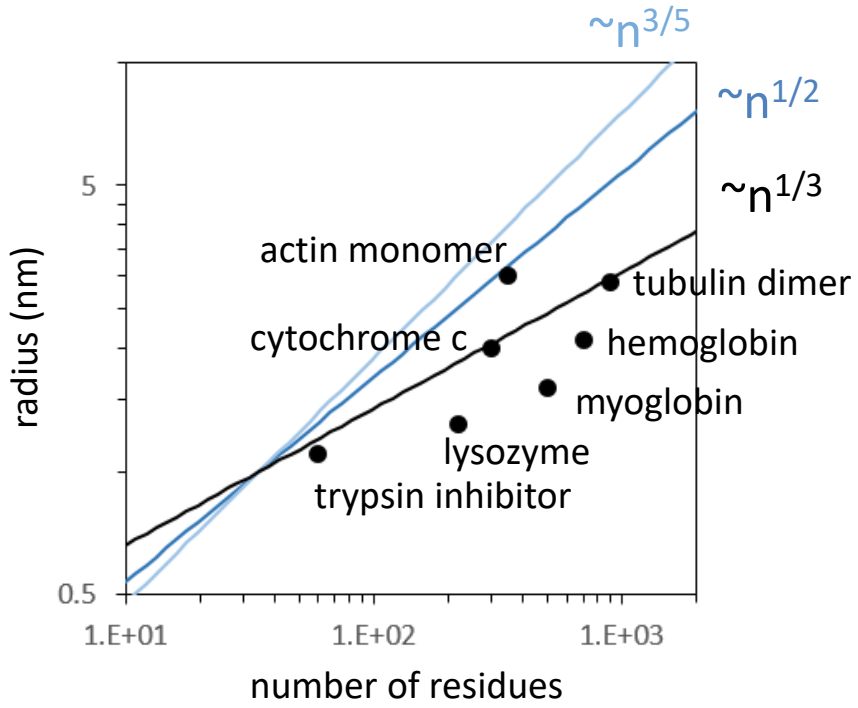


Adapted from Physical Biology of the Cell (Chapter 8)

Scaling laws for protein size



Scaling laws for a random walk
self-avoiding chain: $r \sim n^{3/5}$
random chain: $r \sim n^{1/2}$
compact chain: $r \sim n^{1/3}$



Adapted from Physical Biology of the Cell (Chapter 8)

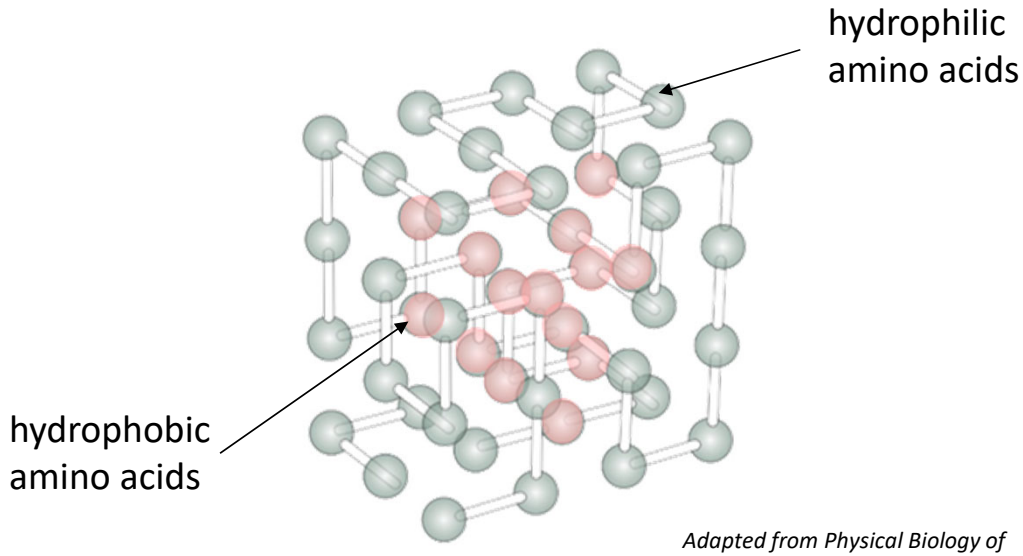
How does a protein find its native state?

Random search:

Even on a lattice (6 degrees of freedom) 100 aa protein has $6^{100} = 6.5 \times 10^{77}$ conformations, random search would take astronomical amount of time

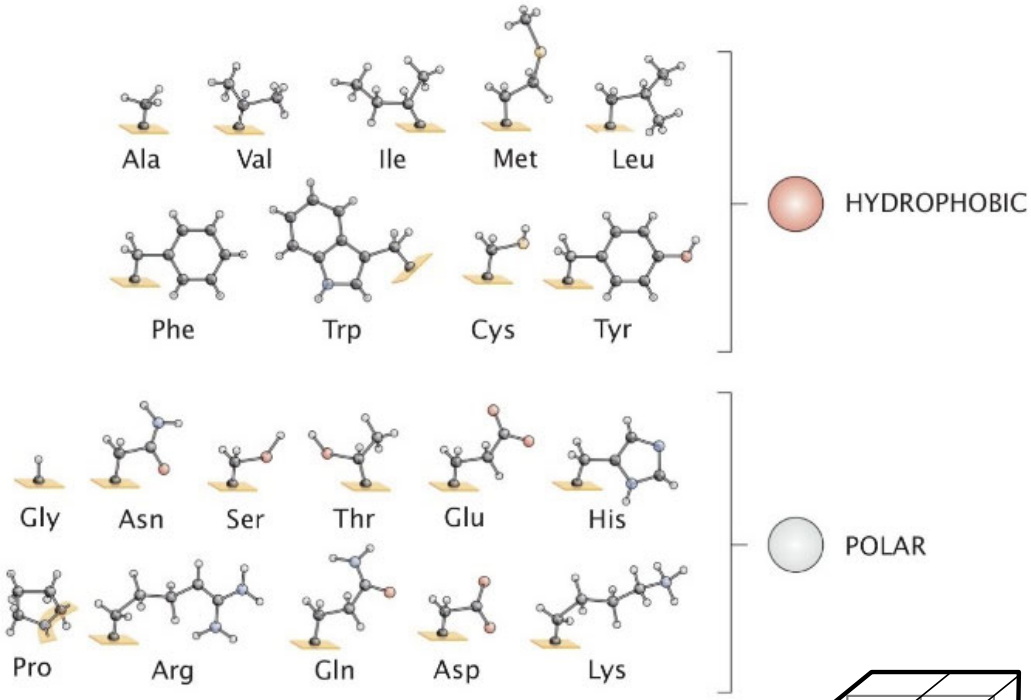
Mechanism:

Hydrophobic collapse

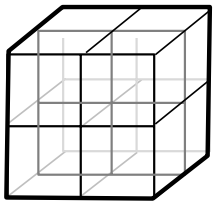


Adapted from Physical Biology of the Cell (Chapter 8)

HP-models of protein folding



Adapted from Physical Biology of the Cell (Chapter 8)



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

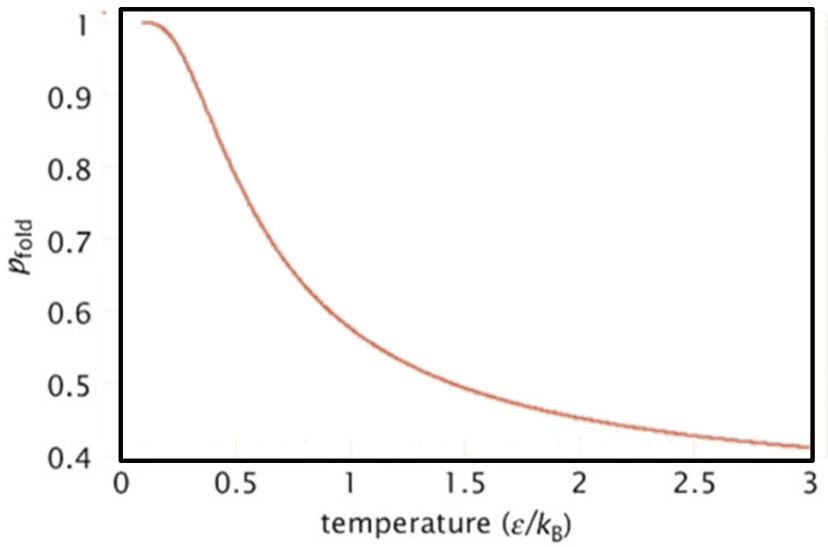
→ **2 letter alphabet allows an abstract model of structure formation**

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



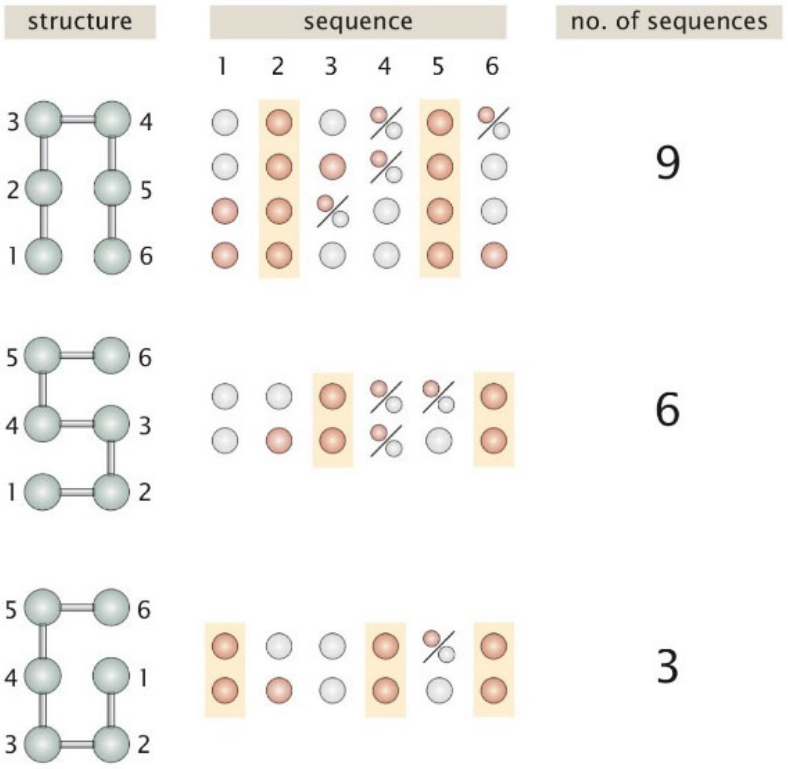
$$p_{fold} = \frac{e^{-2\beta\varepsilon}}{e^{-2\beta\varepsilon} + 2e^{-4\beta\varepsilon}} = \frac{e^{-2\beta\varepsilon}}{Q}$$

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

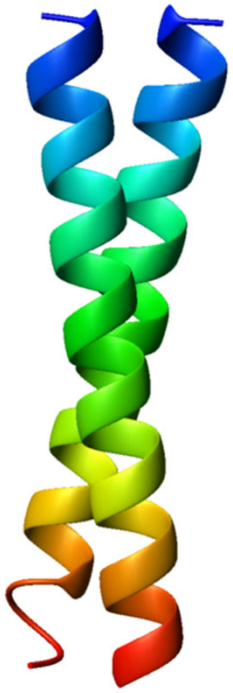
→ determine energy of each

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

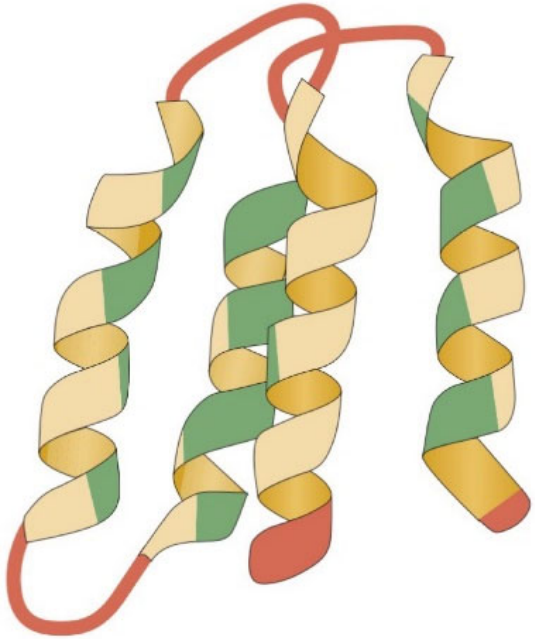
Secondary structure elements form the **elementary motifs** of protein structure.

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



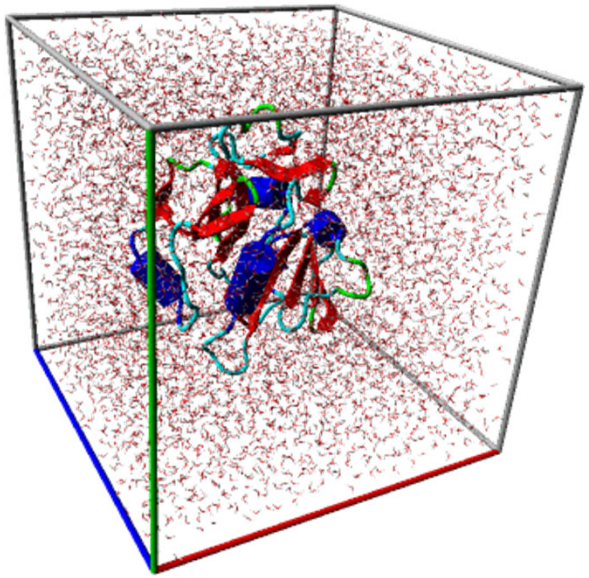
Hecht et al. Protein Sci 2004

Secondary structure
→ fixed, helical

Tertiary structure:
follows HP rules

Provides a platform, on which further functionality can be designed

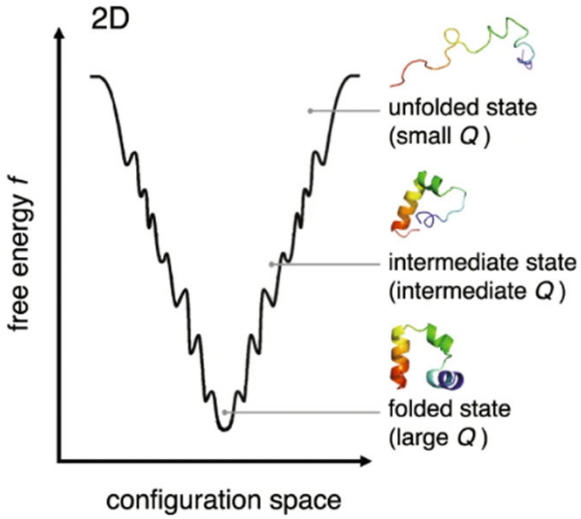
Force field for protein structure



https://in.explara.com/

Force field for protein structure evaluation:

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

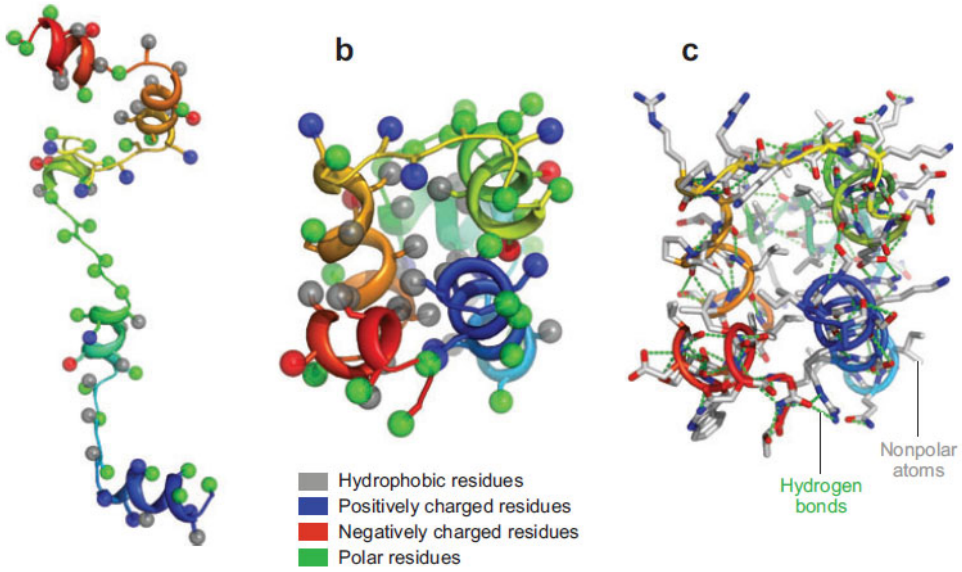


https://www.nature.com/articles/s41598-019-50825-6

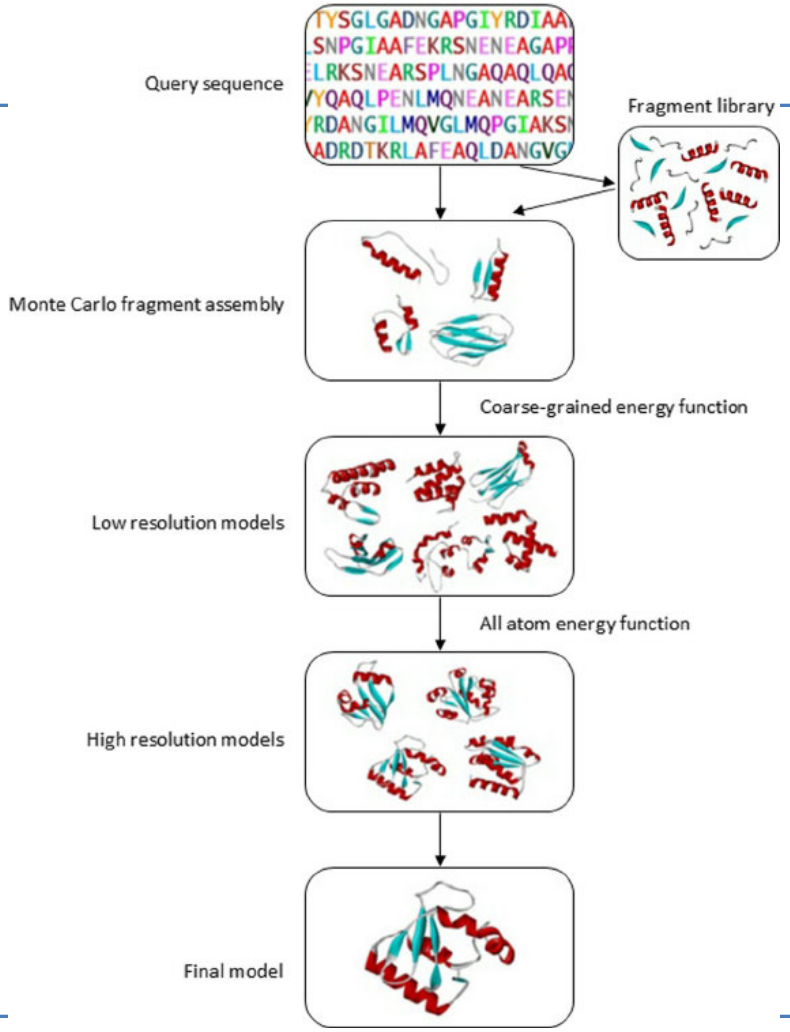
Goal: Minimize the configurational energy, find the global minimum

Protein structure prediction / design

Design from fragment libraries,
followed by refinement (Rosetta)

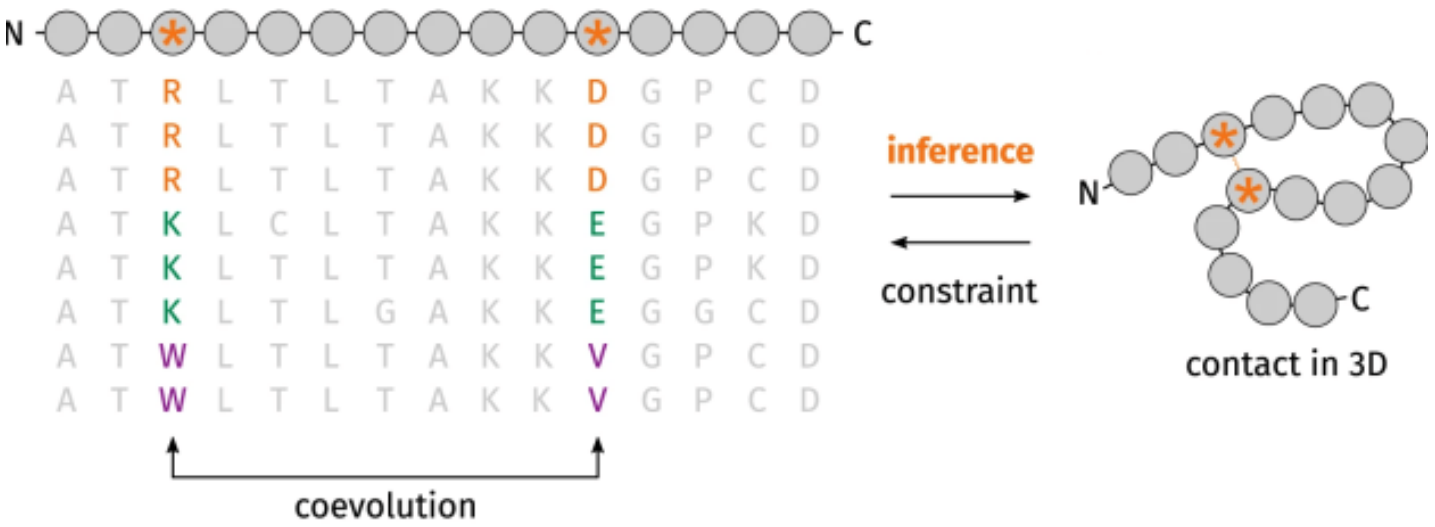


Das & Baker, Annu Rev Biochem 2008



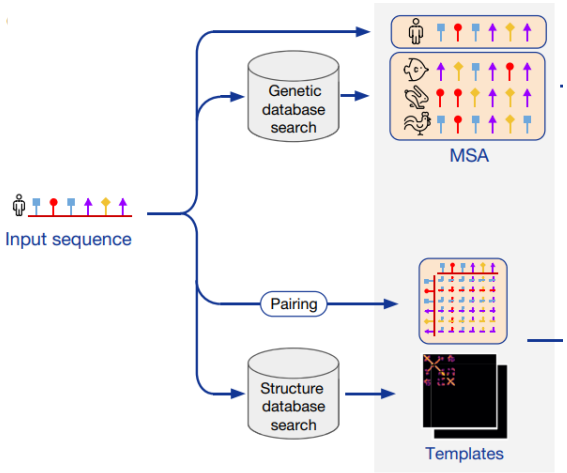
https://tbiomed.biomedcentral.com/articles/10.1186/s12976-015-0014-1

Structural information from sequence: Co-evolution analysis



43146:42mrxuqdd1srqh1335;:99

Protein structure prediction – ML approaches



Input:

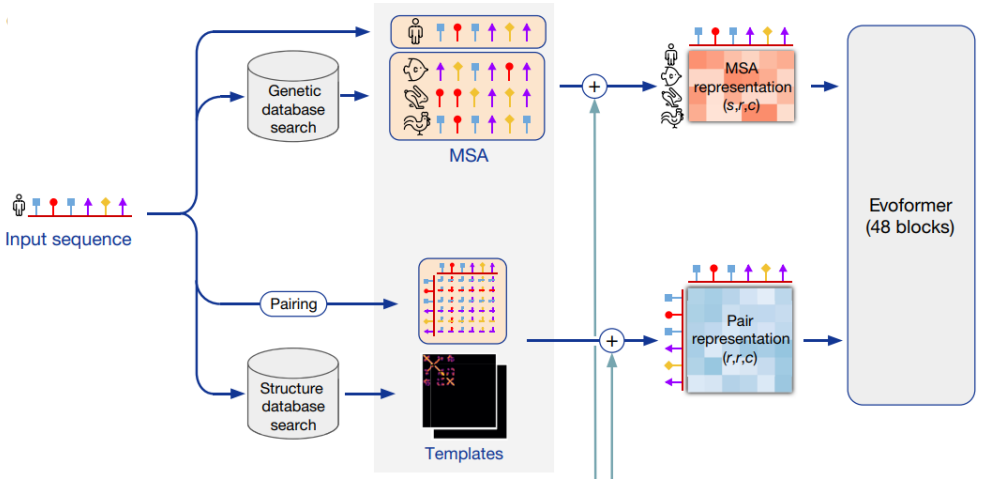
- primary sequence of target protein
- **multiple sequence alignments (MSA)** of homologues

- structural information of homologues (PDB)

Training set 170k structures of the PDB

AlphaFold2 - Nature 2021

Protein structure prediction – ML approaches



Input:

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

Evoformer

MSA Representation Update

Each residue in each sequence in the MSA is treated as a "token."

- Attention layers connect residues **vertically** to capture **co-evolution**
- **Row attention** (along each sequence) → sequence context.
- **Column attention** (across sequences) → correlations between positions.

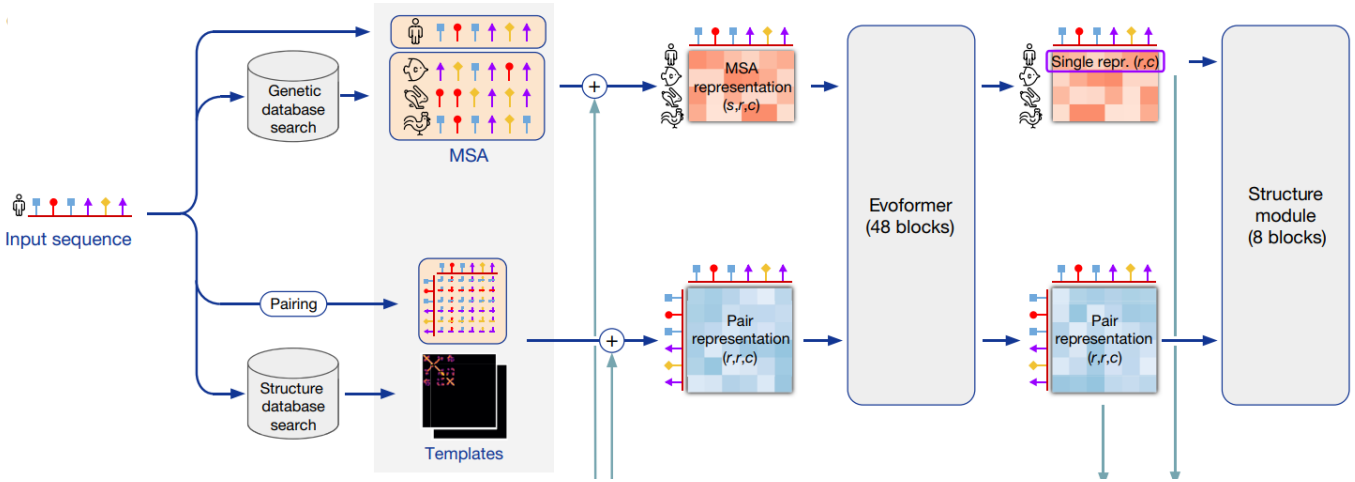
Pair Representation Update

pair representation for each residue pair (i, j), derived from sequence features and positional encodings.

- Pair matrix represents all-to-all spatial relationships.
- It is updated through “triangular updates,” (if A–B and B–C are related, then A–C should be consistent).

AlphaFold2 - Nature 2021

Protein structure prediction – ML approaches



Structure Module translates learned pairwise relations into 3D coordinates:

- **backbone frame** for each residue (a local coordinate system).
- Iteratively refines frames using **invariant point attention (IPA)** — a special geometric attention layer that respects 3D rotations and translations.
- Outputs **atomic coordinates** (N, Cα, C, O atoms, optionally side chains).

→ **Module understands actual 3D geometry, not just pairwise distances.**

Input:

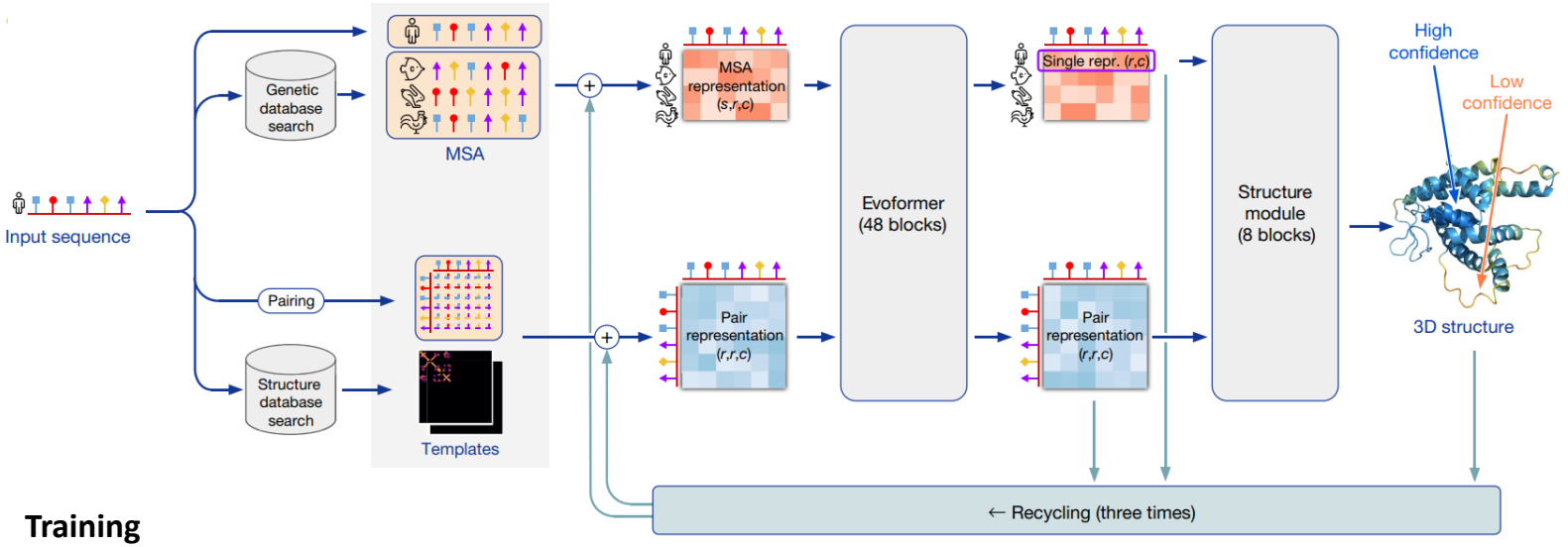
- primary sequence of target protein
- multiple sequence alignments of homologues
- structural information of homologues (PDB)

Evoformer:

- MSA representation
- Residue pair representation

AlphaFold2 - Nature 2021

Protein structure prediction – ML approaches



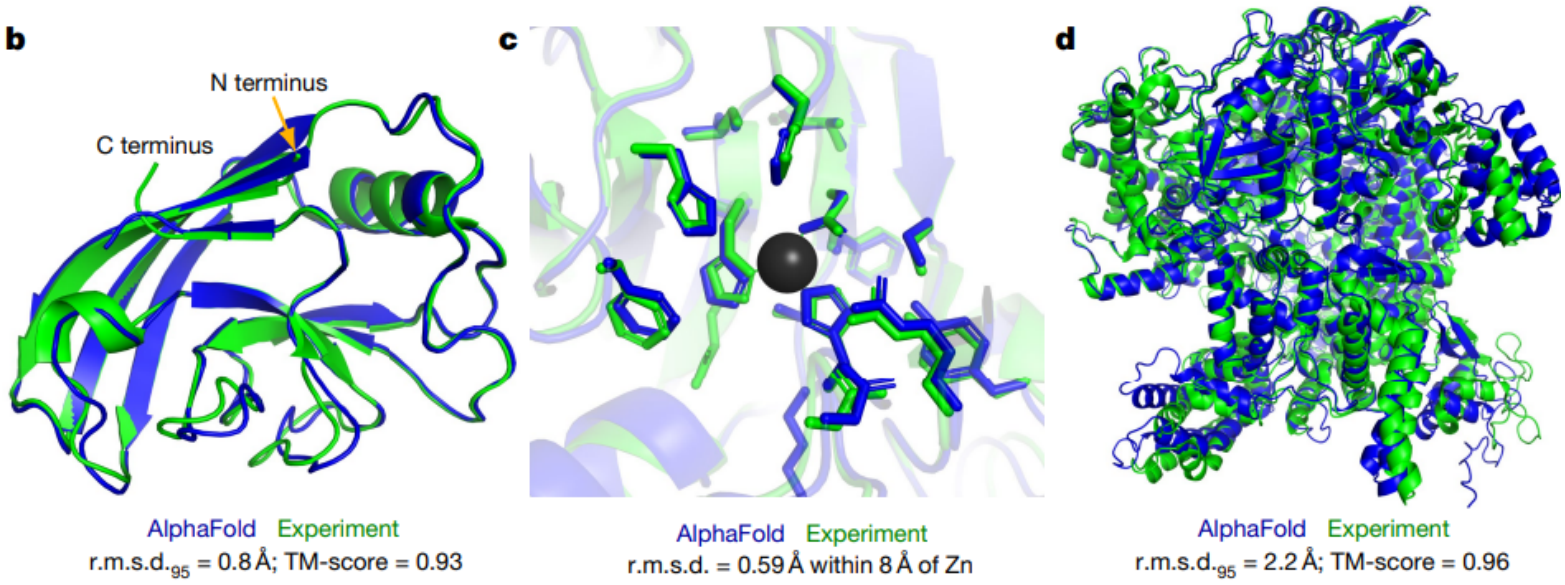
Training

The model was trained on ~170k experimentally solved PDB structures. **Key losses include:**

- **Frame Aligned Point Error (FAPE)**: geometric difference between predicted and true relative atomic coordinates.
- **Distogram loss**: cross-entropy on inter-residue distances and orientations.
- **Predicted LDDT** (local distance difference test): measures model confidence per residue.
- **Auxiliary losses** on MSA and pair outputs.

AlphaFold2 - Nature 2021

Protein structure prediction – ML approaches



Result:

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

AlphaFold - Nature 2021

Difference to earlier methods

structure prediction pipelines included **hand-coded structural constraints**, such as:

- Allowed ϕ/ψ backbone torsion angles (Ramachandran maps)
- Fixed bond lengths and bond angles
- Steric clash penalties
- Hydrogen-bond potentials
- Energy terms for hydrophobic burial or electrostatics

→ explicit **energy functions** to search for lowest-energy conformation

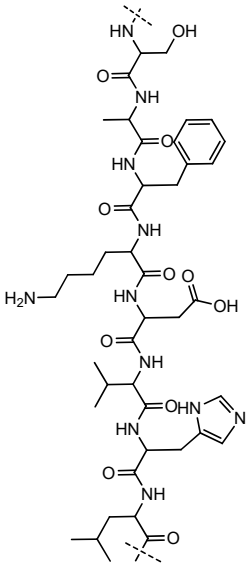
AlphaFold2 does not have any of these explicit physics-based terms

→ it learns everything directly from training set

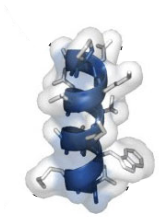
- The **Evoformer** and **Structure Module** learn what geometrically valid proteins *look like*.
- Outputs are 3D coordinates trained against **experimental PDB structures** via loss function
- If impossible torsion angles or clashes are predicted → loss on atomic coordinates and local frames penalizes

→ network *learns* to avoid unphysical geometry

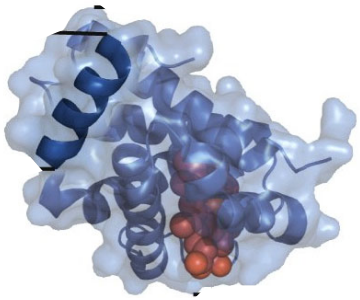
Protein structure – A thermodynamic view



Amino acid chain
primary structure



α-helix
secondary
structure



Protein domain
Tertiary structure

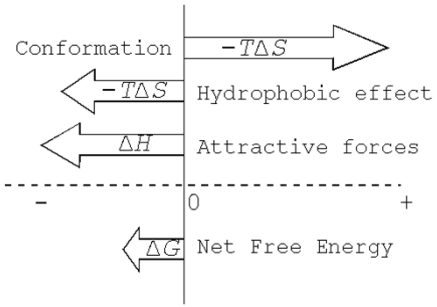
Protein complex
Quarternary structrue

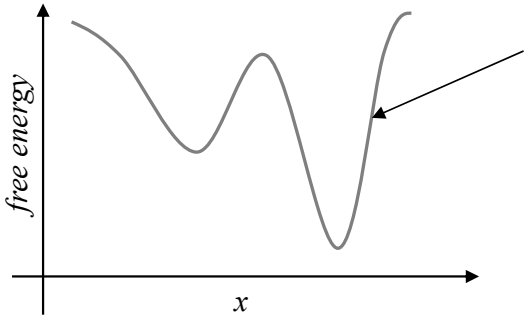
transitions

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

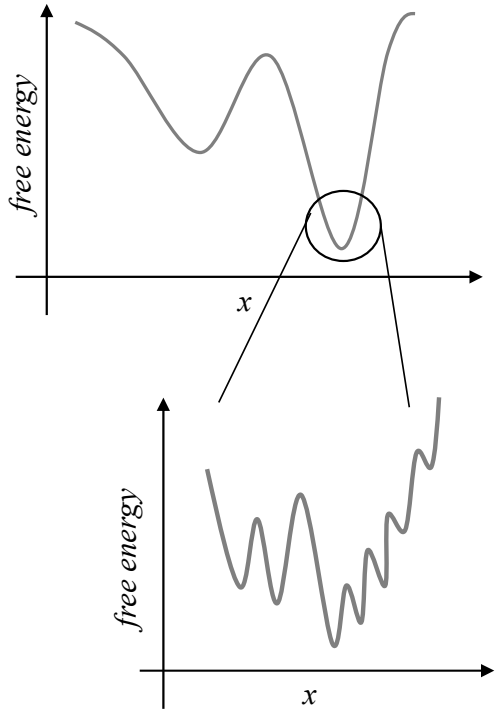
Global states in proteins

For a biophysical understanding → simplifications are required.

Global state can be
Coarse grained characterized

ensemble

***i* Microstates** interchanging
Atom positions rapidly
Dihedral angles
Bond vibrations characterization
... difficult



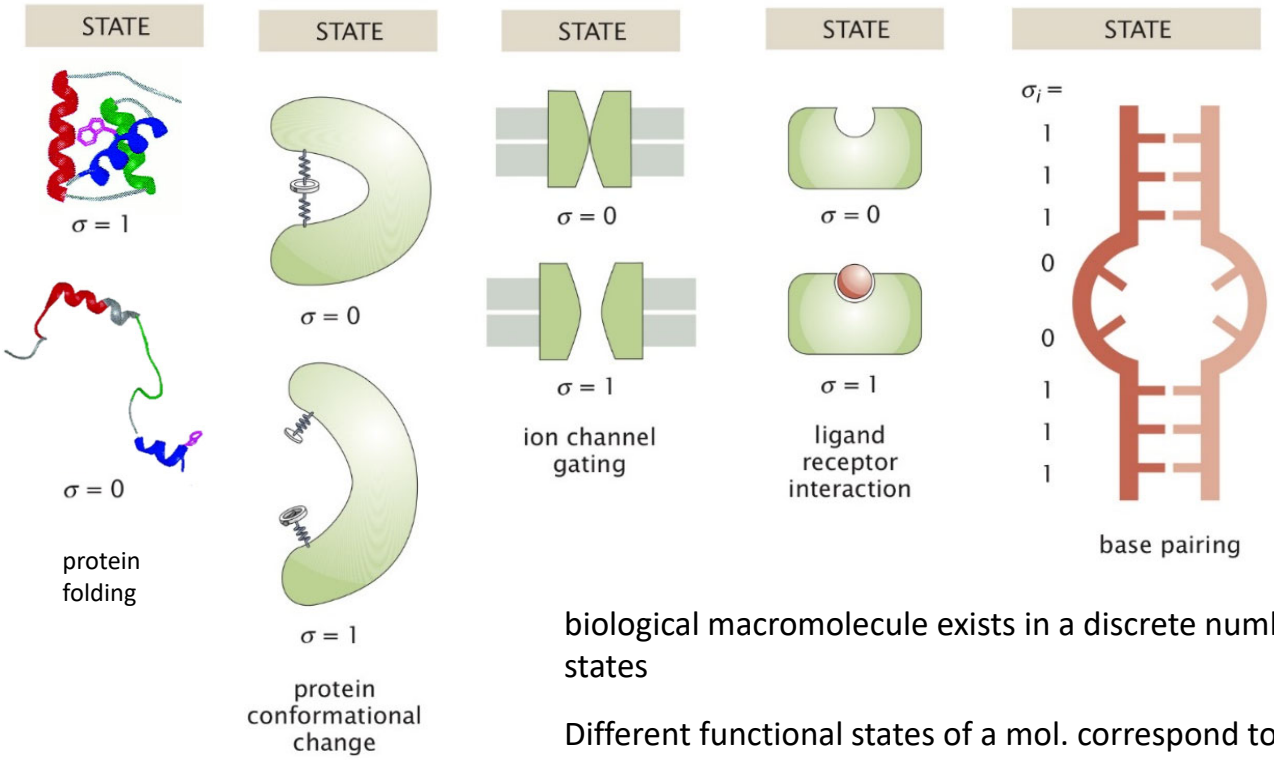
Global states in proteins

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

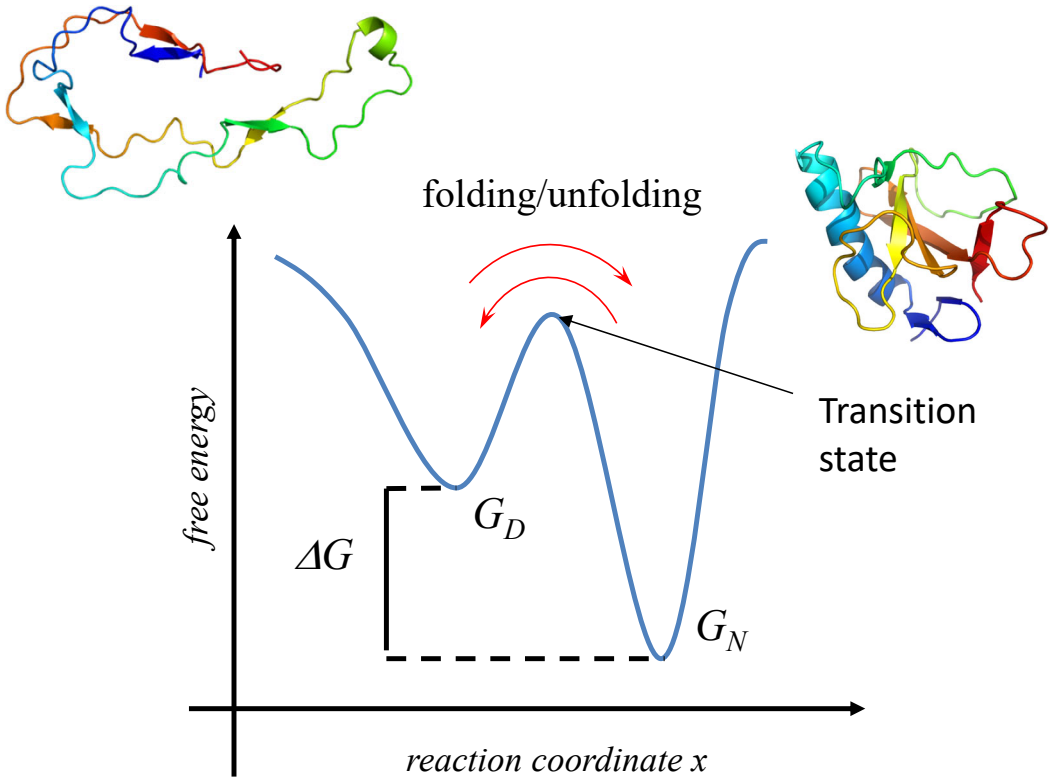


biological macromolecule exists in a discrete number of global states

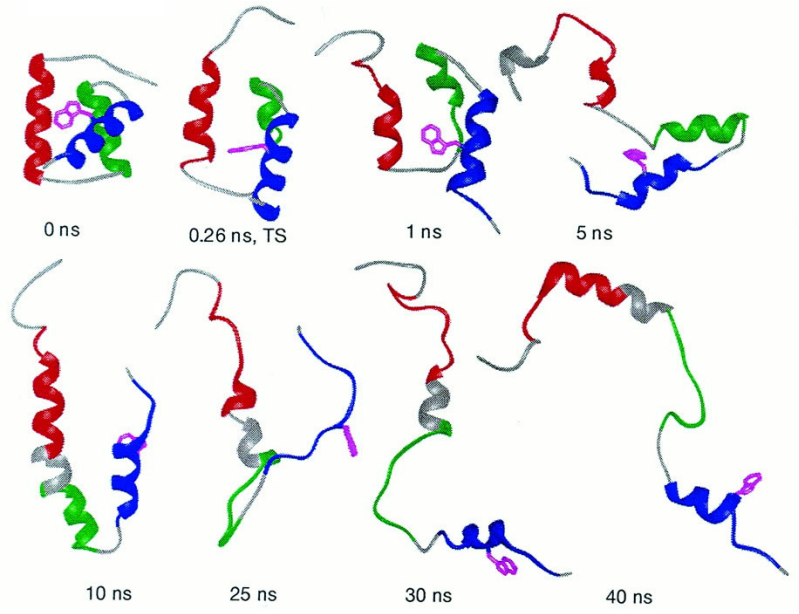
Different functional states of a mol. correspond to different global states

Changes in global states could be related to function

Global states in a protein: N and D



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)

→ unfolding eq.

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)

Absorption processes

Definition of light absorption:

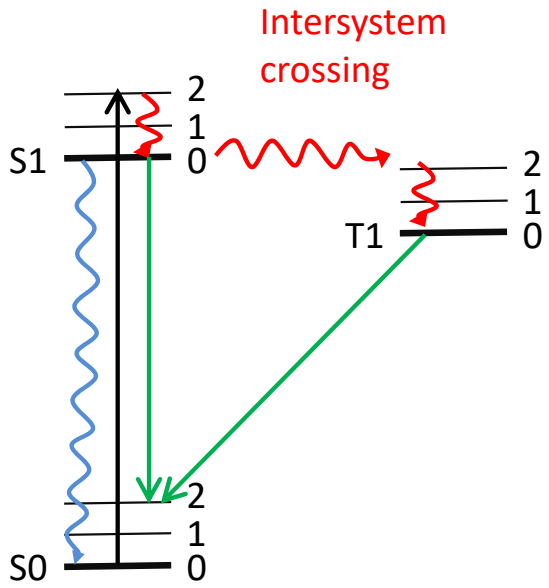
a photon of energy $E = h\nu = \frac{hc}{\lambda}$

is absorbed by a molecule:

Fate of the excited state:

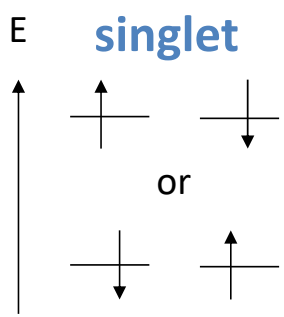
- internal conversion: heat
- emission of a photon: **fluroescence**, **phosophorescence**

Jablonski diagram



Part II: Emission spectroscopy

Emission from a singlet state:
Fluorescence

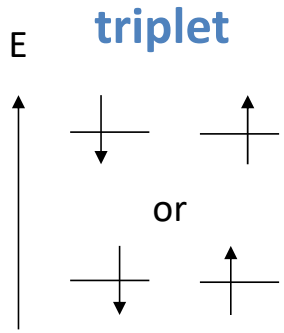


spin = 0

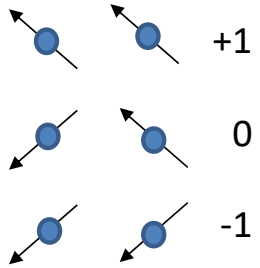
ways for achieving total spin

from a triplet state:
Phosphorescence

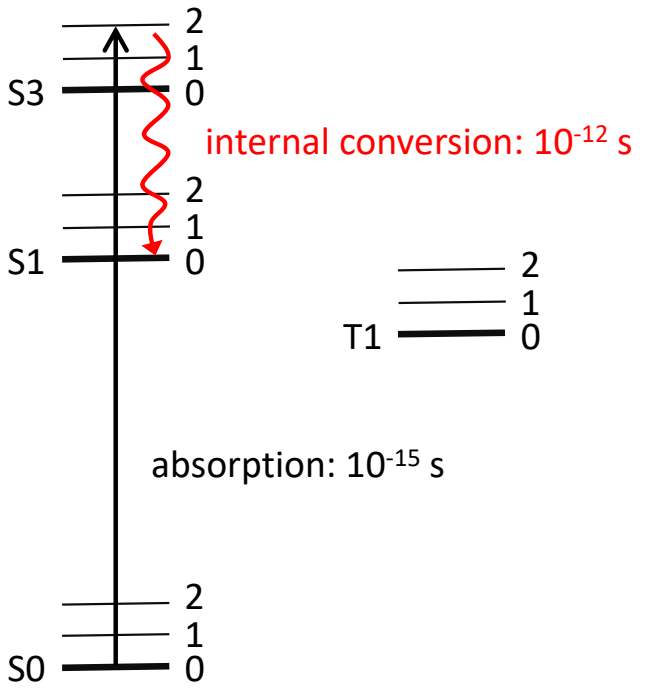
- excitation into triplet requires spin-change in electron
- similarly, emission from a triplet state requires spin change



spin = 1



Lifetimes of excited states and emission



Absorption

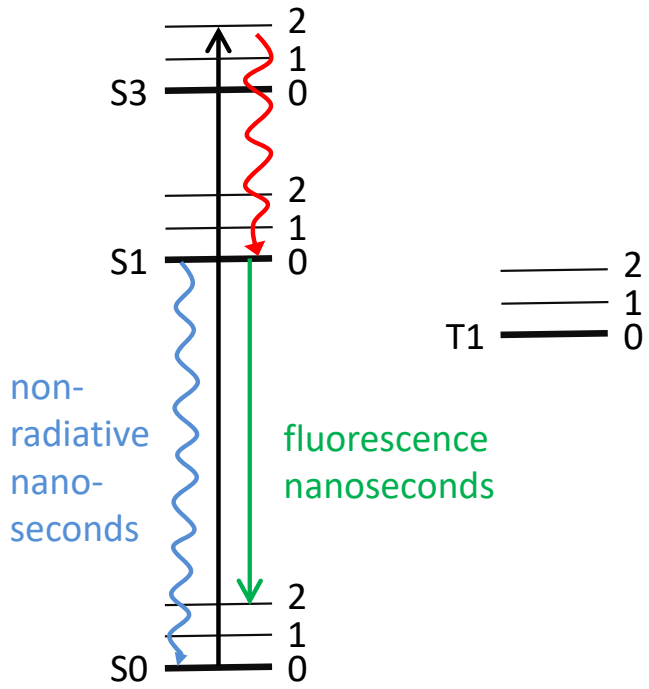
time: 10^{-15} s, corresponds to the time it takes for light wave to pass molecule

Internal conversion

Molecule returns to lowest singlet state, lowest vibrational niveau

time: 10^{-12} s

Lifetimes of excited states and emission



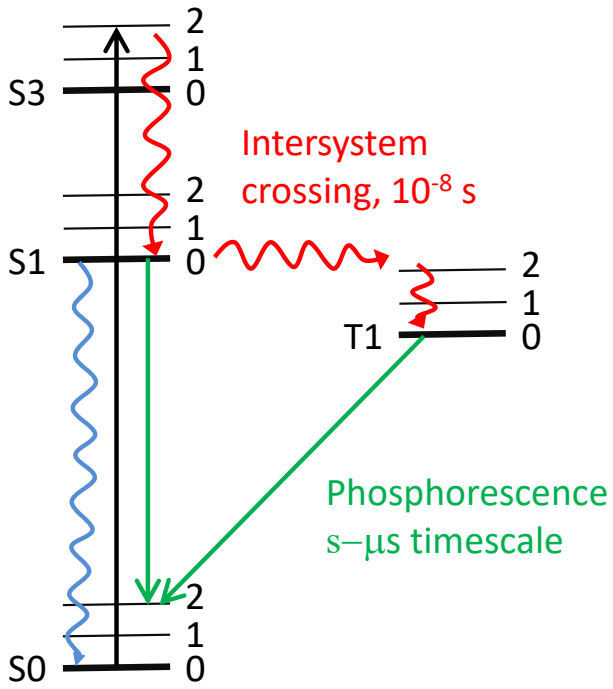
Fluorescence

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

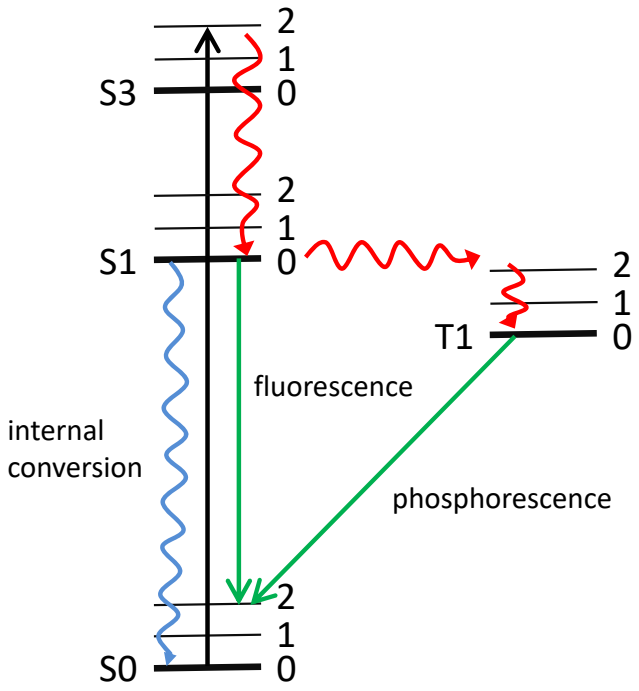
Phosphorescence

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

requires again inversion of e- spin

forbidden transition, thus long timescale

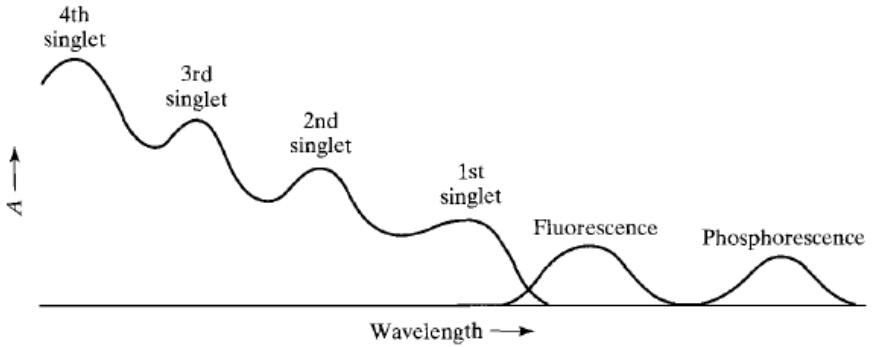
Absorption and emission spectra



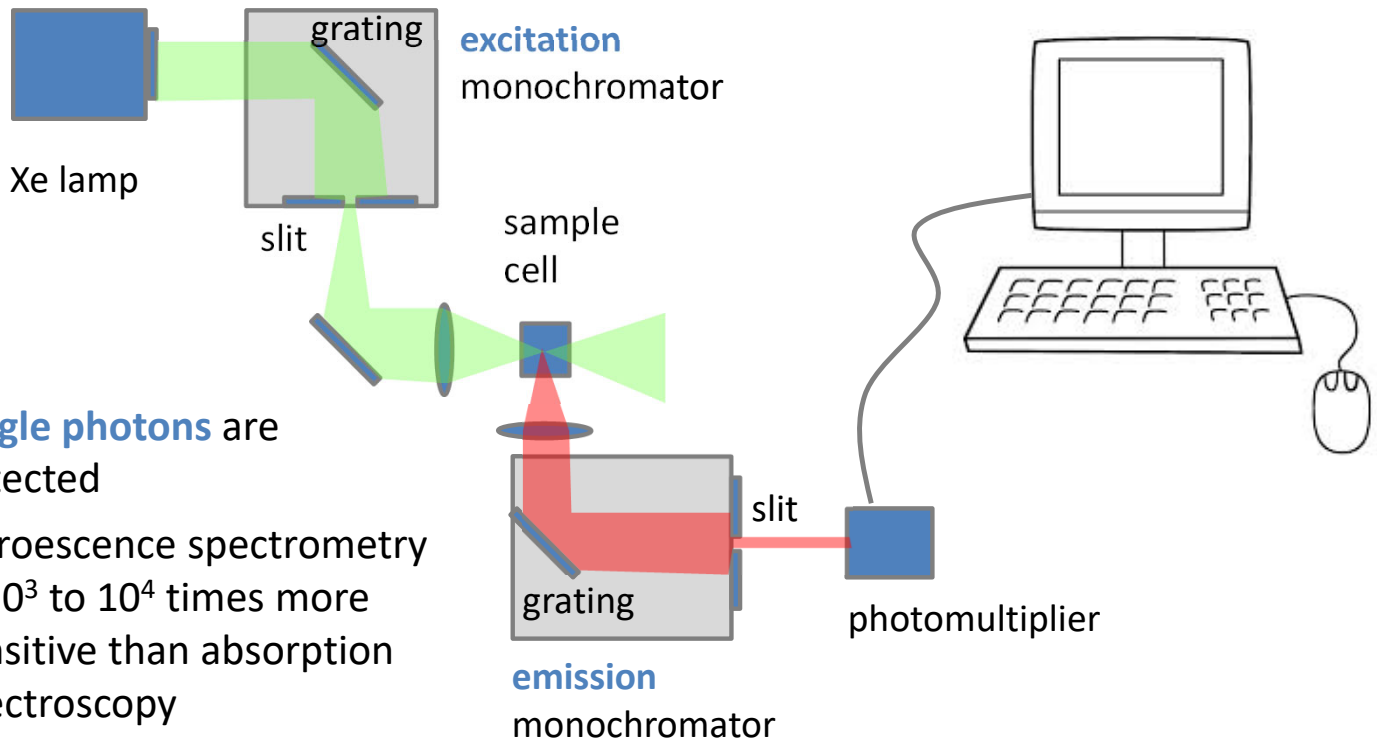
Spectra:

Fluorescence is red-shifted to the absorption spectrum

phophorescence is at even higher wavelengths



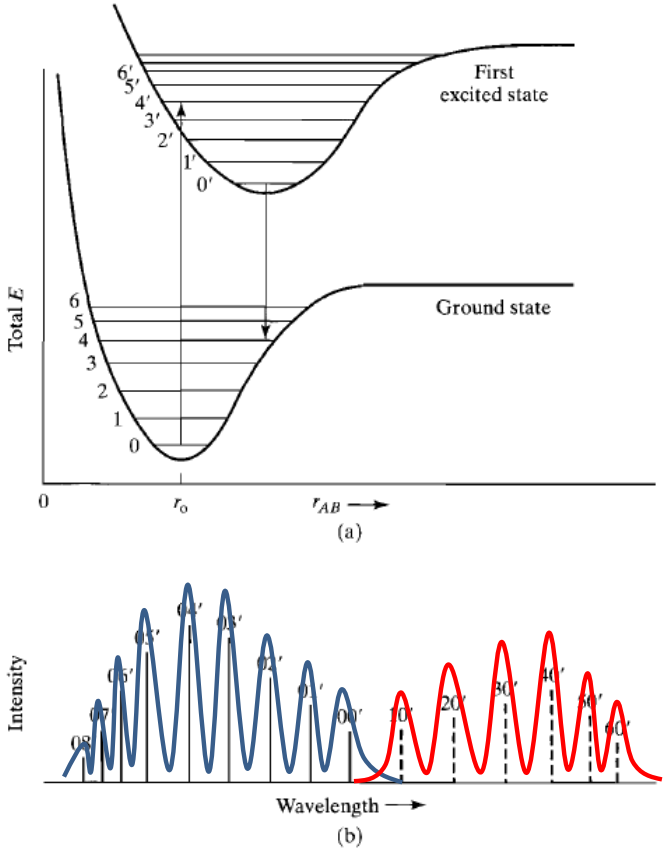
Measuring fluorescence



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

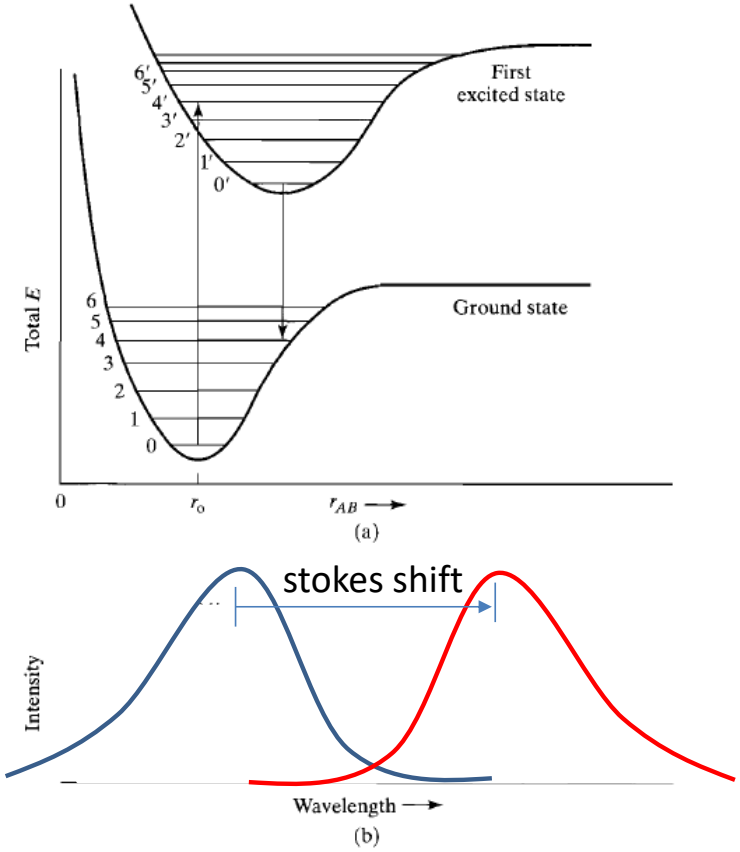
ground state -> s1

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**

Quantum yield

The quantum yield is defined as:

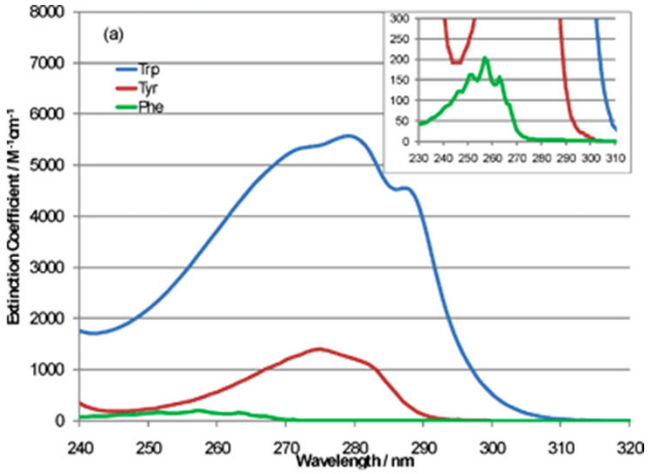
$$Q_F = \frac{No.\ emitted\ photons}{No.\ absorbed\ photons}$$

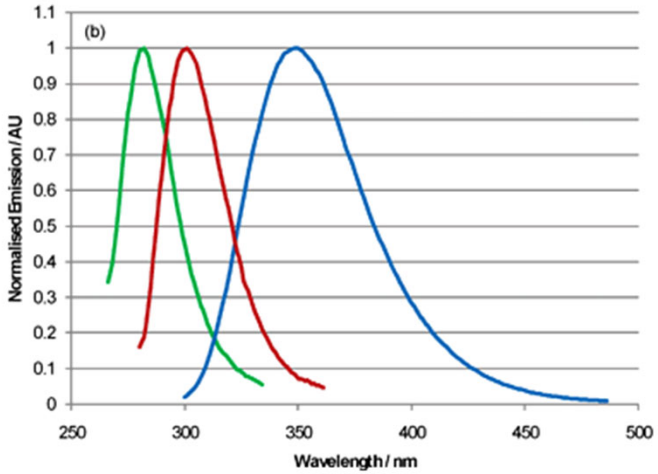
Even with a quantum yield $Q_F = 1$, the energy output is below 100%, as the fluorescence is at a higher wavelength.

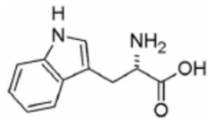
Strong fluorescence observed with molecules with strong π-$\pi*$ absorption, n-π* transitions often do not fluroesce.

Amino acids: Fluroescence spectra

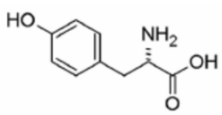
Absorbance and fluroescence spectra of aromatic amino acids



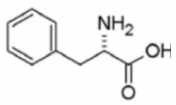




Trp



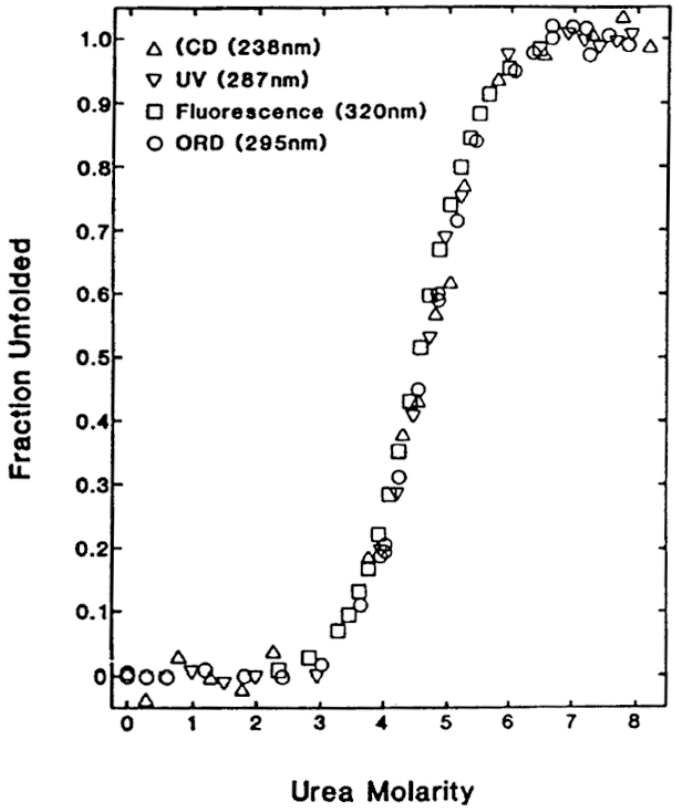
Tyr



Phe

Quantum yield for tryptophan: $Q_F=0.2$

Denaturation of globular, monomeric protein : RNase T1



$$N \rightleftarrows D$$

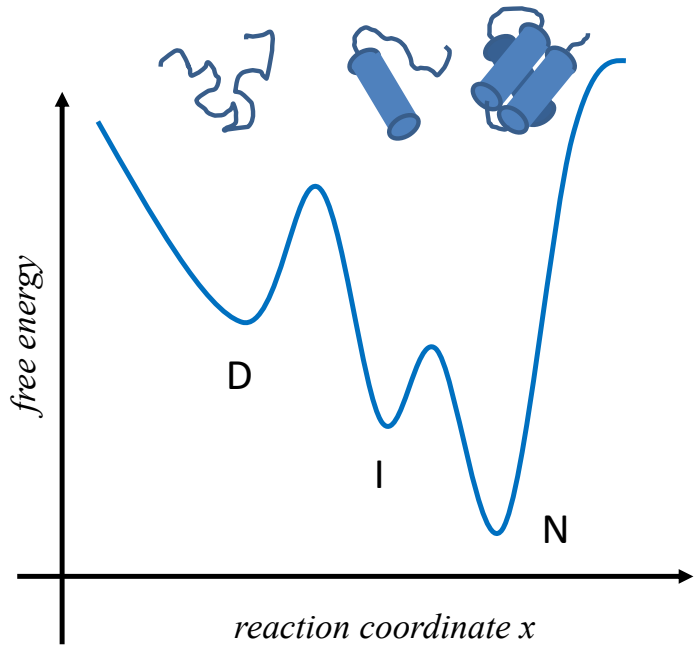
$$y = [\mathrm{N}]\,\mathrm{y_N} + [D]\,\mathrm{y}_D$$

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

different parameters completely overlap

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

Three (or more) state equilibrium transitions

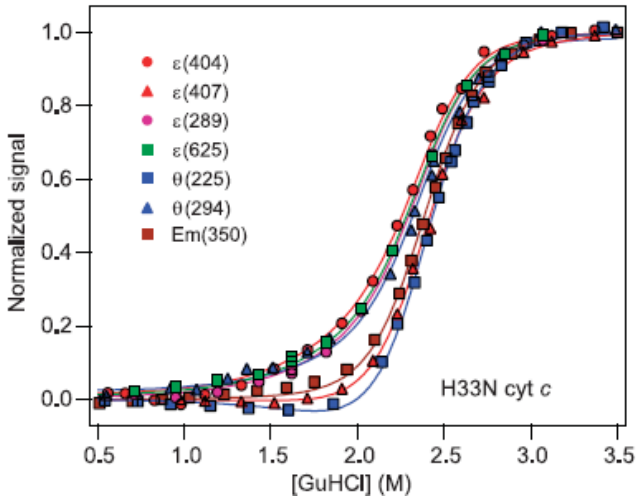


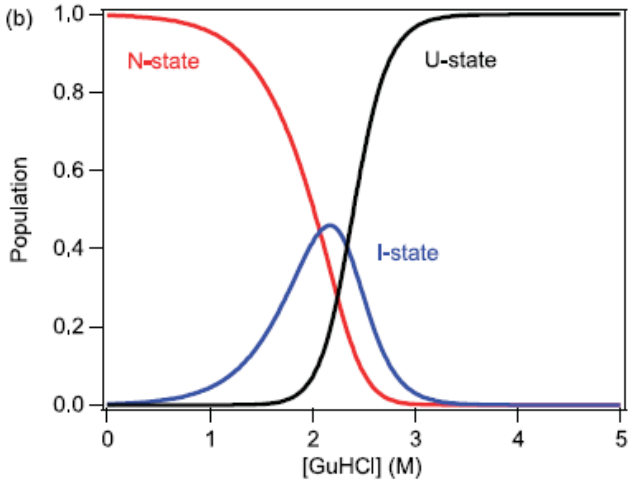
In many proteins, intermediates are populated.

Intermediates include proteins, where only partial structure has formed.

This complicates the analysis.

Multistate transitions in cytochrome c unfolding



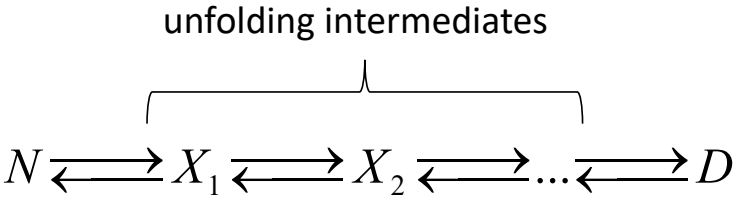


Latypov et al.
JMB 2006

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

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

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

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

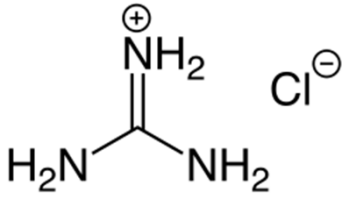
Denaturation with chemical agents

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

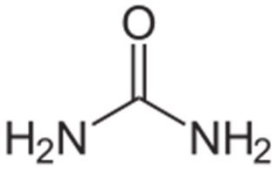


guanidinium chloride

2. Indirect effect:

   Alteration of water structure and thus diminishment of the hydrophobic effect

   Facilitation of the exposure of hydrophobic groups.



urea

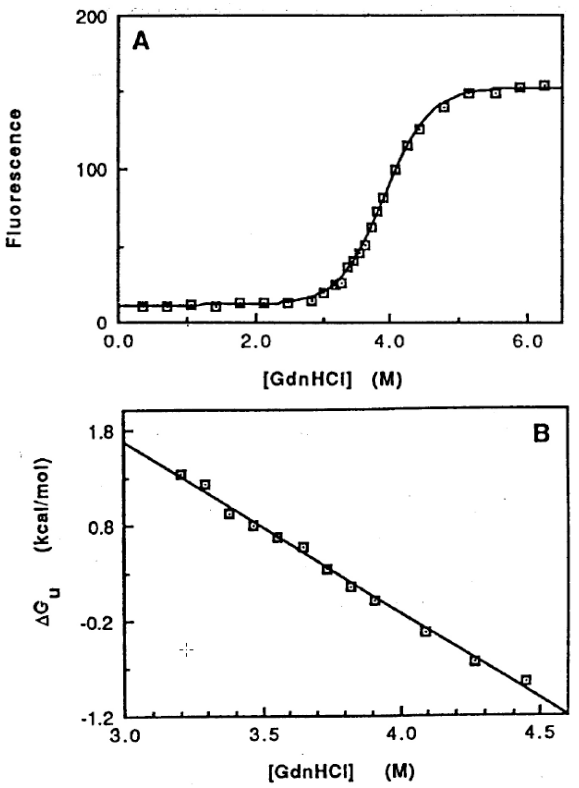
Protein denaturation with denaturants

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

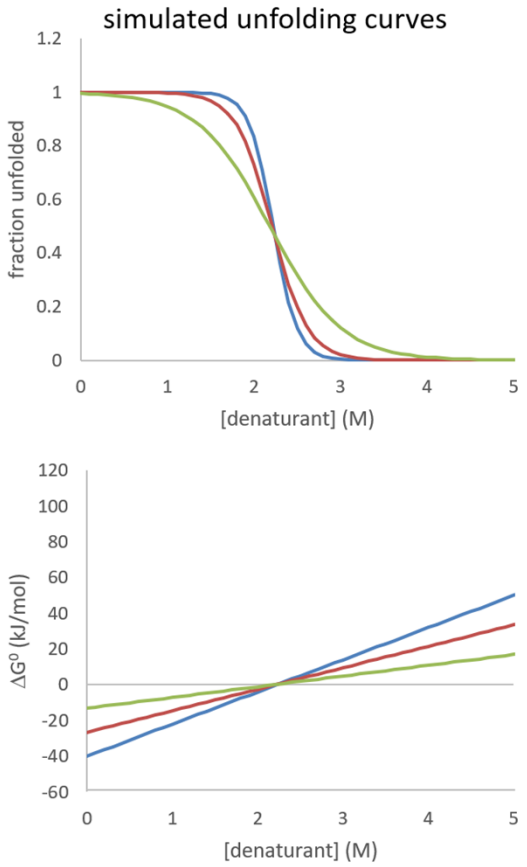
$$\Delta G^0 = \Delta G^0_{H_2O} + m\,[\text{denaturant}]$$

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

Example: Chymotrypsin inhibitor 2 (CI2)



Sensitivity to denaturant



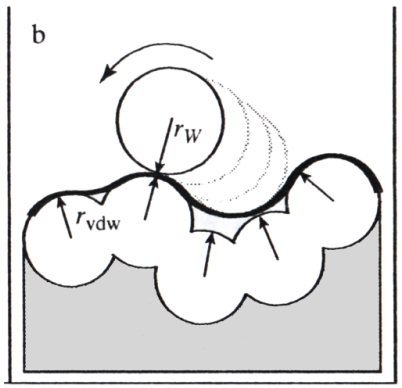
Parameters

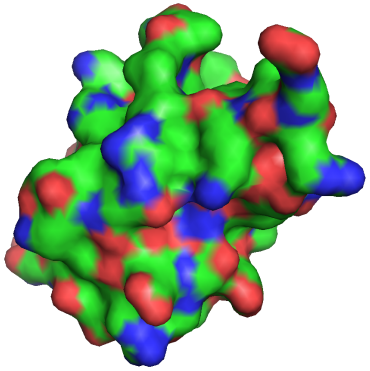
- ΔG_0 = -40 *kJ/mol*, m = 18 *kJ/mol/M*
- ΔG_0 = -26.5 *kJ/mol*, m = 12 *kJ/mol/M*
- ΔG_0 = -13.1 *kJ/mol*, m = 6 *kJ/mol/M*

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:

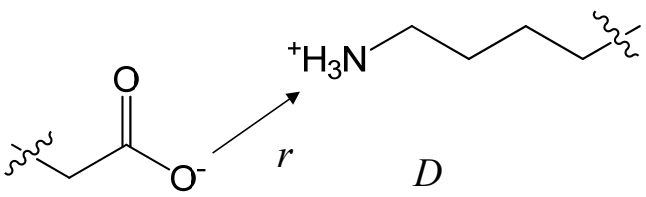
- proportional to change in ASA
- proteins with large hydrophobic core exhibit high m-value
- the higher the m-value the stronger the dependence of a folding transition to denaturant (steepness)

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?

Protein denaturation by pH

- Proteins can also be denatured by pH



Contribution to protein stability:

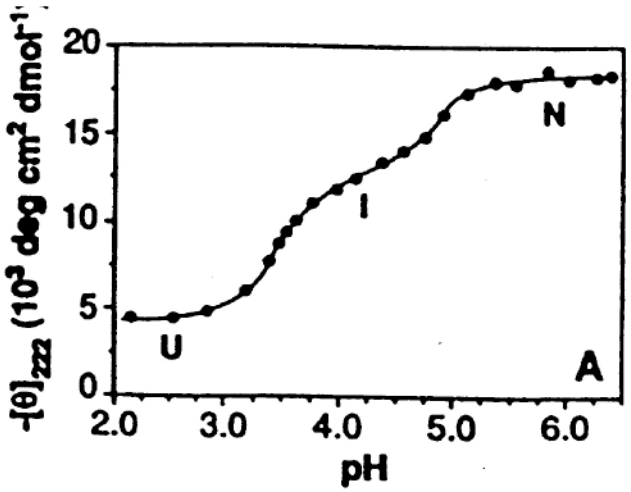
- ion pairs are mostly on the surface of the protein
- the interactions are short range due to **screening**
- disruption of water shell reduces stability contribution

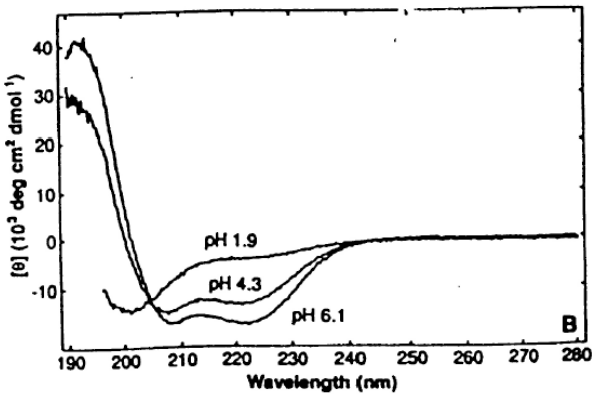
Potential from the ion-pair interaction results in a **shift in the pKa** of the involved residues:

$$\Delta G^0 = -2.303\, RT\, \Delta\mathrm{pK}$$

ΔpK = pK(D) – pK(N) for acidic side chains
ΔpK = pK(N) – pK(D) for basic side chains

pH denaturation of proteins





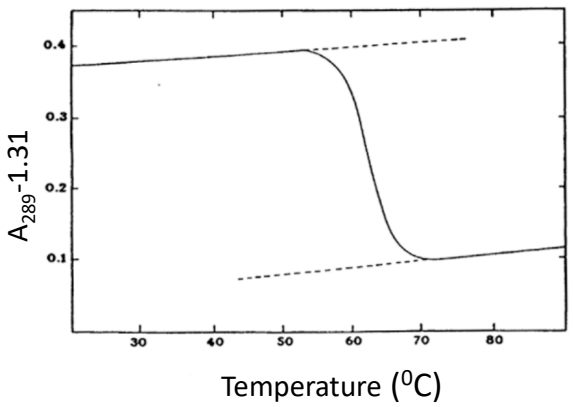
pK values of specific groups

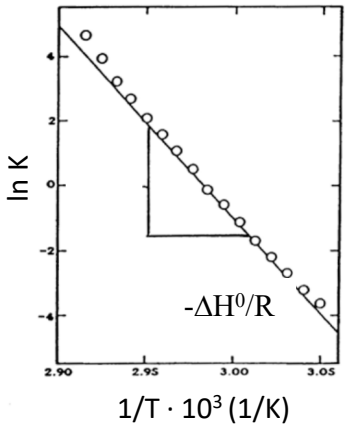
	pK'_1 α-COOH	pK'_2 α-NH_3^+	pK'_R R-Gruppe
Glycin	2,34	9,6	
Alanin	2,34	9,69	
Leucin	2,36	9,60	
Serin	2,21	9,15	
Threonin	2,63	10,43	
Glutamin	2,17	9,13	
Asparaginsäure	2,09	9,82	3,86
Glutaminsäure	2,19	9,67	4,25
Histidin	1,82	9,17	6,0
Cystein	1,71	10,78	8,33
Tyrosin	2,20	9,11	10,07
Lysin	2,18	8,95	10,53
Arginin	2,17	9,04	12,48

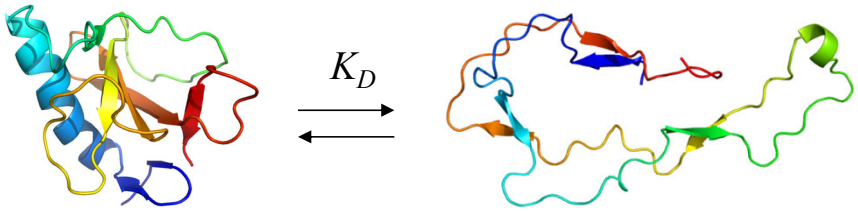
pH denaturation of apo-myoglobin

Thermal denaturation

Ribonuclease T1







N *D*

the T-dependence of K is given by:

$$\ln K(T) = -\frac{\Delta H^0}{RT} + \frac{\Delta S^0}{R}$$

derivative for T:

$$\frac{d \ln K}{dT} = \frac{\Delta H^0_{vH}}{RT^2}$$

$$\frac{d \ln K}{d1/T} = -\frac{\Delta H^0_{vH}}{R}$$

van't Hoff equation

Thermodynamic parameters of protein denaturation

Enthalpy:

Determined from slope of Van't Hoff plot

$$\frac{d \ln K}{d1/T} = -\frac{\Delta H^0_{vH}}{R}$$

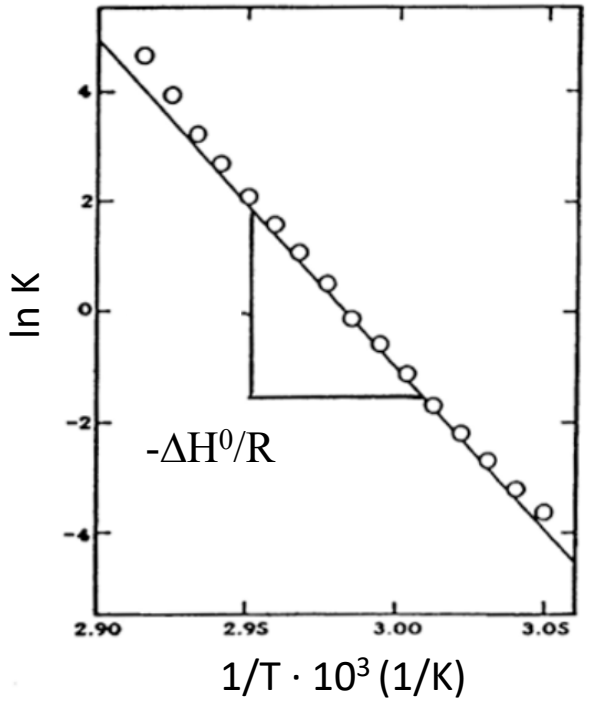
within window of linear relationship

Entropy:

At T_m, the temperature of the mid-point of the transition ($\Delta G^0 = 0$)

$$\Delta G^0 = \Delta H^0 - T\Delta S^0$$

$$\Delta S^0 = \Delta H^0 / T_m \quad \text{(for } \Delta G^0 = 0\text{)}$$



Quiz 2: Thermal transitions

- Another small protein is 99% folded at 328K and 1 % folded at 340 K
- what is the standard enthalpy (ΔH^0) of its folding transition?
- also estimate the entropy of the transition (ΔS^0)!

Heat capacity changes in protein denaturation

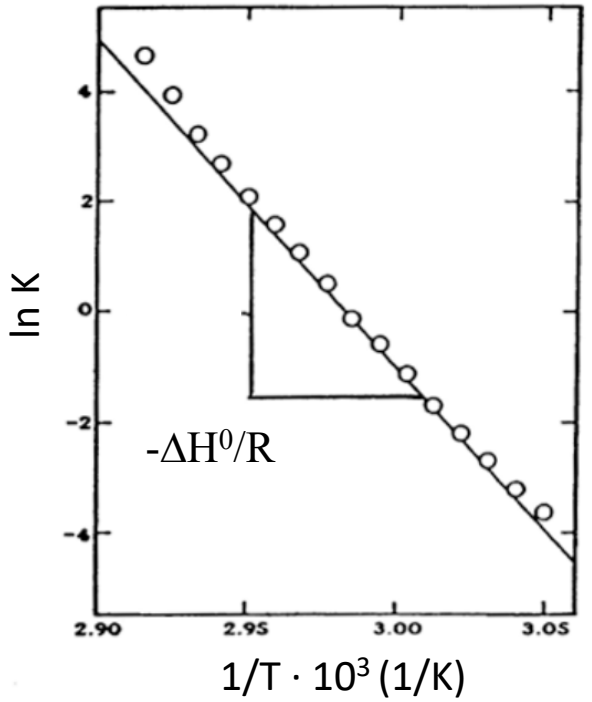
Heat capacity: $$C_P = \left(\frac{\partial H}{\partial T}\right)_P = \left(\frac{\partial Q}{\partial T}\right)_P$$

Change of internal energy upon heating, property of each global state

$$\Delta C_P^0 = C_{P,unfolded}^0 - C_{P,folded}^0$$

Curvature in Van't Hoff relations: The native state and the denatured state have different heat capacities.

Protein unfolding heat capacity is large and positive



$$\Delta C_p^{\,0} > 0$$

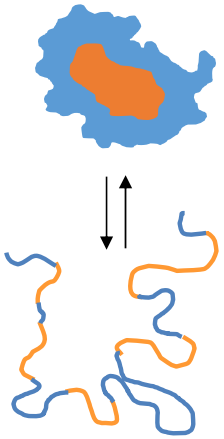
Heat capacity of protein denaturation

Definition of heat capacity:

$$\Delta C_p{}^0 = \frac{d\Delta H^0}{dT}$$

$$\Delta C_p{}^0 / T = \frac{d\Delta S^0}{dT}$$

both, ΔH^0 and ΔS^0 are temperature dependent



The molecular origin of $\Delta C_p{}^0 > 0$:

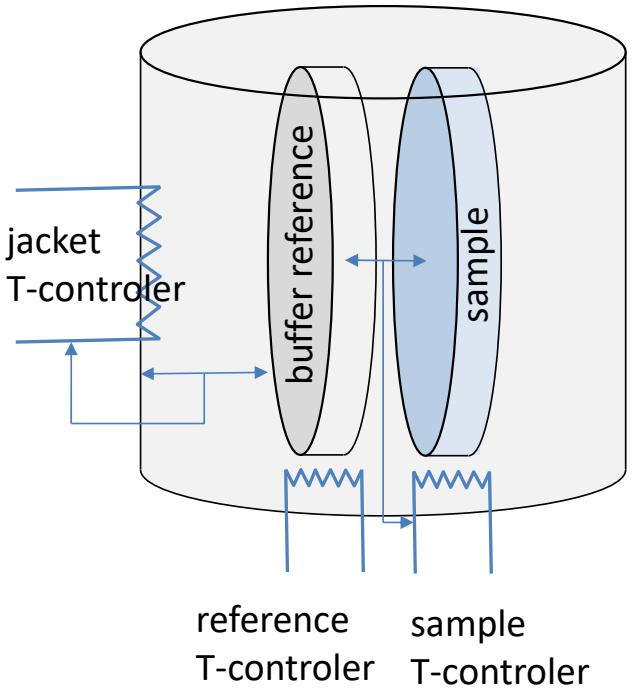
The native and the denatured state exhibit differences in **solvation**

in D, hydrophobic residues are exposed. The **water structure** leads to high heat capacity (compare to Lecture 1)

$\Delta C_p{}^0$ is usually 40-80 J mol^{-1} K^{-1} per residue

related to m-value (also dependent on ASA)

Measuring thermodynamic parameters: Differential Scanning Calorimetry (DSC)



Measures **amount of heat** required to change the **temperature** in the sample vs. the reference.

For an adiabatic isolated microcalorimeter, no heat exchange takes place with the environment

$$\delta Q = 0$$

according to first law

$$dU = dW$$

changes in internal energy solely depend on work within instrument

The partial molar heat capacity Cp

The heat capacity is defined as the **amount of heat** (Q) required to **change the temperature** by dT with constant pressure.

$$C_P = \left(\frac{\partial Q}{\partial T}\right)_P = \left(\frac{\partial H}{\partial T}\right)_P = \frac{dH}{dT}$$

Heat capacity of a protein solution $C_{P,sol}$ is a composite of **partial molar heat capacity** terms:

- Heat capacity of the solvent $C_{P,1}$
- Heat capacity of the protein (e.g. non-covalent interactions) $C_{P,2}$

From $C_{P,2}$ we can determine the enthalpy of protein folding/unfolding (n denotes the molar amounts)

$$C_{P2} = \left(\frac{\partial C_{PSol}}{\partial n_2}\right)_{T,p,n_i \neq 2}$$

Using calorimetry to determine protein stability

Problem: $C_{P,2}$ cannot be directly measured

From calorimetry, the **apparent molar heat capacity** ($C_{P,app}$) is obtained

this is the **difference** between the heat capacity of the solution ($C_{P,sol}$) and of the solvent ($C_{P,1}$)

$$C_{Papp} = \frac{C_{PSol} - n_1 C_{P1}}{n_2}$$

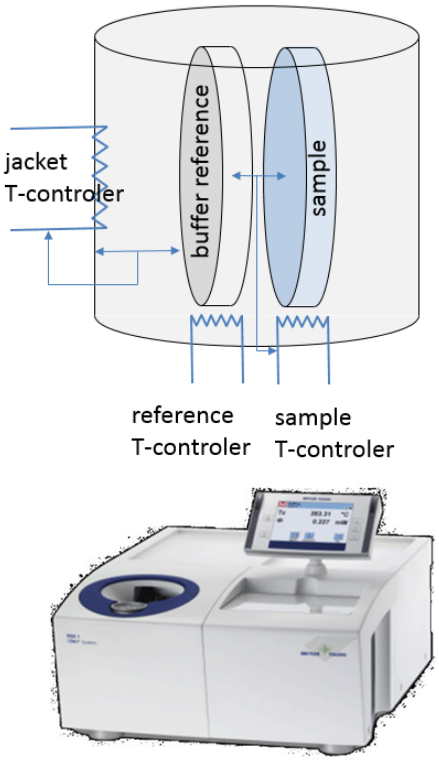
with: $$C_{P2} = \left(\frac{\partial C_{PSol}}{\partial n_2} \right)_{T,p,n_i \neq 2}$$

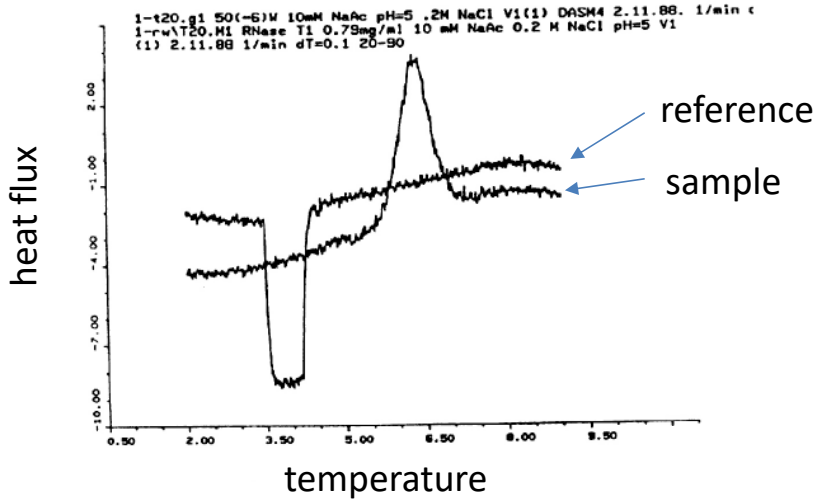
$$C_{P2} = C_{Papp} + \underbrace{n_2 \left(\frac{\partial C_{Papp}}{\partial n_2} \right)_{T,p,n_i \neq 2}}$$

this term can usually be neglected for employed protein concentrations

$$\boxed{C_{P2} = C_{P1} - C_{Papp}}$$

Determining heat capacity from calorimetry





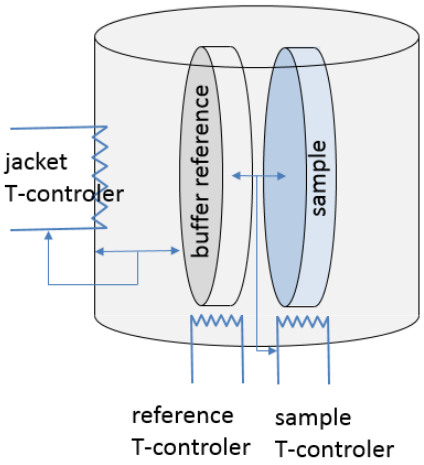
offset correction between sample and reference

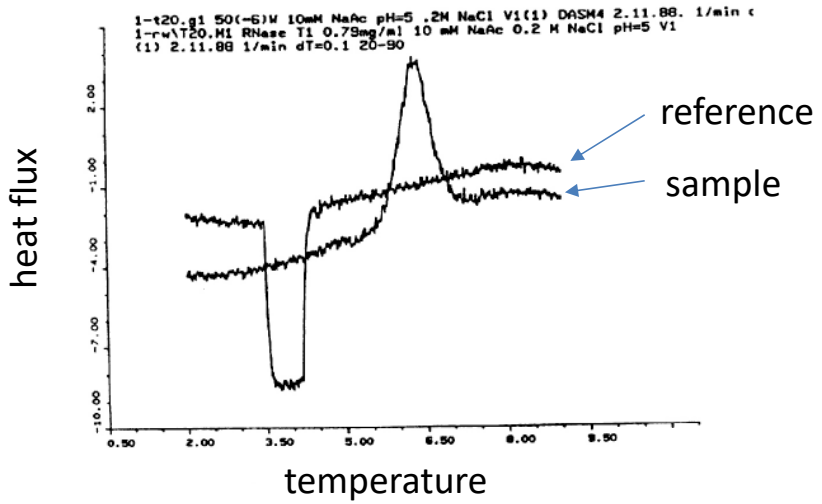
from normalized and calibrated difference follows the **heat capacity change of protein unfolding**

$$C_{P2} = C_{P1}\frac{v_P}{v_1} - \frac{Ck}{m_p}h$$

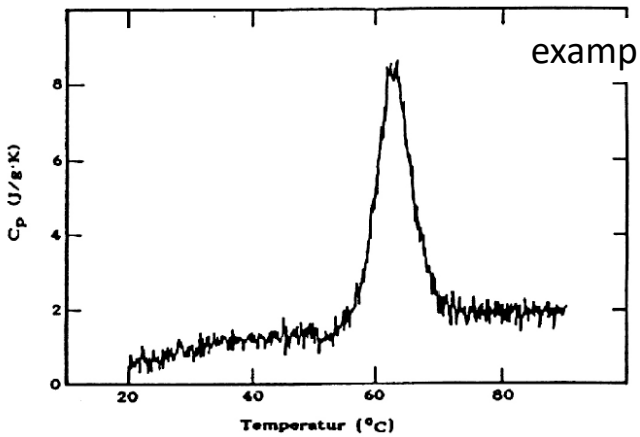
v_P : partial specific volume of protein
v_1 : partial specific volume of solvent
h: normalization
k: calibration constant

Determining heat capacity from calorimetry

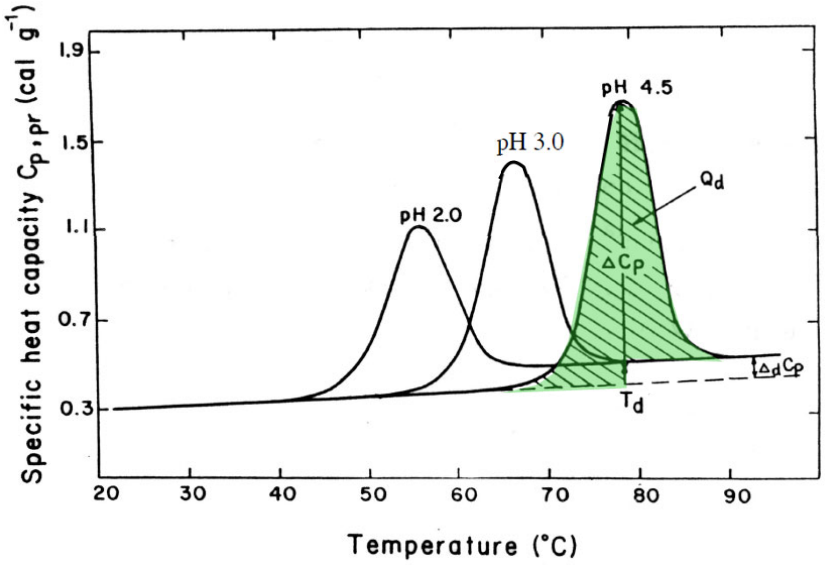




example: RNase T1



Enthalpy of conversion in protein denaturation



$\Delta H_m(cal)$ of lysozyme denaturation as a function of pH

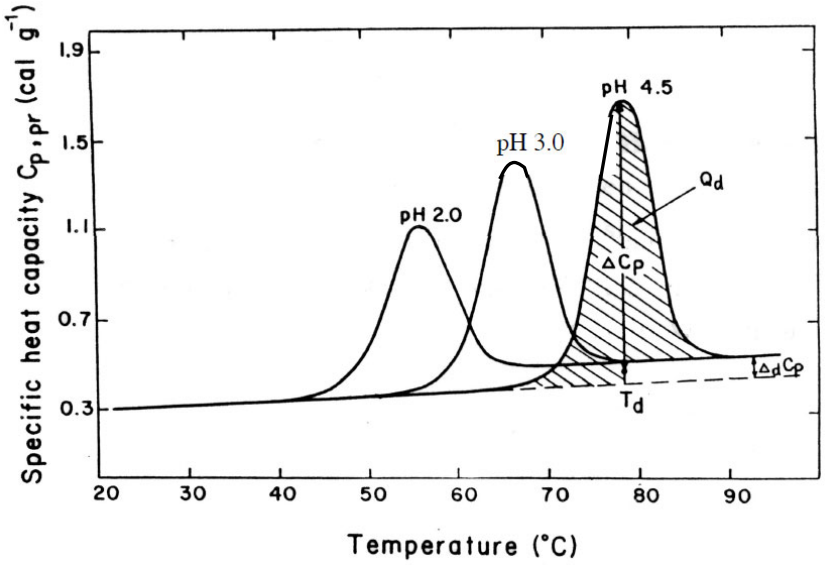
From **integration of the curve** the enthalpy of conversion (folded to unfolded) $\Delta H_m(cal)$ is determined

$$C_P = \frac{dH}{dT}$$

$$Q_d = \Delta H_m(cal) = \int C_p dT$$

model-free determination of enthalpy of state transition

Two-state folding or multistate transition?



$\Delta H_m(cal)$ of lysozyme denaturation as a function of pH

comparing the areas under the curve, an equilibrium constant can be determined.

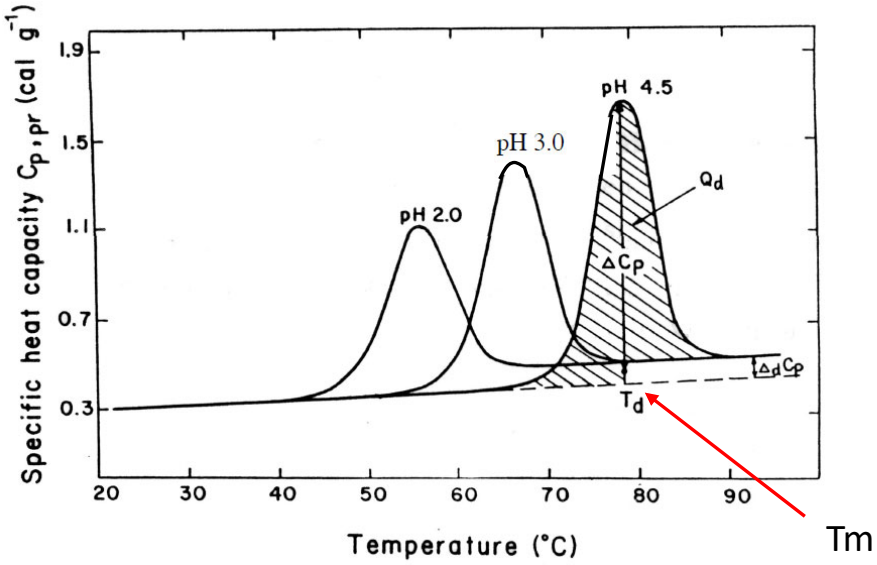
From K, Van't Hoff Enthalpy can be determined immediately, using:

$$\left(\frac{d \ln K}{dT}\right)_p = \frac{\Delta H^0_{v.H.}}{RT^2}$$

This assumes a **two-state transition**

Only in this case, $\Delta H_m(cal) = \Delta H^0_{vH}$

Entropy of conversion



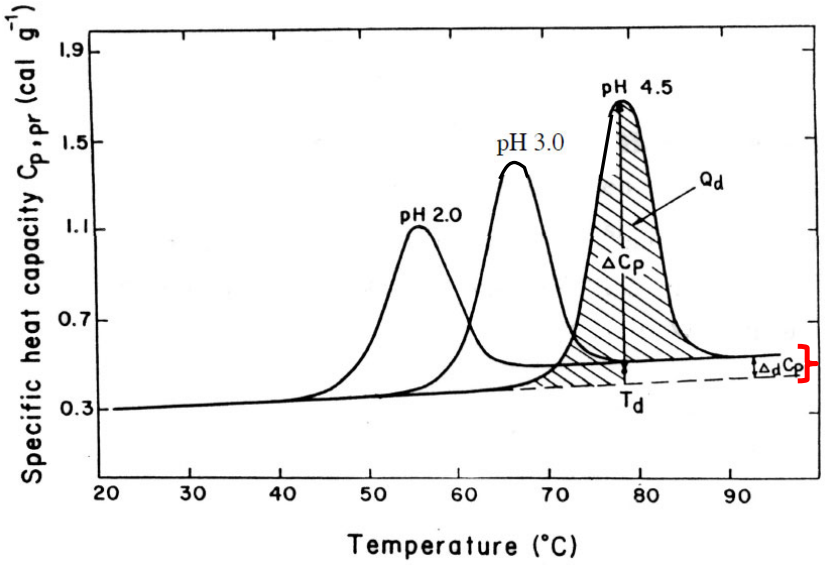
$\Delta H_m(cal)$ of lysozyme denaturation as a function of pH

At the midpoint of the transition (T_m), the entropy can be determined

$$\Delta G = \Delta H - T\Delta S = 0$$

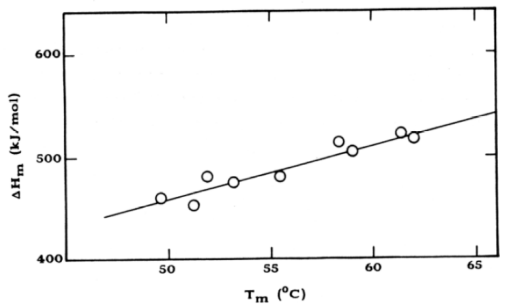
$$\Delta S_m = \frac{\Delta H_m}{T_m}$$

Heat capacity difference of unfolding



ΔC_p can directly be obtained from the measured curve

Can also be determined from the temperature dependence of ΔH_m, e.g. measured under different pH



Rnase T1 denaturation data

The protein stability curve

With heat capacity:

$$\frac{d\Delta H}{dT} = \Delta C_P \qquad \Delta H(T) = \Delta H(T_m) + \Delta C_P (T - T_m)$$

$$\frac{d\Delta S}{dT} T = \Delta C_P \qquad \Delta S(T) = \Delta S(T_m) + \Delta C_P \ln\left(\frac{T}{T_m}\right)$$

With heat capacity, the T-dependence of ΔG^0 becomes:

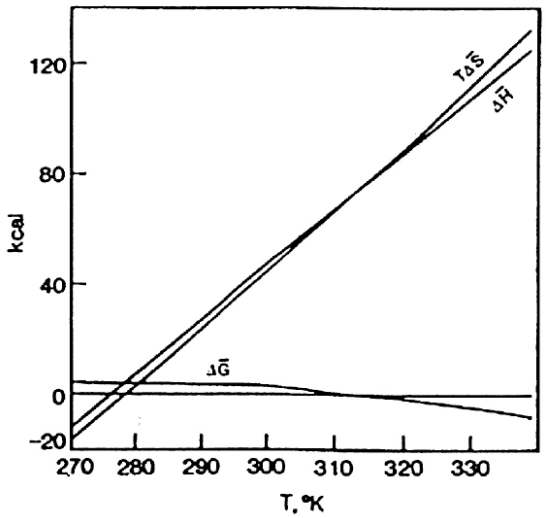
$$\Delta G^0(T) = \Delta H^0(T_m) - \mathrm{T}\,\Delta S^0(T_m) + \Delta C_p{}^0 \left(T - T_m - T \ln\frac{T}{T_m}\right)$$

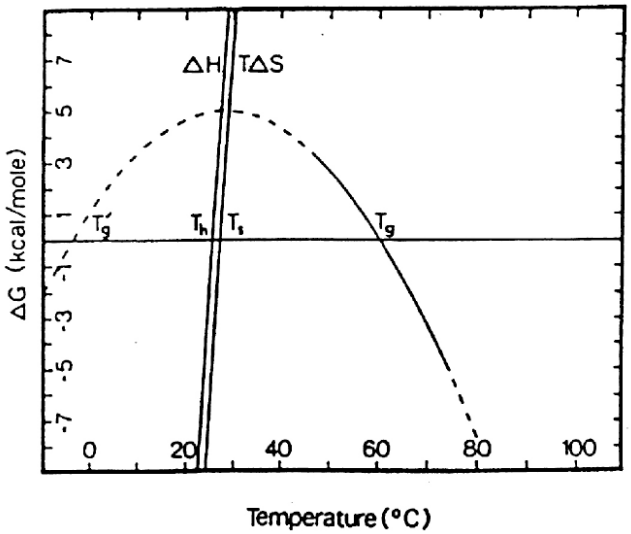
The protein stability curve

With heat capacity, the T-dependence of ΔG^0 becomes:

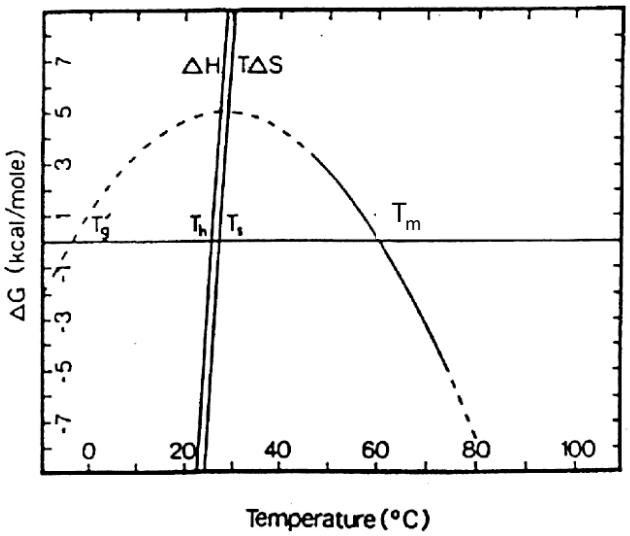
$$\Delta G^0(T) = \Delta H^0(T_m) - \mathrm{T}\,\Delta S^0(T_m) + \Delta C_p{}^0 \left(T - T_m - T \ln \frac{T}{T_m} \right)$$

Example





The protein stability curve



Protein stability is maximal at $\Delta S^0 = 0\ (T_S)$

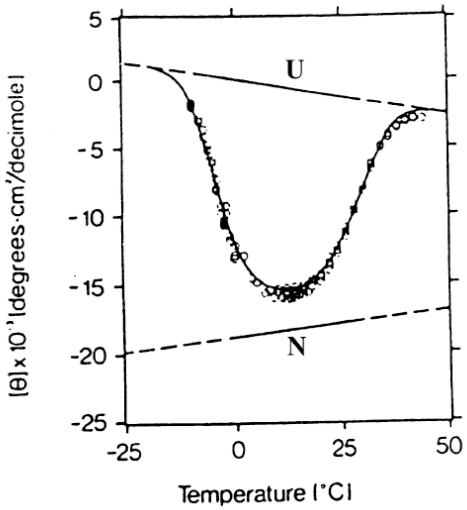
K_D (equilibrium constant) is maximal at $\Delta H^0 = 0\ (T_H)$

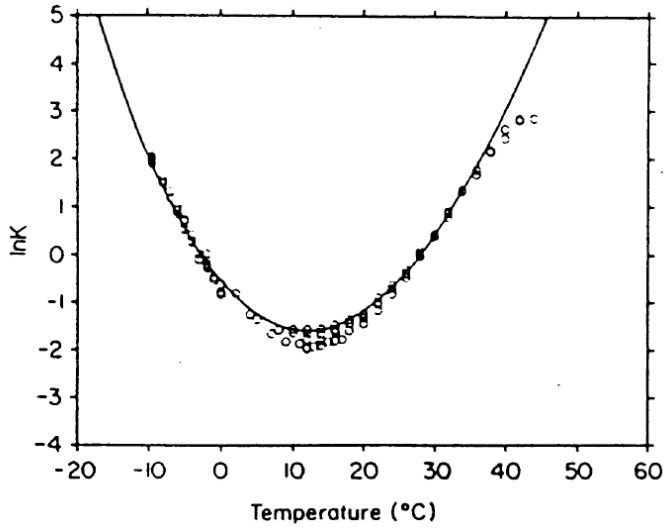
The transition mid-point (T_m) is at $\Delta G^0 = 0$ and $f_N = f_U = 0.5$

Protein cold denaturation

Protein stability curves show second Tm' at low temperature! → cold denaturation

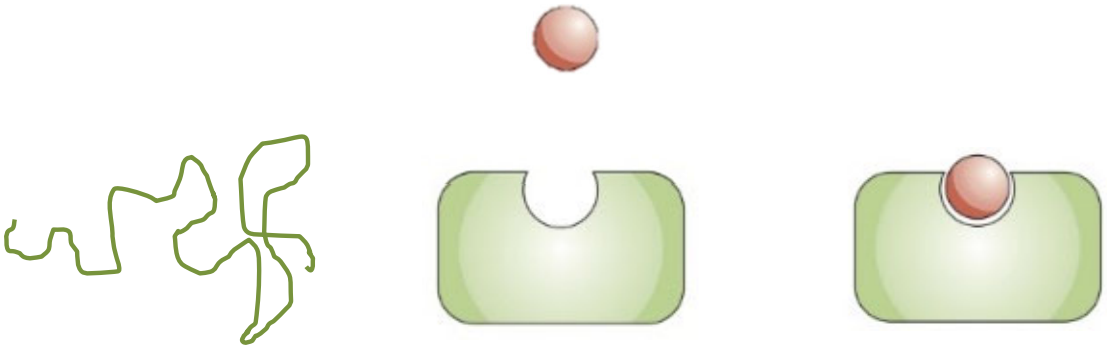
Can be observed for some proteins / mutants:
Cold denaturation of lysozyme (destabilized mutant)






Icefish

Proteins are stabilized by ligand binding

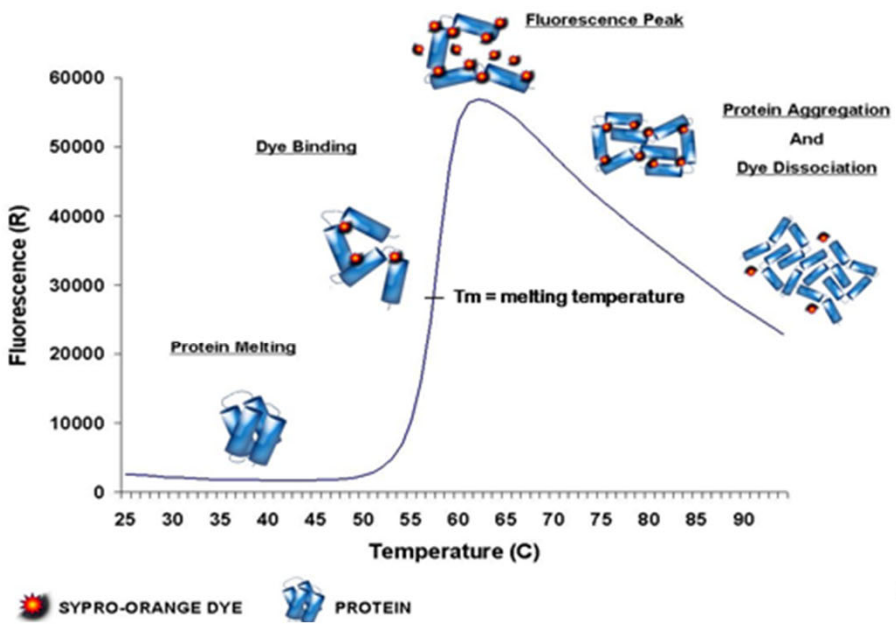


$$R_D \rightleftharpoons R_N + L \rightleftharpoons R_N : L$$

Protein stabilization by
free energy of ligands

→ measuring protein stability as a screening tool

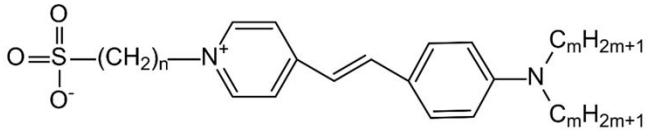
Differential Scanning Fluorimetry: A high-throughput assay to determine protein stability



source: Wikipedia

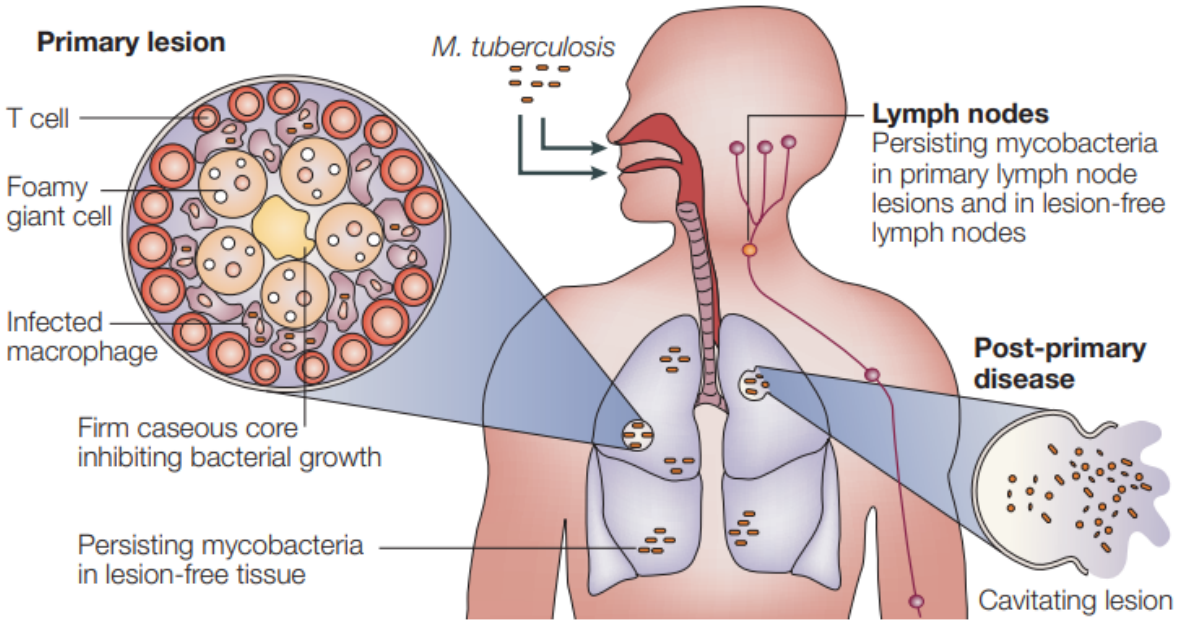
Differential Scanning Fluorimetry: Thermal shift assay

- Addition of a fluorophore
- binding to exposed hydrophobic sites: increase in fluorescence
- dyes: SYPRO orange
- High-throughput method
- ligand binding shifts Tm
- DSC: binding thermodynamics



SYPRO Orange

Case study: protein stability and pharmaceutical research for Tuberculosis



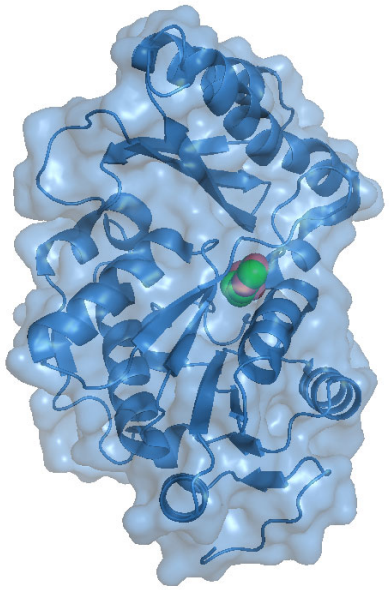
Nat Rev Microbiology 2003

Screening for pantothenate synthetase inhibitors

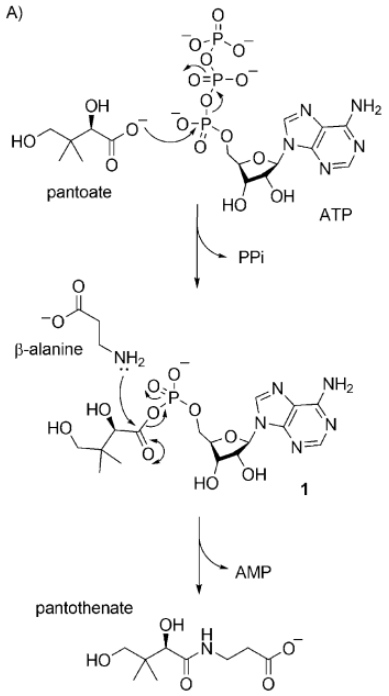
Pantothenate (vitamin B5)

- essential precursor to coenzyme A
- all enzymes absent in mammals
- de novo synthesis important for bacteria, including **M. tuberculosis**
- mutant can get pantotheate through salvage pathway, but no disease

→ **pantotheate synthase is a good target**

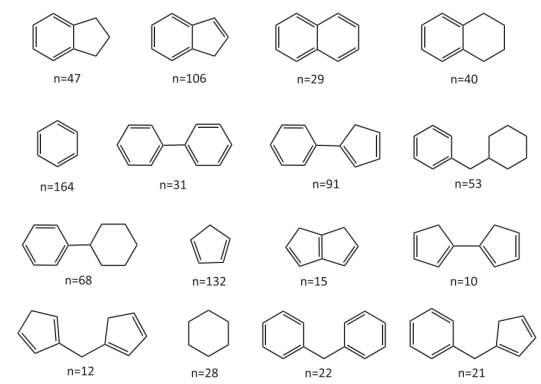


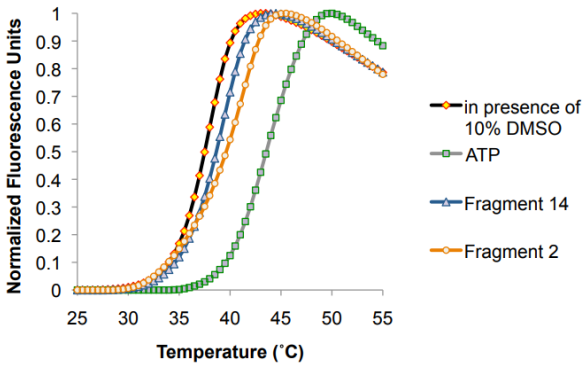
pantothenate synthetase: new target



Ciulli et al. ChemBioChem 2008

Fragment screen to identify binders





Silvestre et al. PNAS 2013

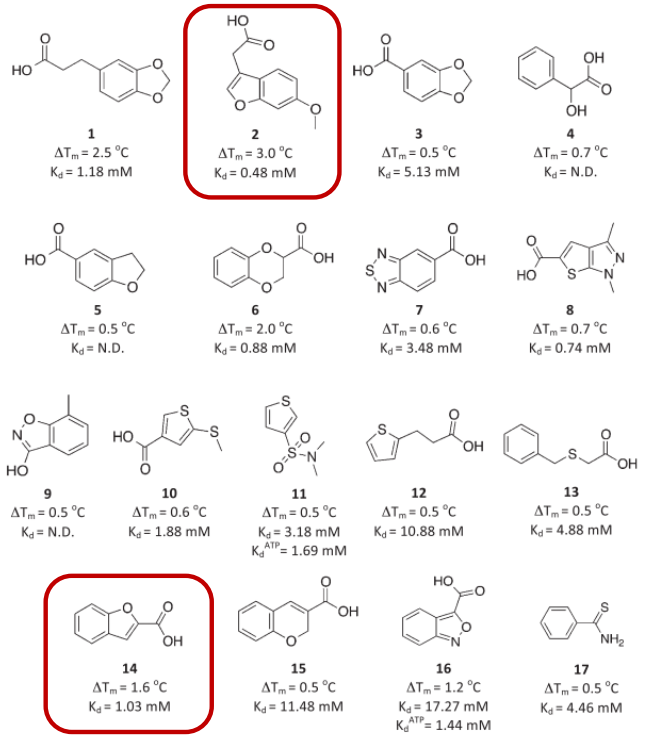


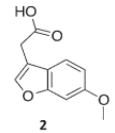
Fig. 1. The hits identified in the primary thermal shift screen and validated by secondary NMR spectroscopy screen, with K_d determined by ITC. The fragments exhibited a varied range in both ΔT_m (from 0.5 to 3.0 °C) and affinities (K_d from 480 μM to 17.3 mM).

Follow – up on hits to obtain thermodynamics of binding

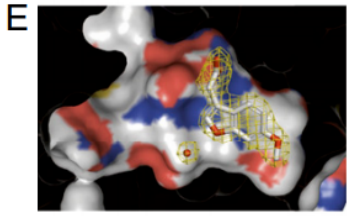
Table 1. Thermodynamic binding parameters determined by ITC

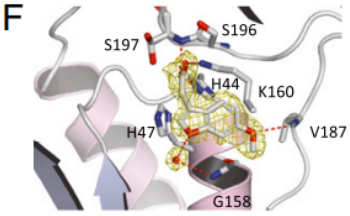
Fragment	ΔH, kcal/mol	ΔG, kcal/mol	K_d, mM	*LE*
1	−13.9 ± 0.1	−4.0 ± 0.1	1.2 ± 0.01	0.29
2	−10.1 ± 0.2	−4.5 ± 0.3	0.5 ± 0.01	0.30
3	−11.6 ± 0.2	−3.1 ± 0.3	5.1 ± 0.2	0.26
6 (racemate)	−10.3 ± 0.05	−4.2 ± 0.1	0.9 ± 0.04	0.32
6 (R-enantiomer)	−7.9 ± 0.1	−4.1 ± 0.1	0.9 ± 0.02	0.32
6 (S-enantiomer)	−12.3 ± 0.2	−4.3 ± 0.2	0.7 ± 0.02	0.33
7	−9.4 ± 0.1	−3.3 ± 0.1	3.5 ± 0.04	0.28
8	−7.51 ± 0.1	−4.2 ± 0.1	0.7 ± 0.02	0.32
10	−7.2 ± 0.1	−3.7 ± 0.1	1.9 ± 0.03	0.37
11	−16.2 ± 0.8	−3.4 ± 0.9	3.2 ± 0.2	0.31
11 (in the presence of ATP)	−1.5 ± 0.05	−3.8 ± 0.1	1.7 ± 0.08	0.34
12	−21.2 ± 1.0	−2.7 ± 1.0	10.9 ± 0.6	0.27
13	−10.7 ± 0.3	−3.1 ± 0.3	4.9 ± 0.2	0.26
14	−8.7 ± 0.2	−4.0 ± 0.2	1.0 ± 0.01	0.33
15	−12.3 ± 0.7	−2.3 ± 0.7	11.5 ± 0.8	0.18
16	−11.3 ± 0.02	−2.6 ± 0.1	17.3 ± 0.8	0.22
16 (in the presence of ATP)	−1.8 ± 0.4	−3.8 ± 0.4	1.4 ± 0.02	0.32
17	−5.3 ± 0.2	−3.1 ± 0.2	4.5 ± 0.2	0.34

binding mode of **2**

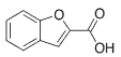


2
ΔT_m = 3.0 °C
K_d = 0.48 mM

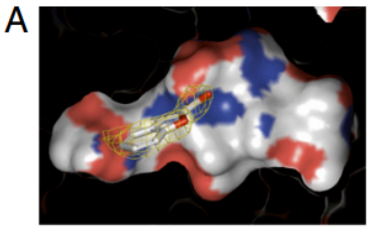


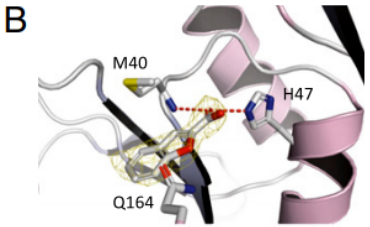


binding mode of **14**

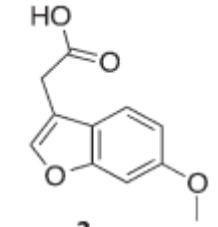


14
ΔT_m = 1.6 °C
K_d = 1.03 mM





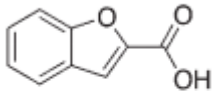
From fragments to inhibitors



2

$\Delta T_m = 3.0\ ^{\circ}C$

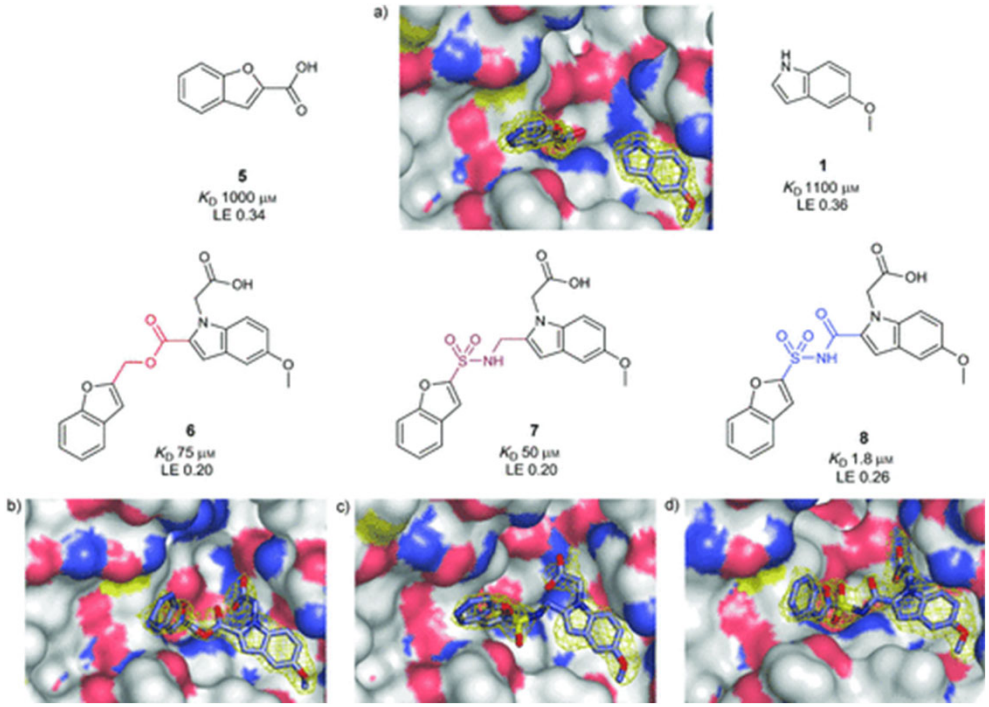
$K_d = 0.48$ mM



14

$\Delta T_m = 1.6\ ^{\circ}C$

$K_d = 1.03$ mM



A fragment-linking strategy designed for *M. tuberculosis* pantothenate synthetase

Hung et al. Angew Chem 2009