Experimental data: s-values & helical propensities

Table 2. *Helix propensity values measured at 273 K*

Residue	s-Value[a]	w-Value[a]	$-RT\ln(w)$ (kcal/mol)	$\Delta\Delta G^0$ (kcal/mol)
Ala	1.54	1.61	−0.258	−1.88
Arg$^+$	1.1[f]	1.2[f]	−0.047	−1.67
Leu	0.92	0.96	0.022	−1.60
Lys$^+$	0.78	0.82	0.108	−1.52
Glu^0	0.63[b]	0.66[b]	0.225	−1.40
Met	0.60	0.63	0.251	−1.37
Gln	0.53	0.56	0.314	−1.31
Glu$^-$	0.43[b]	0.45[b]	0.433	−1.20
Ile	0.42	0.44	0.445	−1.18
Tyr	0.37–0.50[c]	0.39–0.53[c]	0.344[c]–0.511	−1.28[c] to −1.11
His^0	0.36[d]	0.38[d]	0.525	−1.10
Ser	0.36	0.38	0.525	−1.10
Cys	0.33	0.35	0.570	−1.06
Asn	0.29	0.31	0.635	−1.00
Asp$^-$	0.29[e]	0.31[e]	0.635	−1.00
Asp^0	0.29[e]	0.31[e]	0.635	−1.00
Trp	0.29[c]–0.36	0.30[c]–0.38	0.525–0.653[c]	−1.10 to −0.97[c]
Phe	0.28	0.29	0.672	−0.95
Val	0.22	0.23	0.797	−0.83
Thr	0.13	0.14	1.07	−0.56
His$^+$	0.06[d]	0.06[d]	1.53	−0.10
Gly	0.05	0.05	1.62	0
Pro	≈0.001	≈0.001	≈4	>5

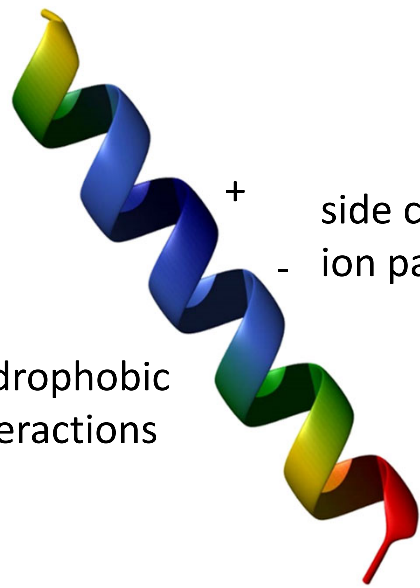
Chakrabartty, Kortemme, Baldwin, Protein Sci **1994**

Measured from fit of theory to host-guest measurements

allows now the **prediction** of the helicity of a given amino acid sequence

[a] Values obtained by applying Lifson-Roig theory modified to include either N-capping or charged group–helix macrodipole interactions. Conditions: 273 K, 1.0 M NaCl for uncharged residues and Lys, and 273 K, 10 mM NaCl for Arg, Asp, Glu, and His.
[b] Values from Scholtz et al. (1993).
[c] Values corrected for error in fraction helix measurement caused by aromatic contribution.
[d] Values from Armstrong and Baldwin (1993).
[e] Values from Huyghues-Despointes et al. (1993).
[f] Values from Huyghues-Despointes and Baldwin (unpubl.).

Side chain interactions: Stabilizing and destabilizing



\+

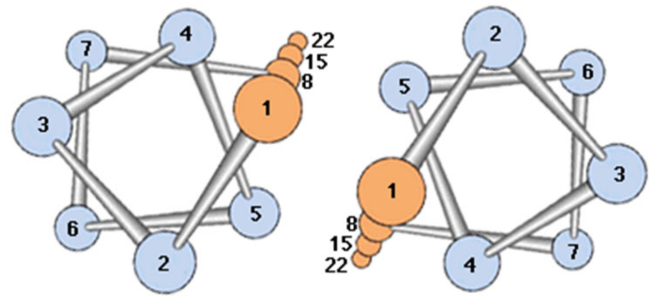
side chain

\-

ion pairs

hydrophobic
interactions

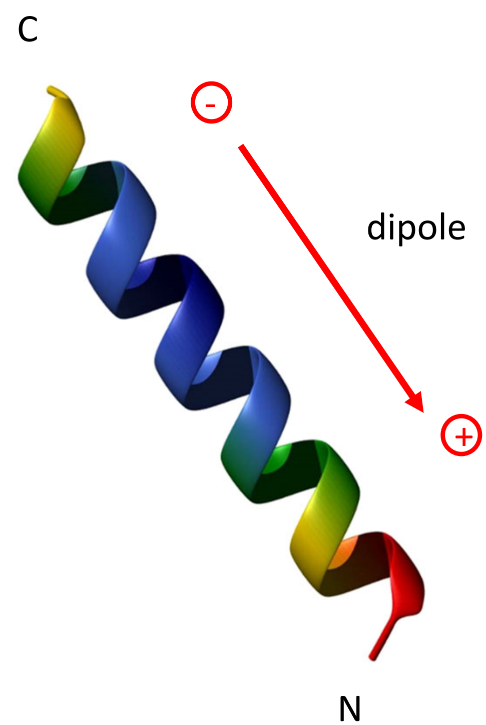
The helical wheel:



Mapping side chains on the same face of the helix

interactions possible
stabilizing or destabilizing

Interactions with the helix-dipole



Negative charges at the C-terminus

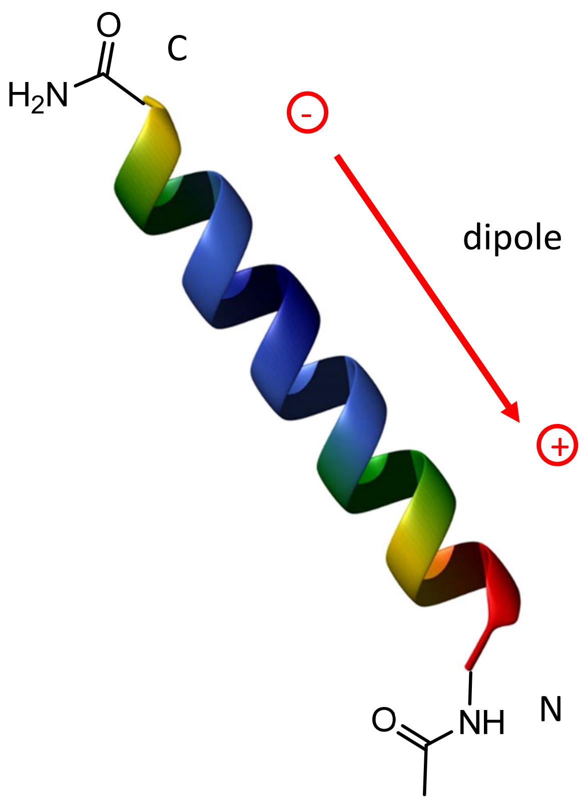
or

positive charges at the N-terminus

destabilize helices

Capping groups can overcome this effect

Capping groups in isolated helices



C-terminal **amide**

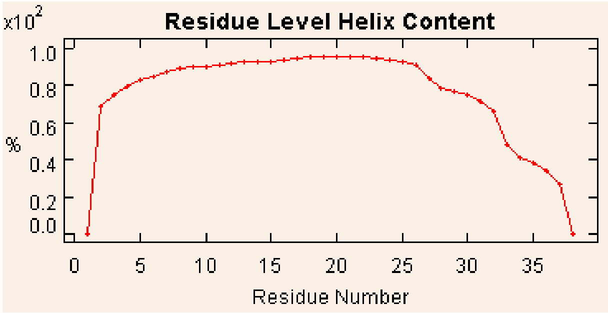
N-terminal **acetyl** (can also accept another H-bond)

Helix prediction: AGADIR

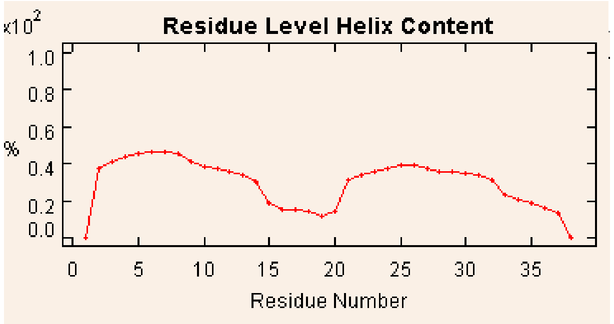
Secondary structure prediction based on helix-coil theory

takes into account further empirical parameters

Peptide 1
AAAAAAAQQAAAAQWAAA
AAAAAAAQQAAAAQWAAA



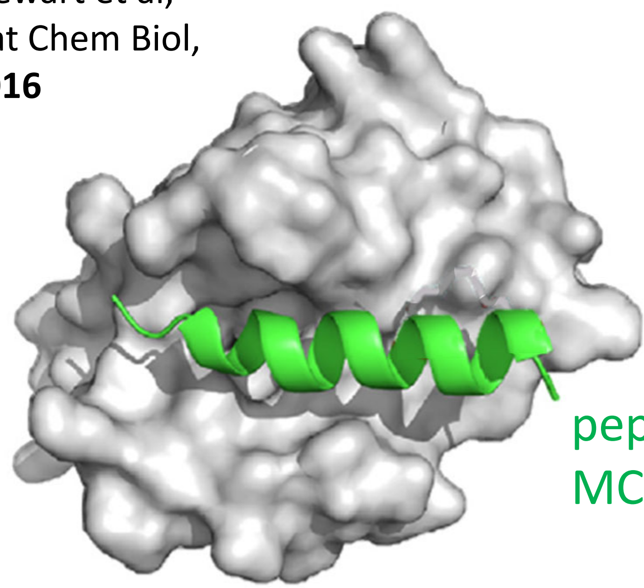
Peptide 1
AAAAAAAQQAAAAQWAGG
AAAAAAAQQAAAAQWAAA



Muñoz, V. & Serrano, L. (1994a), Nature: Struct. Biol. 1, 399-409

How to stabilize helix enough to make an efficient drug?

Stewart et al,
Nat Chem Biol,
2016



peptide from BCL-2, a MCL-1 binding partner

MCL-1: resistance factor in human cancer → prevents apoptosis of cancer cells

Targeting protein-protein interactions:

- very difficult with small molecules
- peptides are a possible solution

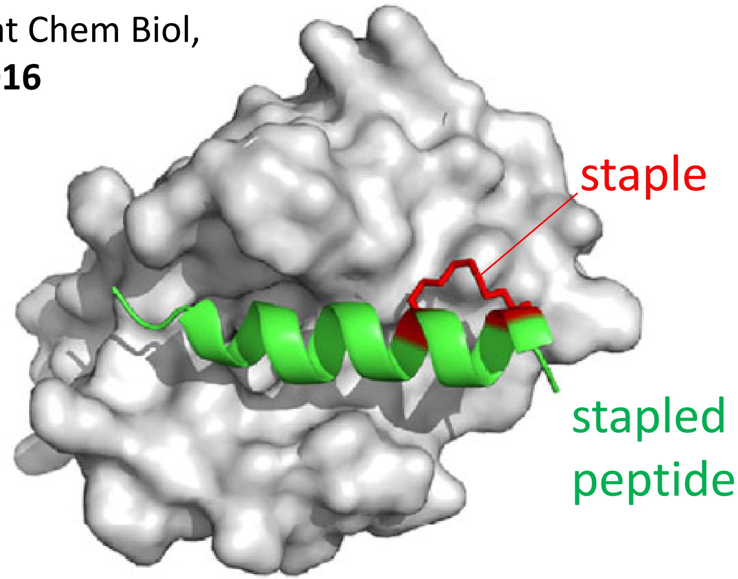
→ **Problems:**

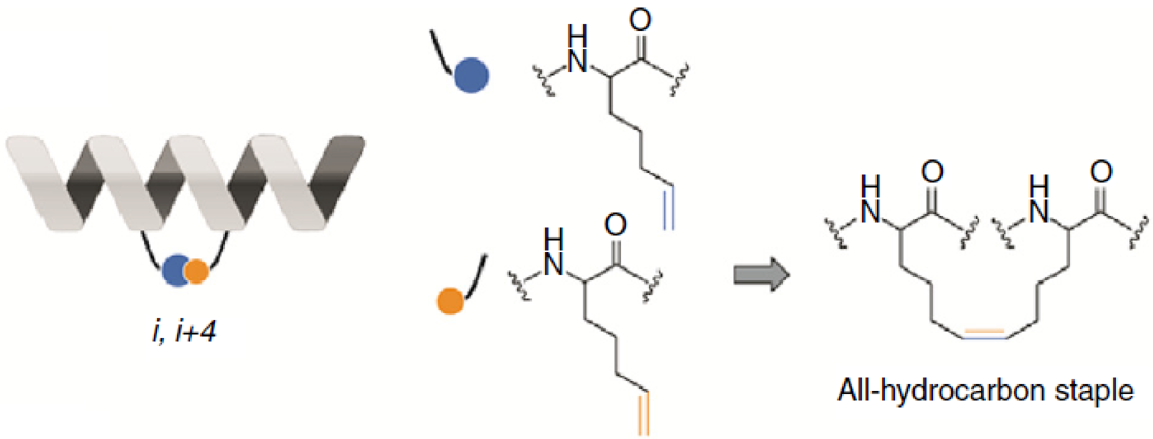
structural stability, cell permeability, stability to proteases

→ Solution: peptides with stabilized structure (a-helix)

Approach: 'Stapling' the peptides

Stewart et al,
Nat Chem Biol,
2016





Araghi & Keating, Curr Opin Struct Biol **2016**

MCL-1: resistance factor in human cancer → prevents apoptosis of cancer cells

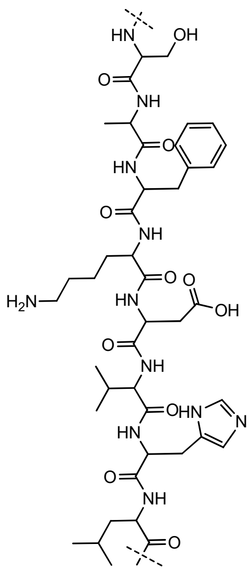
- synthesis via ring-closure metathesis
- strong increase in helicity → why?
- better cell permeability and serum stability

→ drug like properties possible

Summary

- Using simple assumptions and statistical thermodynamics the cooperative conformational change from coil to helix can be analyzed
- The cooperativity of the transition depends on the nucleation parameter σ
- Experiments produced values of the elongation parameter s for all amino acid residues, these serve as a predictor of helix stability
- Charges at helix termini stabilize the helix if they favorably interact with the helix dipole, otherwise the weaken the helix
- Capping motifs stabilize helices through hydrogen bonding with termini, optimal charges, conformations

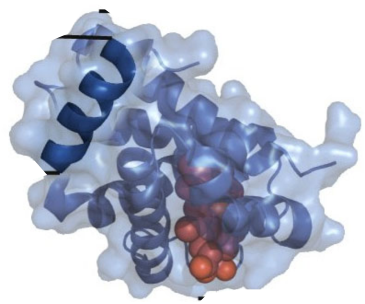
Proteins – Structural hierarchy



Amino acid chain
primary structure

α-helix
secondary
structure

transitions

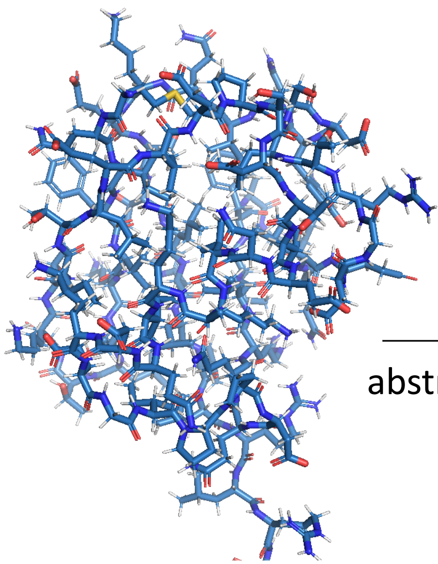


Protein domain
Tertiary structure

Protein complex
Quarternary structrue

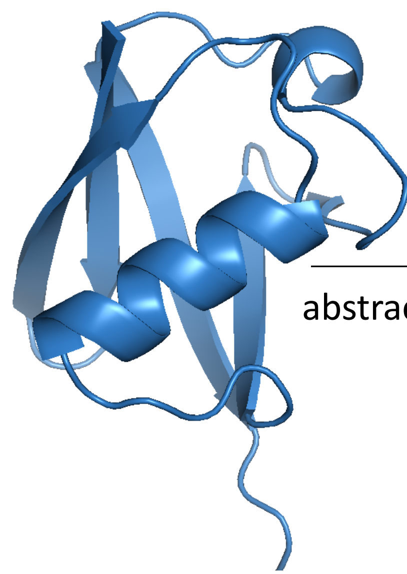
Protein organization in 3D

molecular representation



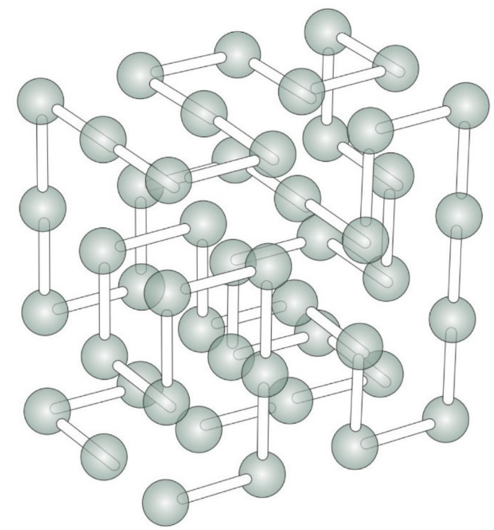
abstraction

cartoon
(secondary structure)



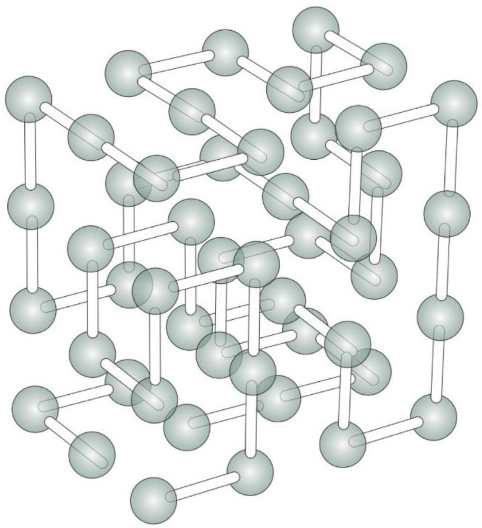
abstraction

lattice model

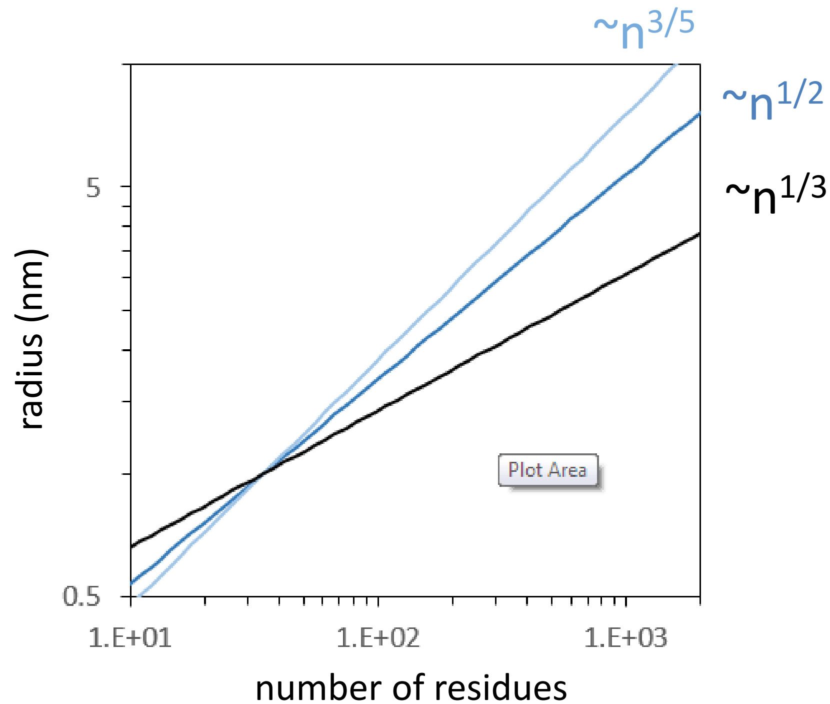


random walk on a 3D lattice

Scaling laws for protein size

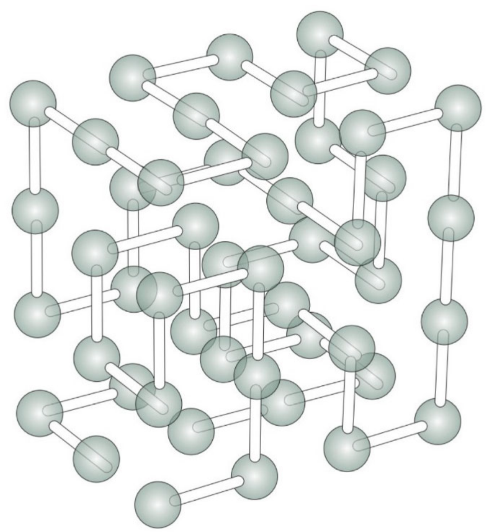


Scaling laws for a random walk
self-avoiding chain: $r \sim n^{3/5}$
random chain: $r \sim n^{1/2}$
compact chain: $r \sim n^{1/3}$

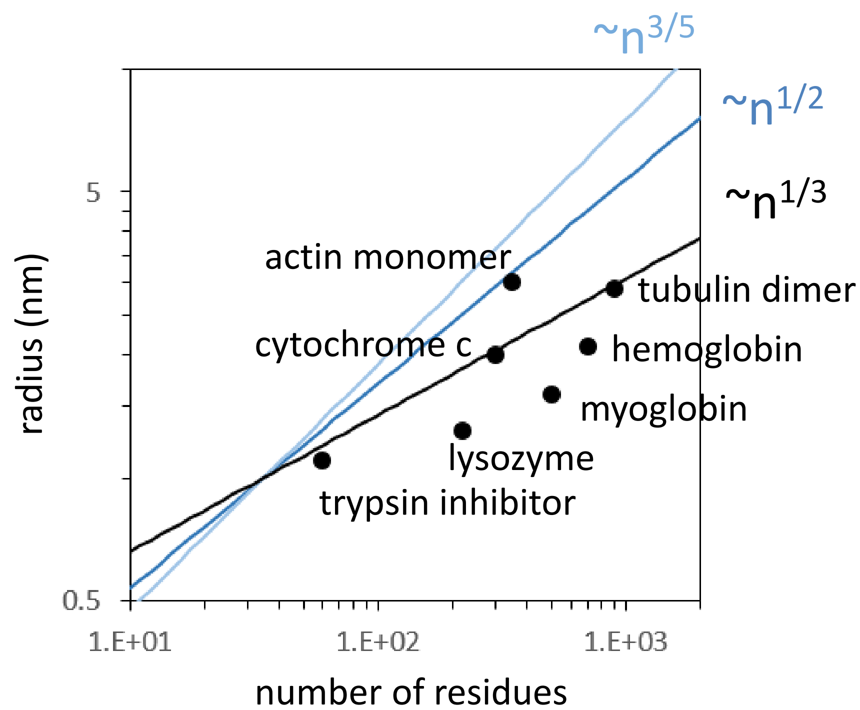


Adapted from Physical Biology of the Cell (Chapter 8)

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)

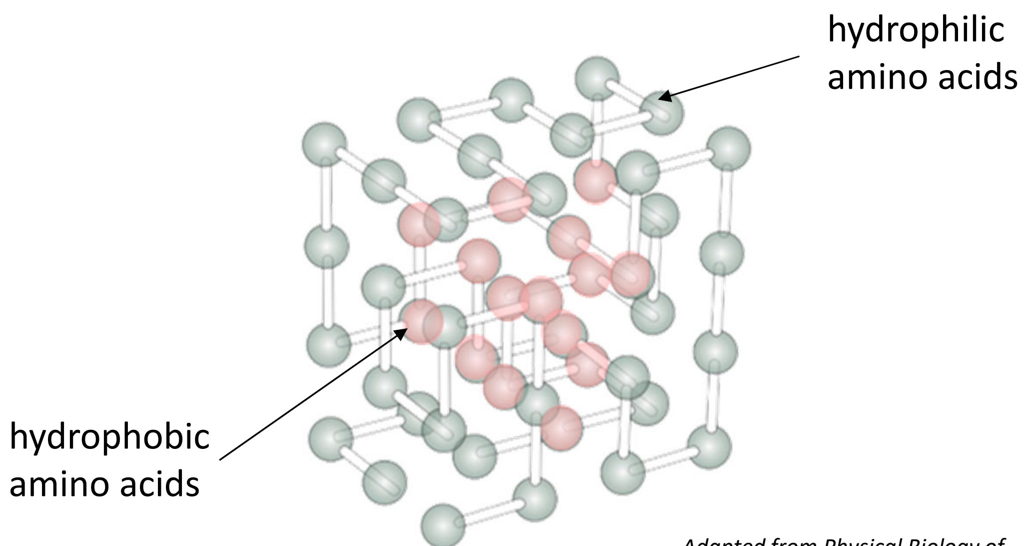
How does a protein find its native state?

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

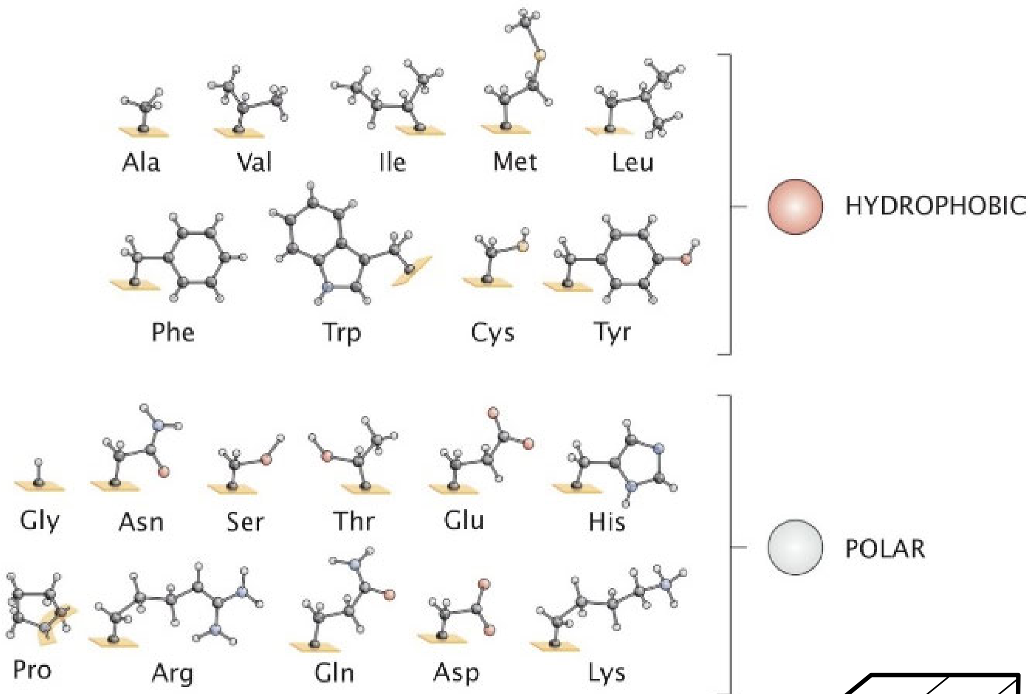
Mechanism:

Hydrophobic collapse

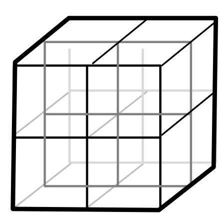


Adapted from Physical Biology of the Cell (Chapter 8)

HP-models of protein folding



Adapted from Physical Biology of the Cell (Chapter 8)



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

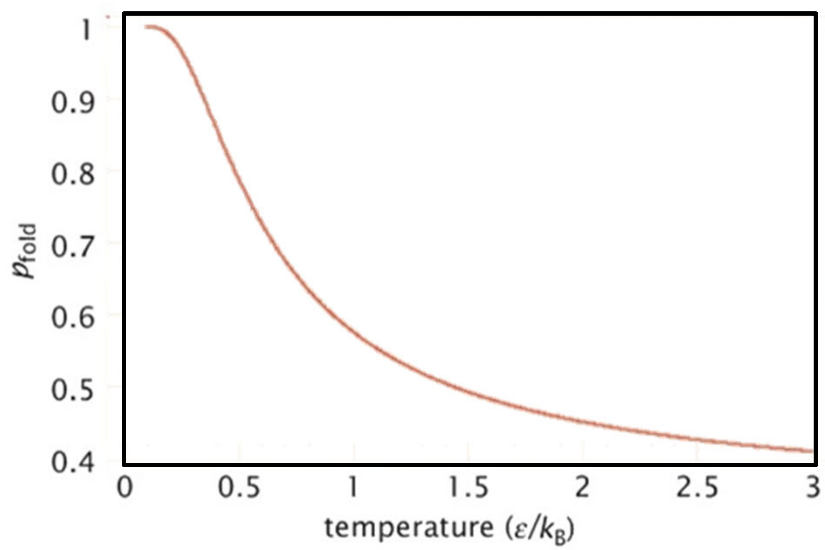
→ **2 letter alphabet allows an abstract model of structure formation**

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



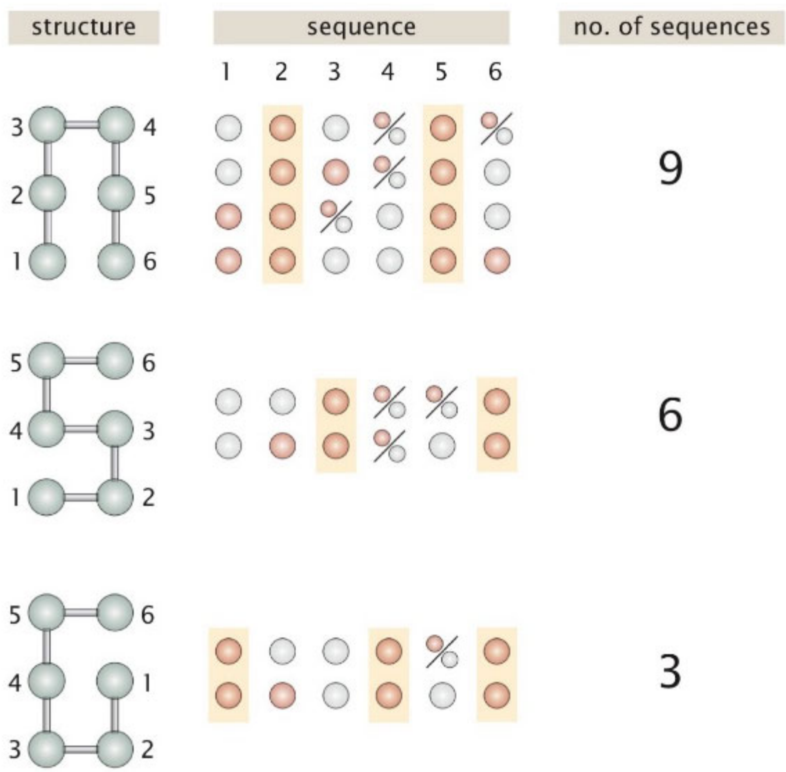
$$p_{fold} = \frac{e^{-2\beta\varepsilon}}{e^{-2\beta\varepsilon} + 2e^{-4\beta\varepsilon}} = \frac{e^{-2\beta\varepsilon}}{Q}$$

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

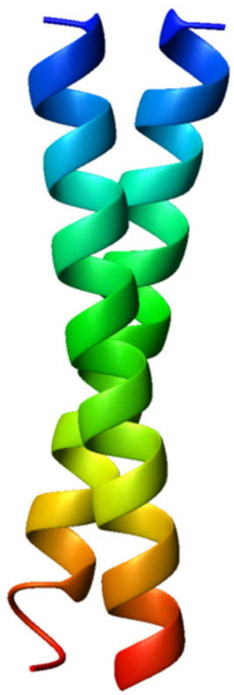
→ determine energy of each

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

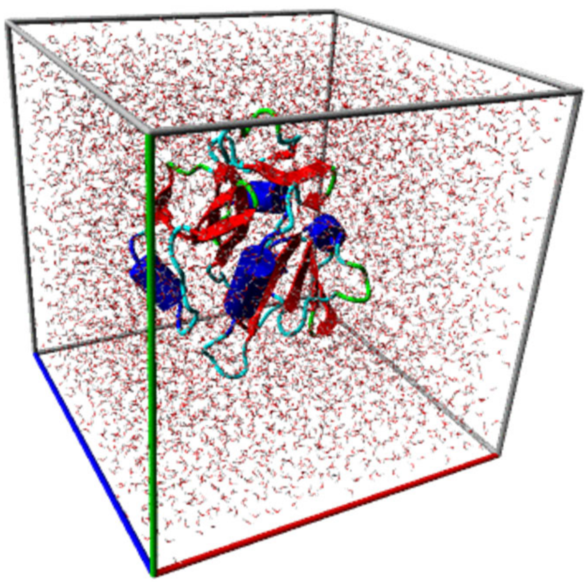
Hecht et al. Protein Sci 2004

Secondary structure
→ fixed, helical

Tertiary structure:
follows HP rules

Provides a platform, on which further functionality can be designed

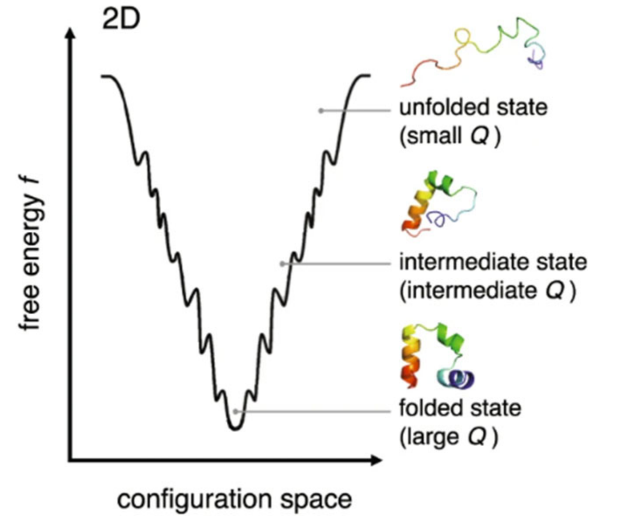
Force field for protein structure



https://in.explara.com/

Force field for protein structure evaluation:

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

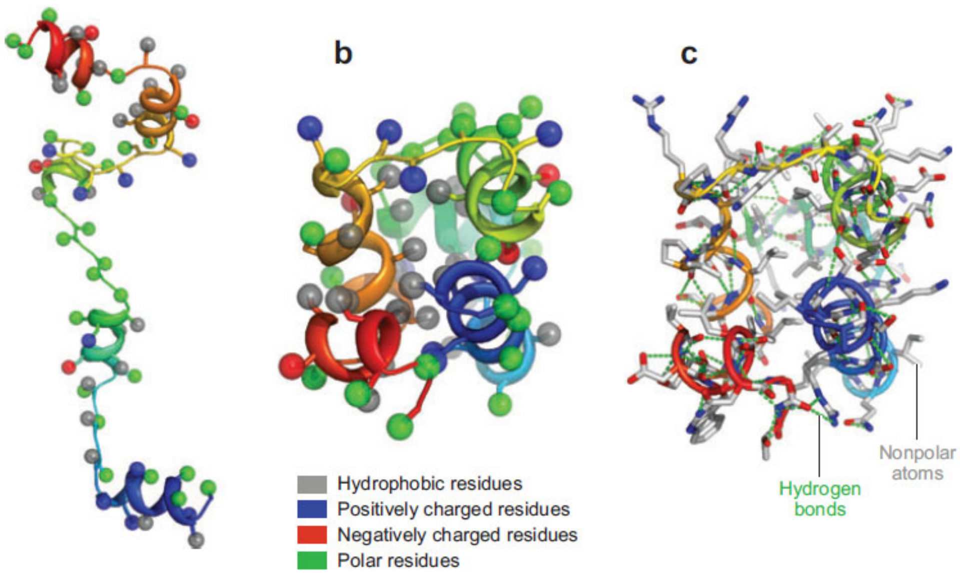


https://www.nature.com/articles/s41598-019-50825-6

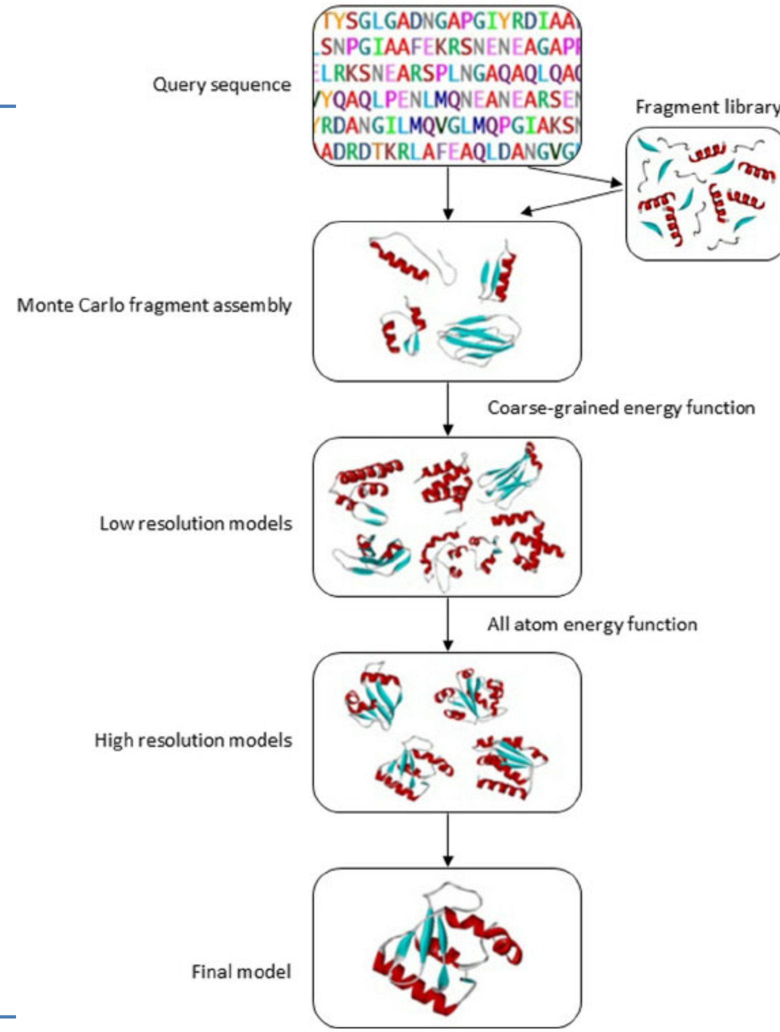
Goal: Minimize the configurational energy, find the global minimum

Protein structure prediction / design

Design from fragment libraries,
followed by refinement (Rosetta)

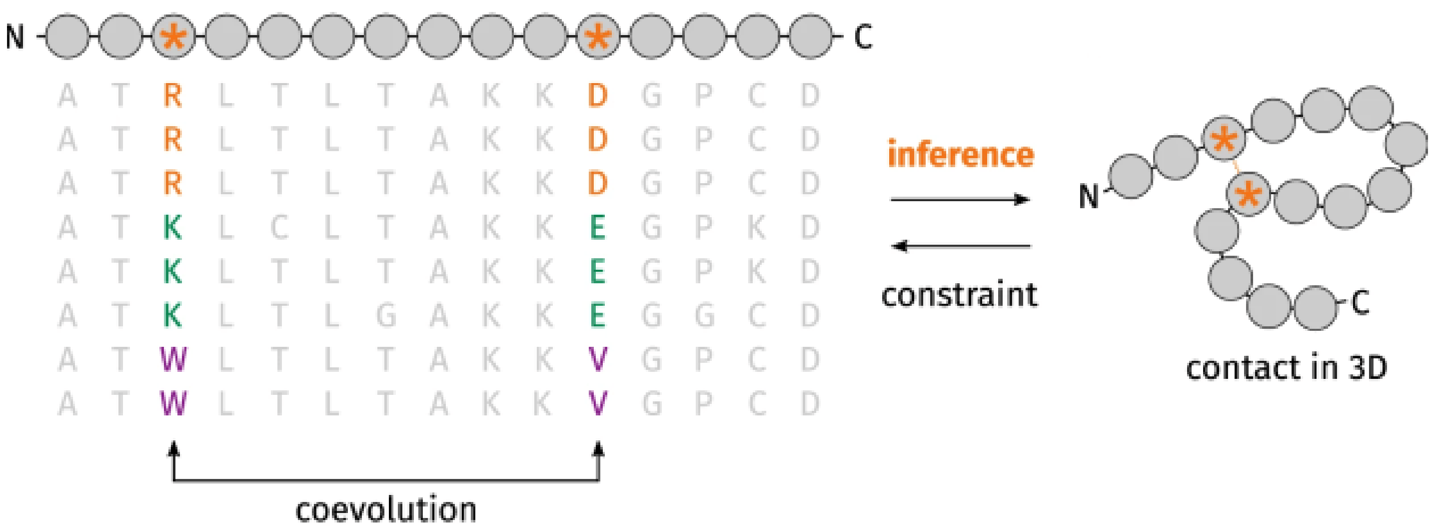


Das & Baker, Annu Rev Biochem 2008



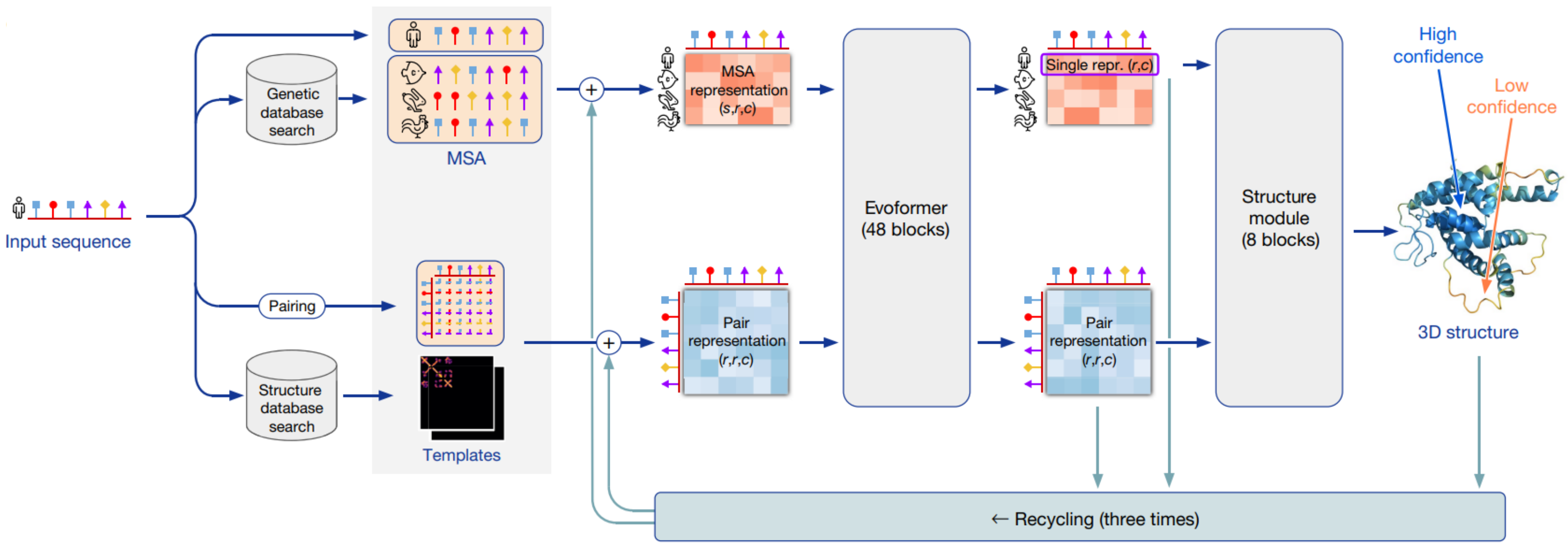
https://tbiomed.biomedcentral.com/articles/10.1186/s12976-015-0014-1

Structural information from sequence: Co-evolution analysis



[43146:42mxuqdo1srqh1335;:99]

Protein structure prediction – ML approaches



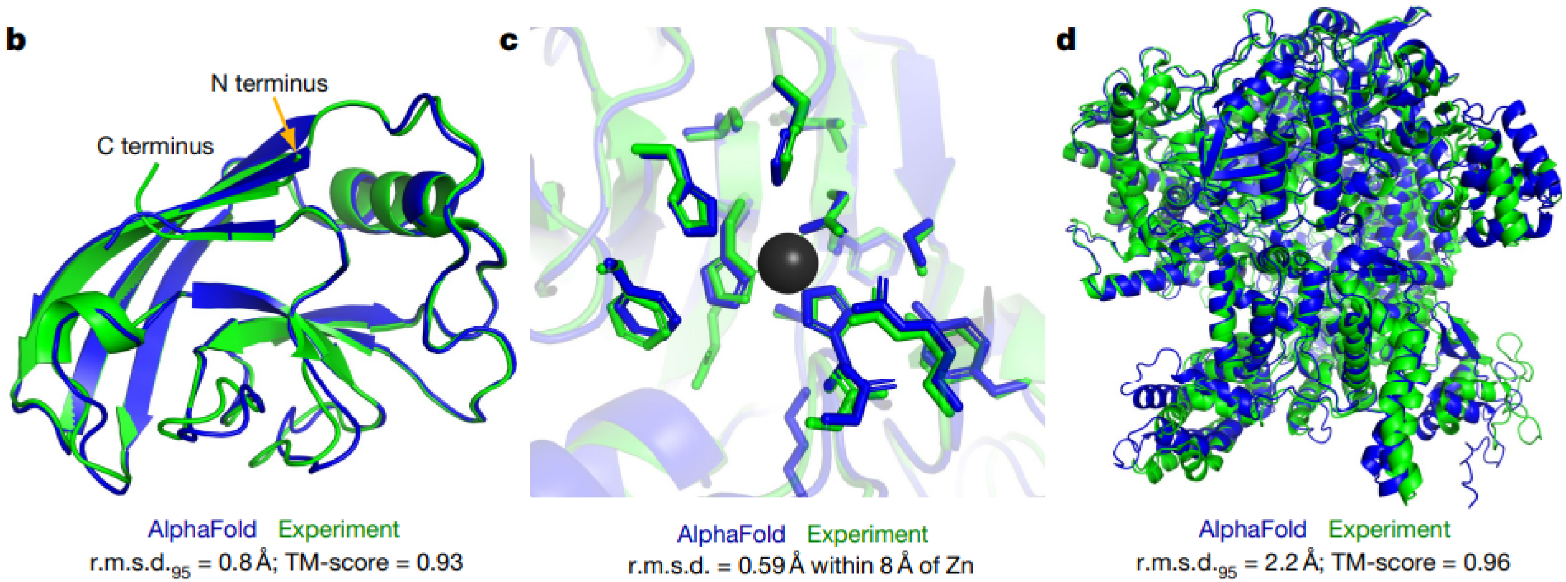
Input:
- primary sequence of target protein
- multiple sequence alignments of homologues
- structural information of homologues (PDB)

Process 1:
- MSA representation
- Residue pair representation

Process 2:
- 3D structure representation: rotation & translation of each residue
- refinement

AlphaFold2 - Nature 2021

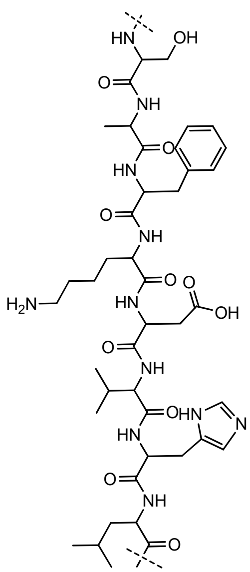
Protein structure prediction – ML approaches



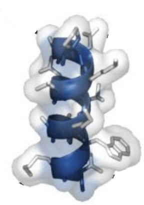
Result:
- highly accurate structure prediction
- low backbone RMSD
- correct residue orientation

AlphaFold - Nature 2021

Protein structure – A thermodynamic view

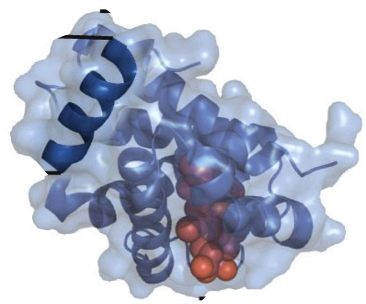


Amino acid chain
primary structure



α-helix
secondary
structure

transitions



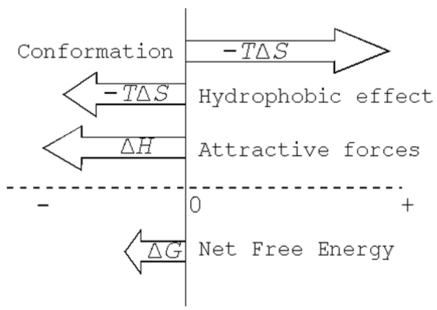
Protein domain
Tertiary structure

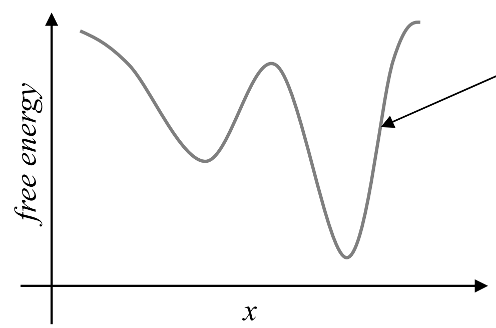
Protein complex
Quarternary structrue

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:





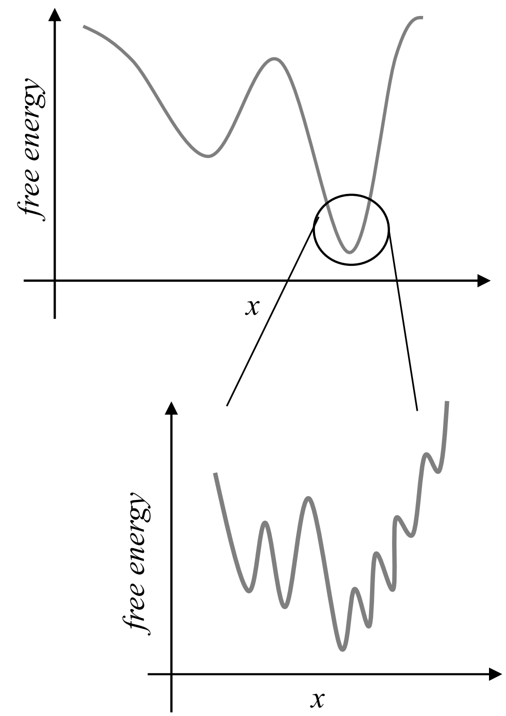
each (populated) protein conformation falls somewhere on this line

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



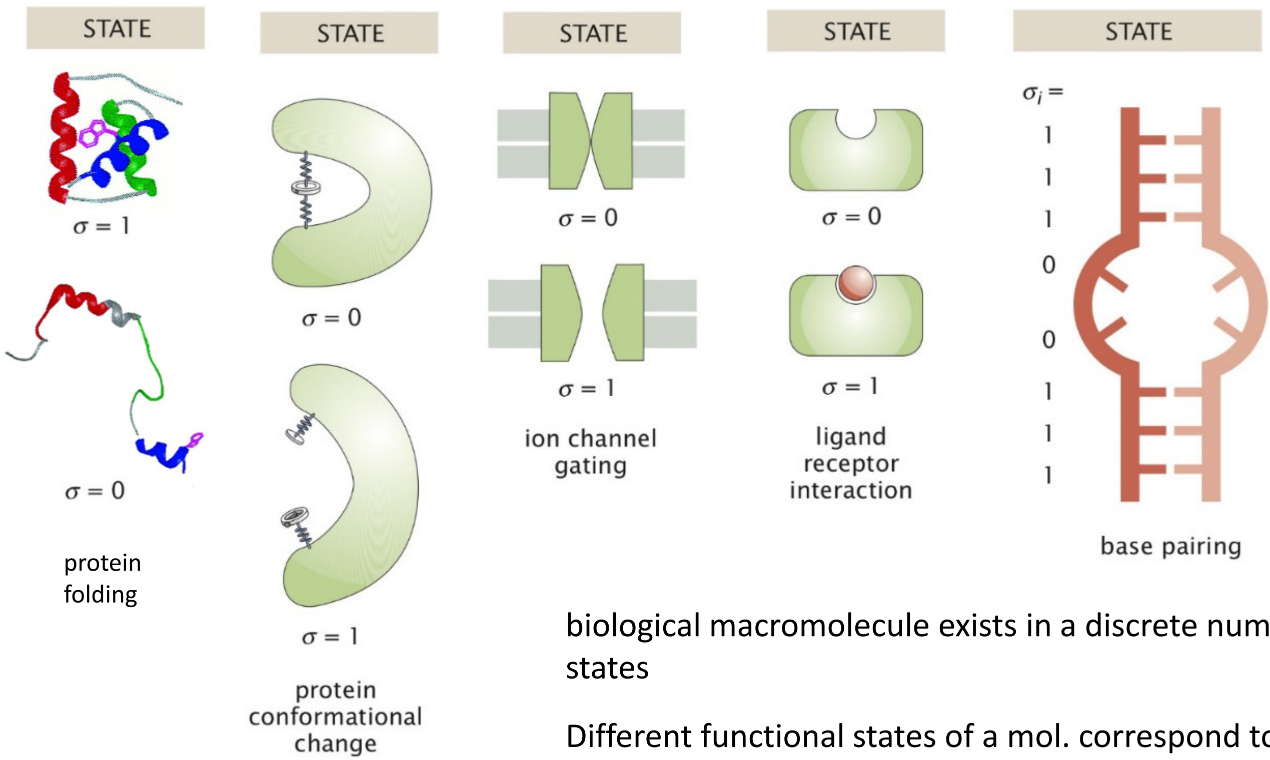
Global states in proteins

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

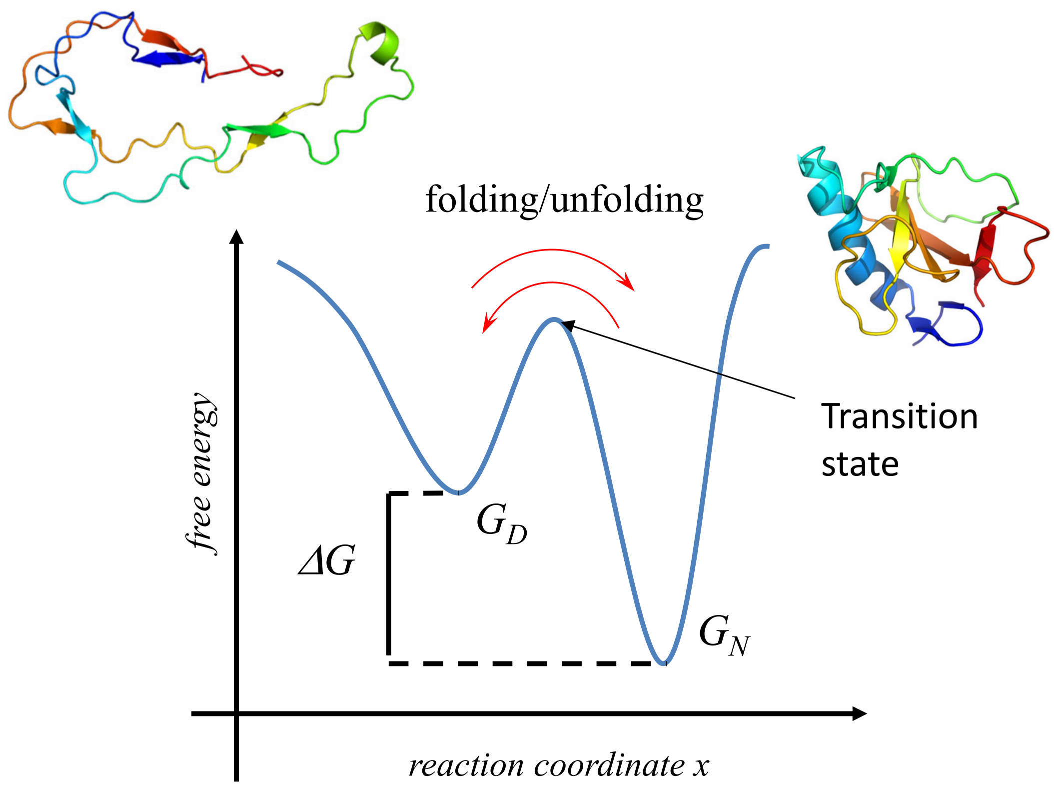


biological macromolecule exists in a discrete number of global states

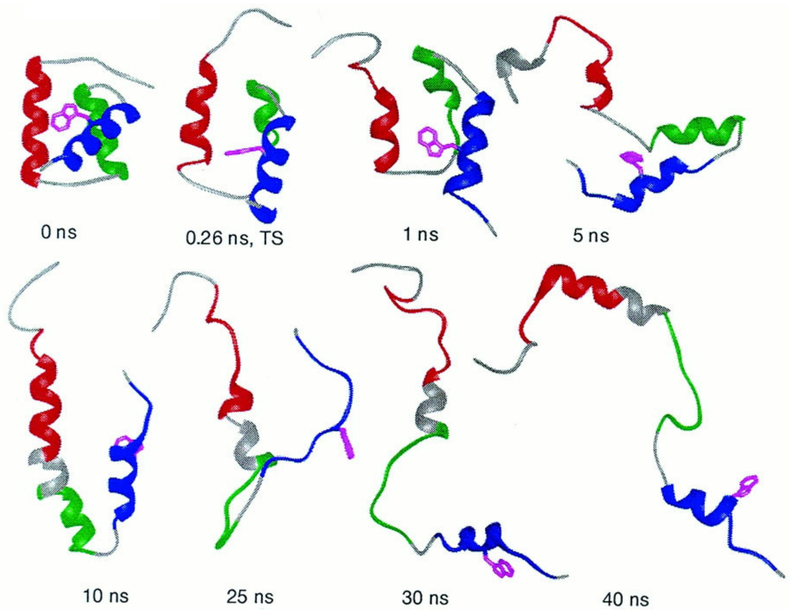
Different functional states of a mol. correspond to different global states

Changes in global states could be related to function

Global states in a protein: N and D



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?

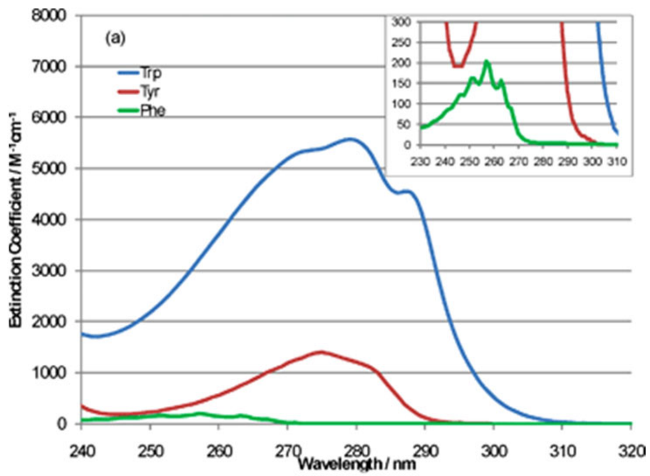
Any methods that can distinguish between U and N

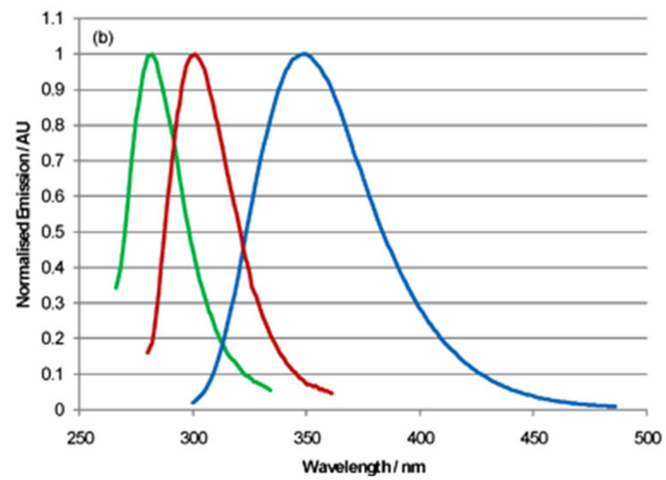
- **Absorbance** (e.g. Trp, Tyr) due to change in the micro-environment
- **Fluorescence** (Trp)-difference in emission spectrum & intensity, due to change in microenvironment
- **Circular dichroism** (far or near UV), due to change in asymmetric environment of fluorophores
- **Calorimetry** (DSC), due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

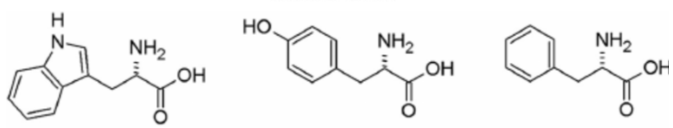
$\boxed{\rightarrow \text{fluorescence}}$

Amino acids: Fluroescence spectra

Absorbance and fluroescence spectra of aromatic amino acids



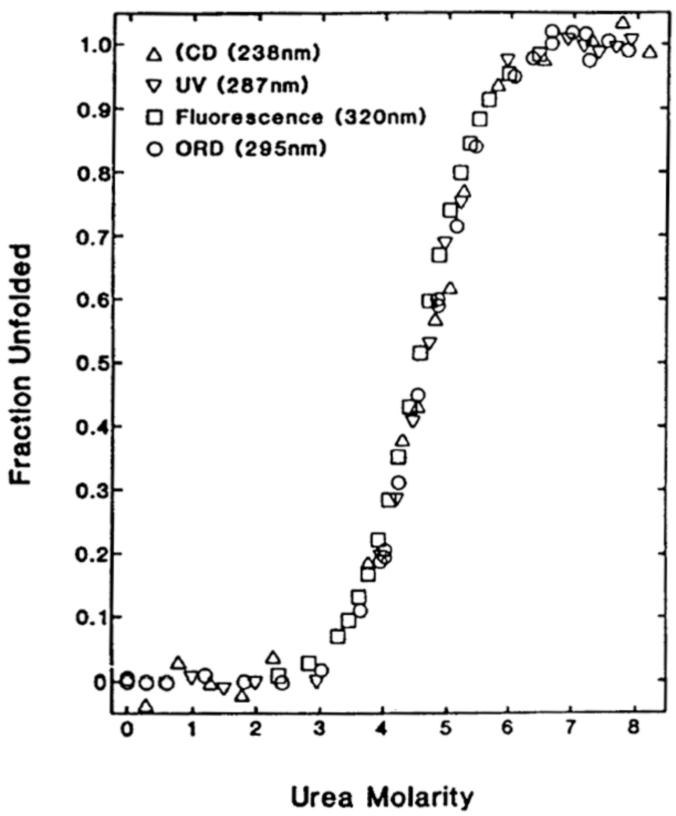




Trp Tyr Phe

Quantum yield for tryptophan: $Q_F=0.2$

Denaturation of globular, monomeric protein : RNase T1



$$N \rightleftarrows D$$

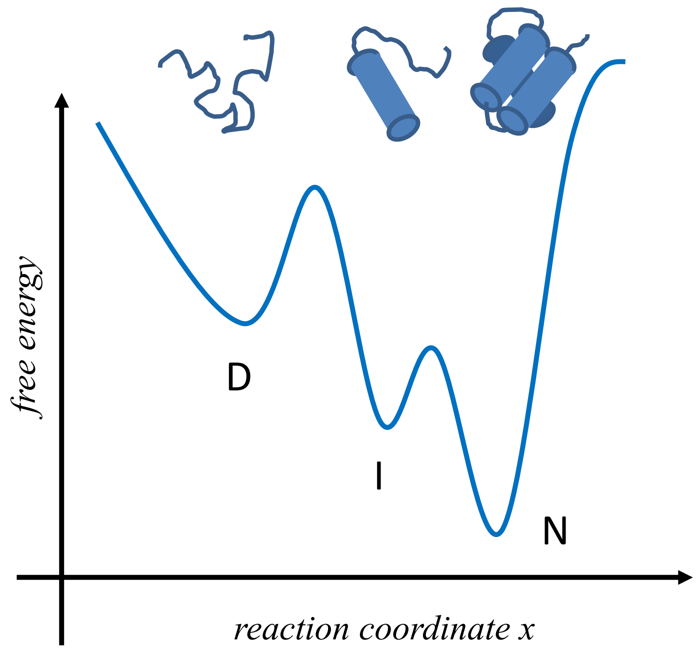
$$y = [\mathrm{N}]\,\mathrm{y}_{\mathrm{N}} + [D]\,\mathrm{y}_{D}$$

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

different parameters completely overlap

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

Three (or more) state equilibrium transitions

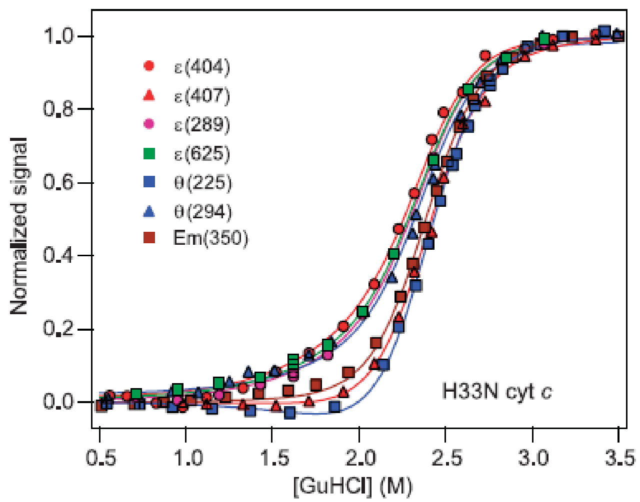


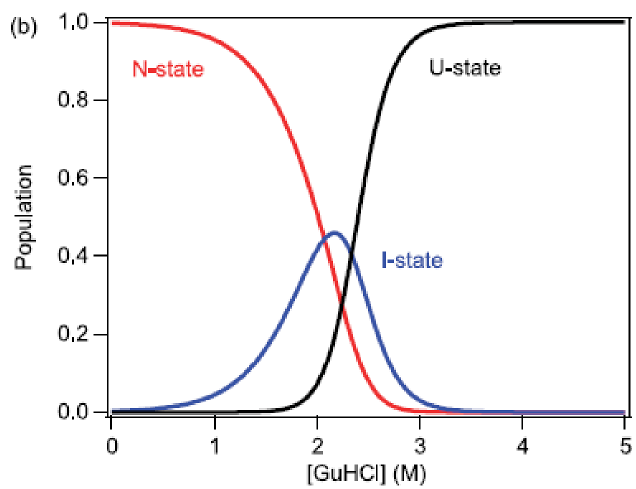
In many proteins, intermediates are populated.

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

unfolding intermediates

$$N \underbrace{\rightleftarrows X_1 \rightleftarrows X_2 \rightleftarrows \ldots}_{} \rightleftarrows D$$

apparent stability, eludicated by summing the individual contributions

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

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

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

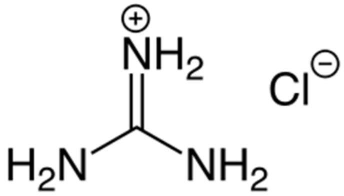
Denaturation with chemical agents

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

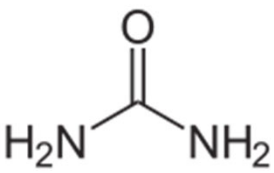


guanidinium chloride

2. Indirect effect:

   Alteration of water structure and thus diminishment of the hydrophobic effect

   Facilitation of the exposure of hydrophobic groups.



urea

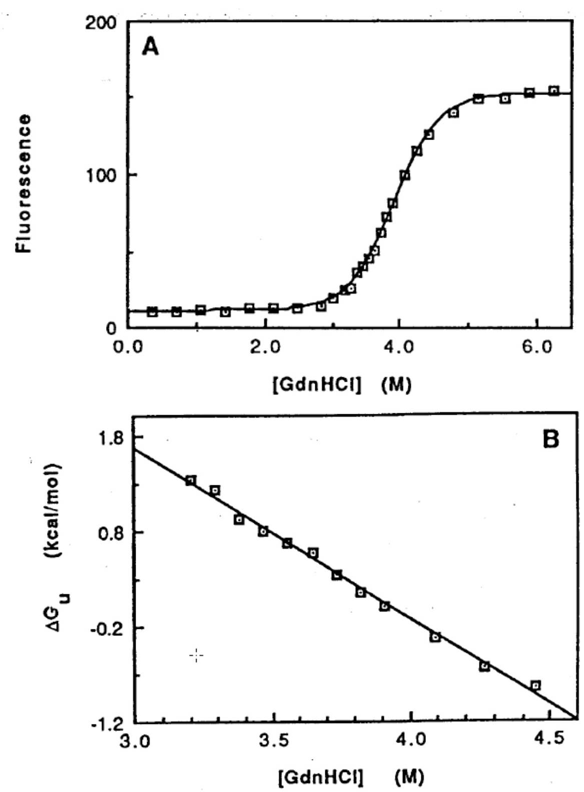
Protein denaturation with denaturants

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

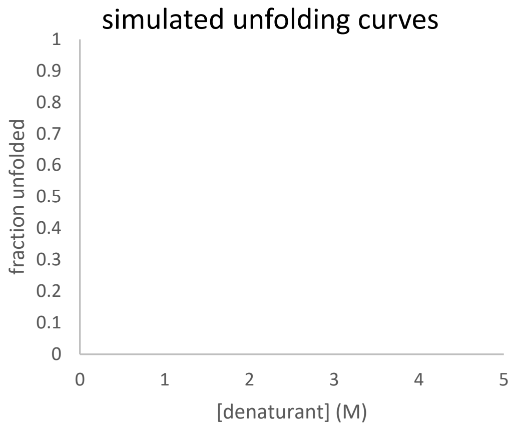
$$\boxed{\Delta G^0 = \Delta G^0_{H_2O} + m\,[\text{denaturant}]}$$

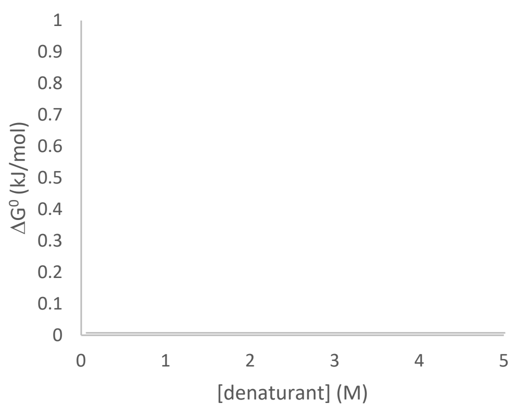
The **free energy of unfolding** can thus be determined by an extrapolation to **0 M denaturant**

Example: Chymotrypsin inhibitor 2 (CI2)



Sensitivity to denaturant





Parameters

— $\Delta G_0 = -40\ kJ/mol, \quad m = 18\ kJ/mol/M$

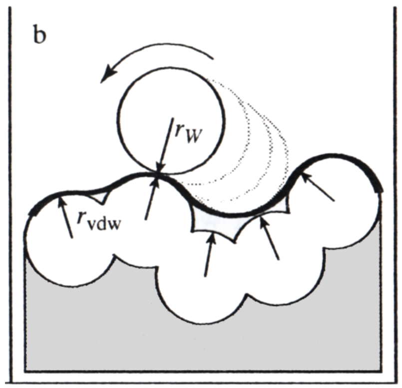
— $\Delta G_0 = -26.5\ kJ/mol,\ m = 12\ kJ/mol/M$

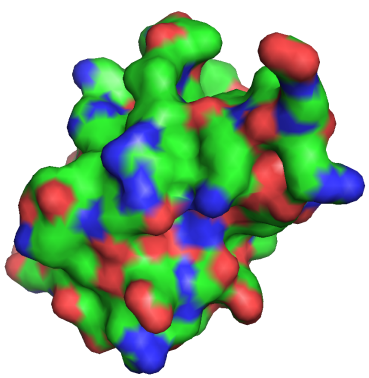
— $\Delta G_0 = -13.1\ kJ/mol,\ m = 6\ kJ/mol/M$

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?