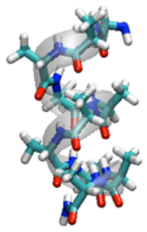


Non-covalent interactions and how macromolecules fold

Lecture 6: Protein folding and misfolding

Dr Philip Fowler

First-year Biophysics course

Objective:

Restate how proteins fold

Examine the thermodynamics of how a protein folds.

Determine what factors affect the stability of the folded, native structure.

Resolve the Levinthal Paradox

Investigate the role misfolded proteins play in disease

Summary:

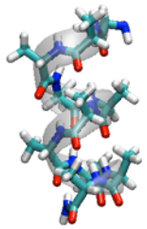
Proteins are marginally stable

We can study the folding of proteins using e.g. denaturants

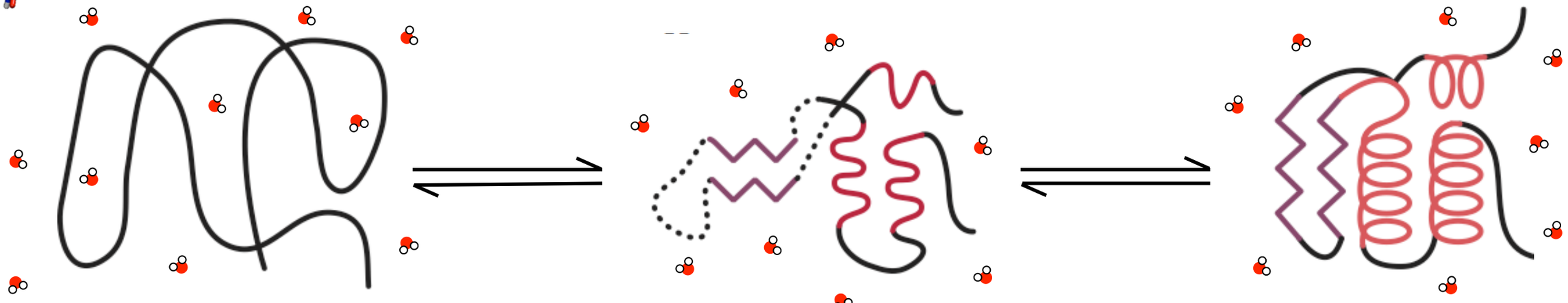
Proteins can be unfolded by changing pH and temperature or by adding denaturants

The concept of the protein folding funnel dispenses with the Levinthal Paradox

Misfolded proteins can form aggregates known as fibrils; prions are infectious proteins



Protein folding



unfolded, extended state (**random coil**)

compact **molten globule**

folded, compact native state

fewer protein—water hydrogen bonds

more water—water hydrogen bonds (can form complete network)

clusters of non-covalent interactions form **cooperatively**

protein

$\Delta H > 0$ ☹

protein can adopt fewer conformations when a molten globule

$\Delta S < 0$ ☹

native fold is less flexible than molten globule

$\Delta H < 0$ ☺☺

$\Delta S < 0$ ☹

much less local ordering of the water

water

$\Delta H < 0$ ☺

$\Delta S > 0$ ☺☺

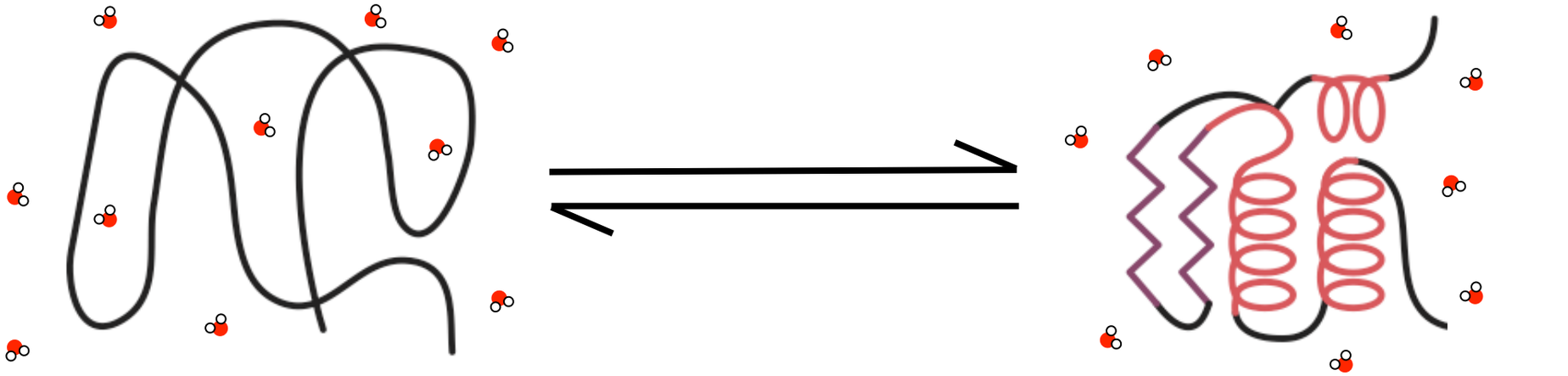
$\Delta H \approx 0$

$\Delta S \approx 0$

from the perspective of the solvent, there is little difference between the molten globule and the folded, native state

The **hydrophobic effect** causes the extended polypeptide chain to collapse and form a compact but dynamic **molten globule**. Clusters of **non-covalent interactions** within the protein then form **cooperatively**.

Thermodynamics of folding lysozyme at 25 °C



unfolded, extended state

folded, compact native state

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

ΔG = -60.9 kJ mol^{-1}

ΔH = -236 kJ mol^{-1}

$T\Delta S$ = -175 kJ mol^{-1}

How much lysozyme is unfolded?

$$\Delta G = -RT \ln K$$

K ~ 4.7 x 10^{10}

i.e. only 1 in 10^{10} proteins will be unfolded.

includes
(a) protein—protein ☺
(b) protein—water ☹
(c) water—water interactions ☺

includes
(a) increase in entropy of the water ☺
(b) decrease in the entropy of the protein ☹

(b) is estimated to be -2 190 kJ mol ☹

The different thermodynamic functions that govern protein folding are in a fine balance.
Proteins are marginally stable* at physiological temperature.

*compare to energy of a hydrogen bond which is 4 - 48 kJ mol^{-1}

data taken from Biophys Chem (1976) **4** 41 and Angew. Chem. Int. Ed. (1998) **37** 868

The tumour suppressor gene: p53

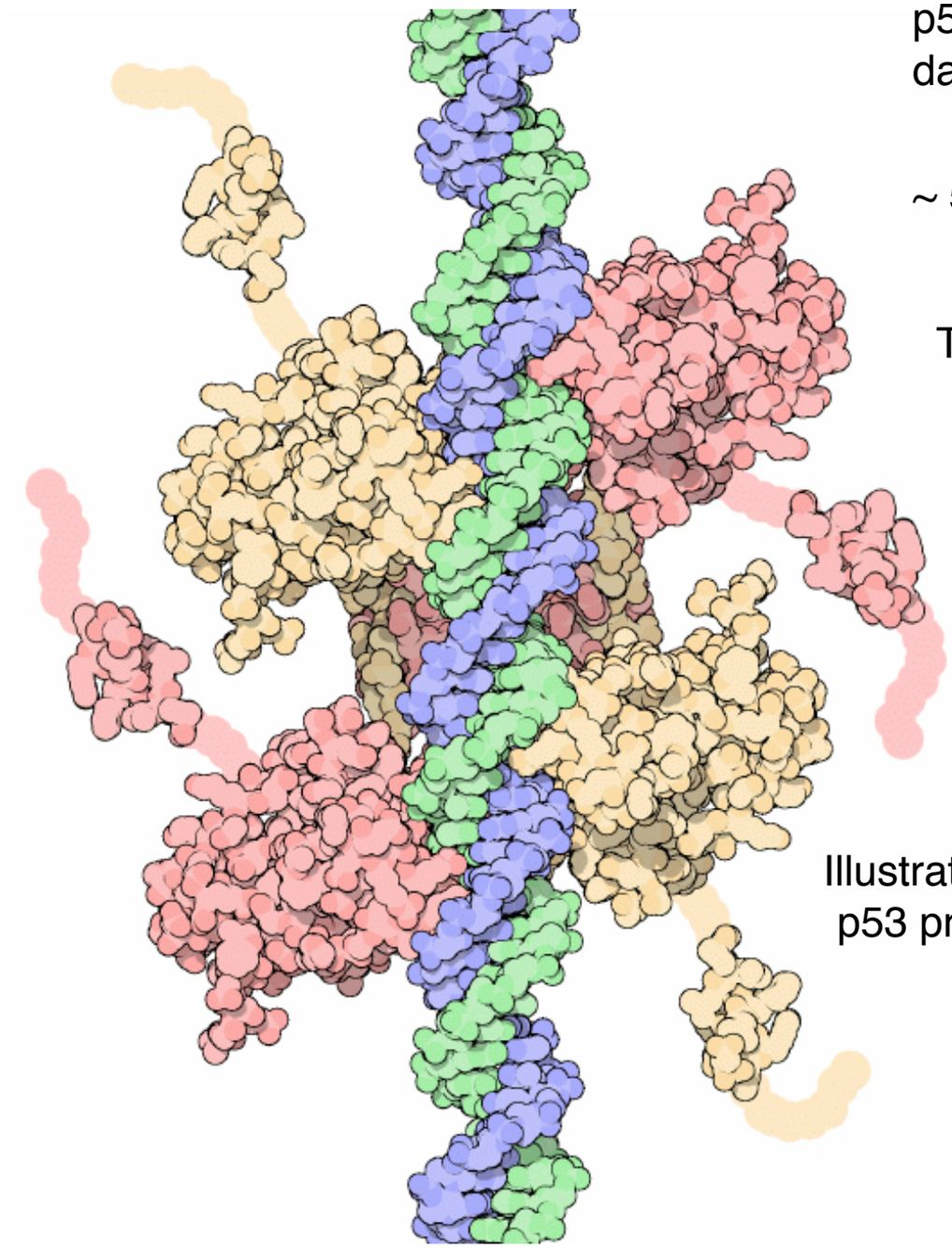


Illustration of p53 protein

DNA

p53 has many functions and for example can initiate both the repair of damaged DNA and apotosis

~ 50 % of human tumours contain a mutation or deletion of the *TP53* gene.

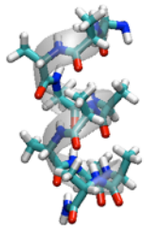
The structure of p53 is only marginally stable

$$\Delta G = \text{-25.1 kJ mol}^{-1}$$

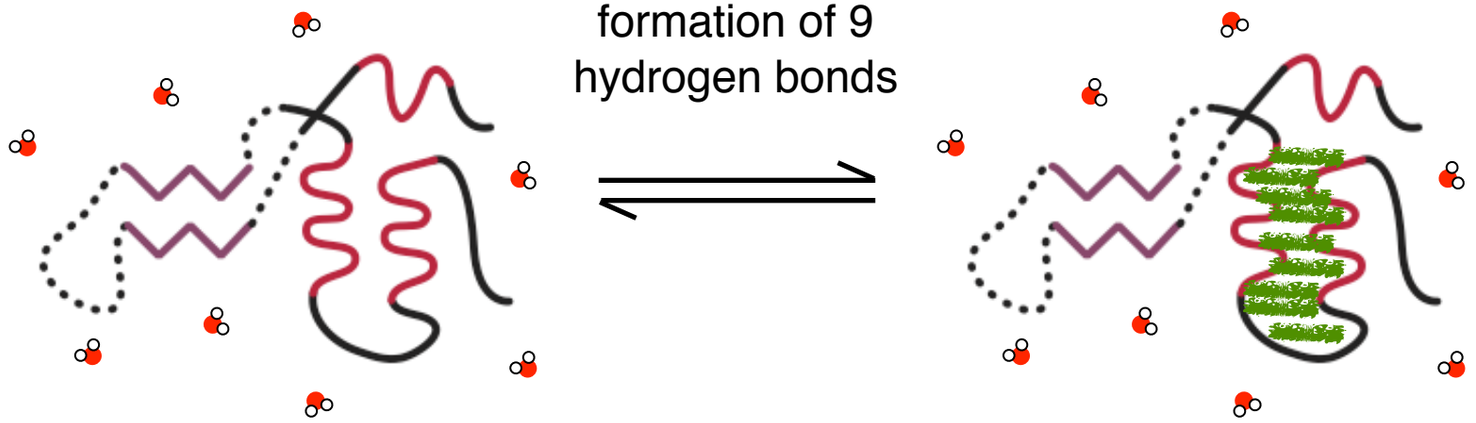
a single mutation, e.g. R175H, can reduce this to

$$\Delta G_{R175H} = \text{-12.5 kJ mol}^{-1}$$

63 615 publications in PubMed with p53 in title or abstract (15 197 more when I started teaching this course in 2010 - 10 per day) PNAS (1997) **94** 14338



Unfolding as well as folding is a cooperative process



The first hydrogen bond significantly reduces the degrees of freedom of the protein and therefore the entropy decreases

$\Delta H_1 < 0$ ☺ $\Delta S_1 << 0$ ☹ ☹

Adding a second hydrogen bond also reduces the degrees of freedom of the protein, but by not as much as the first..

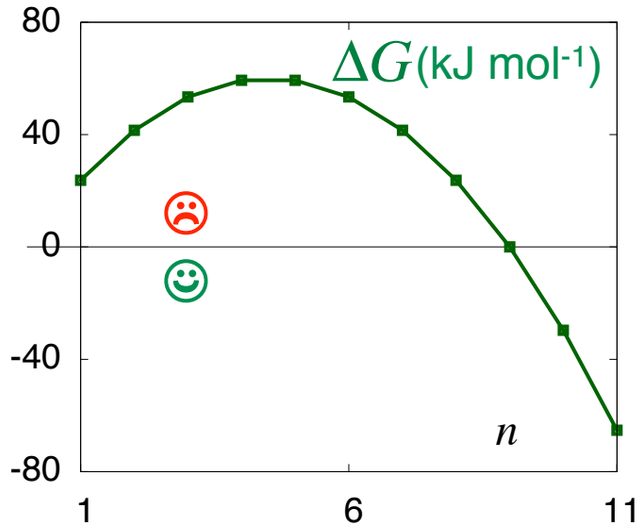
$\Delta H_2 < 0$ ☺ $\Delta S_2 < \Delta S_1 << 0$ ☹ ☹

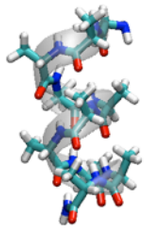
..but this is still unfavourable

Only when *n* hydrogen bonds are made **cooperatively** is the process favourable

This process applies equally in **reverse** i.e. breaking a single hydrogen bond in the native, folded structure makes it easier to break another.

The high degree of cooperativity helps explain why protein folding is a **two-state process**.





protein modification

e.g. disulphide bonds

Temperature

↑T ↑S

Denaturants

e.g. urea

Configurational entropy of the unfolded state

Strength of the hydrophobic effect

Factors affecting the stability of proteins

Formation of co-operative non-covalent interactions in the folded state

Strength

Number

pH

Mutation

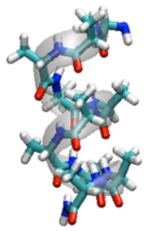
(see earlier p53 example)

extensive network of ion-pairs on the surface of the protein and at the interface between domains and subunits

e.g. Glutamate dehydrogenase*

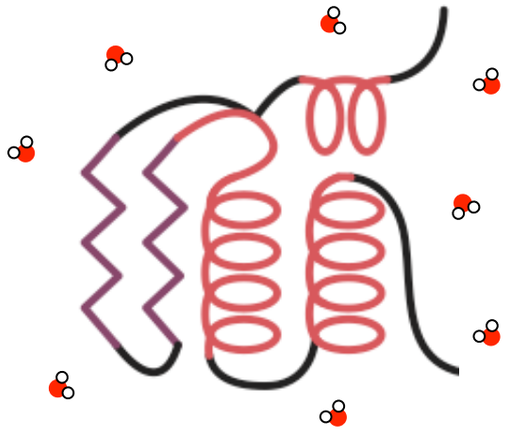
	T (°C)	half-life (hours)
(a) hyperthemophile *Pyrococcus furiosus*	100	12
(b) mesophile *Clostridium symbiosum*	50	0.3

*Structure (1995) **3** 1147



Proteins can be unfolded by mutation, denaturants or changes in temperature and pH

lysozyme



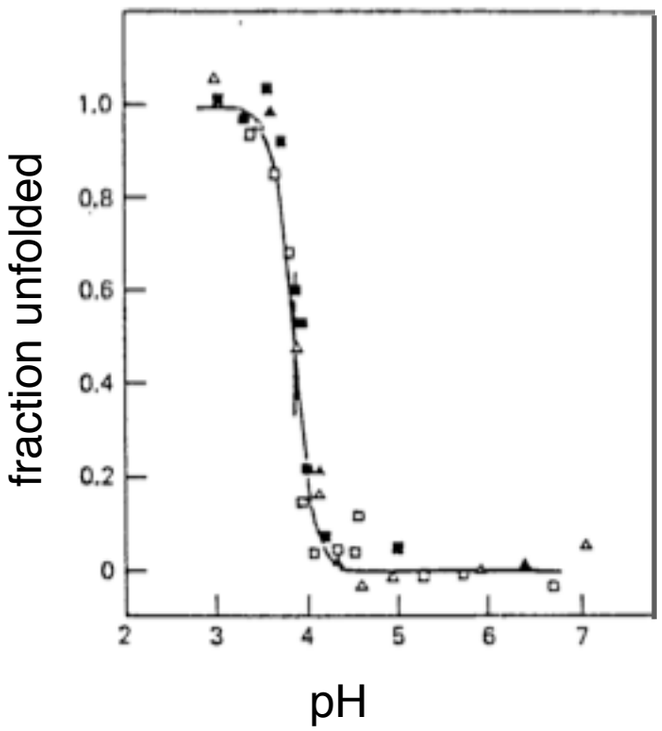
ΔG = -60.9 kJ mol^{-1}

Even though proteins are marginally stable, they are stable enough that there is very little unfolded protein.

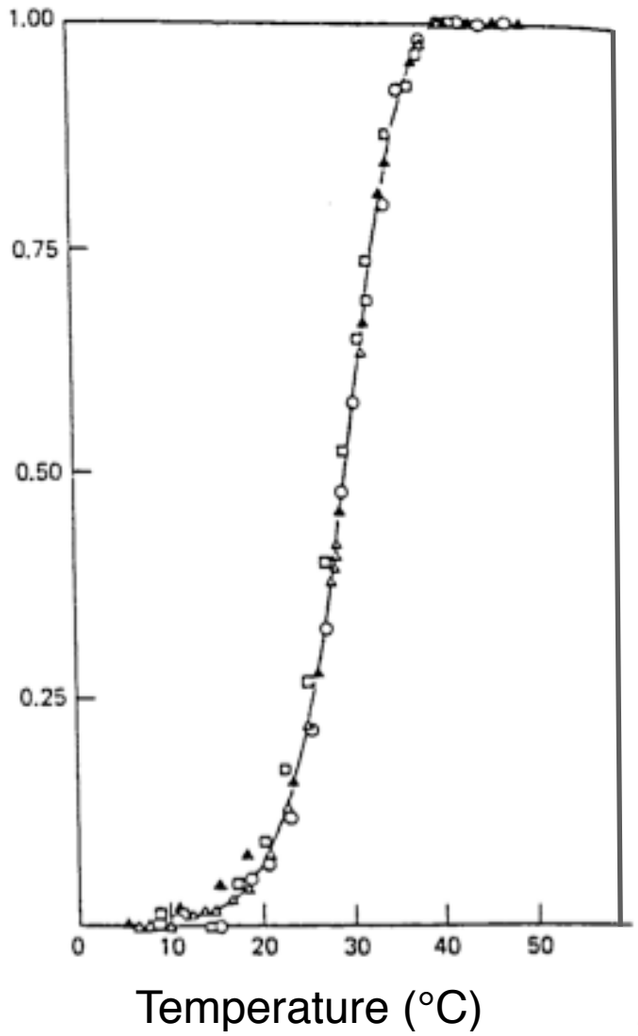
We need to unfold (**denature**) proteins so that we can study how they fold

To do that we reduce the stability of the folded state

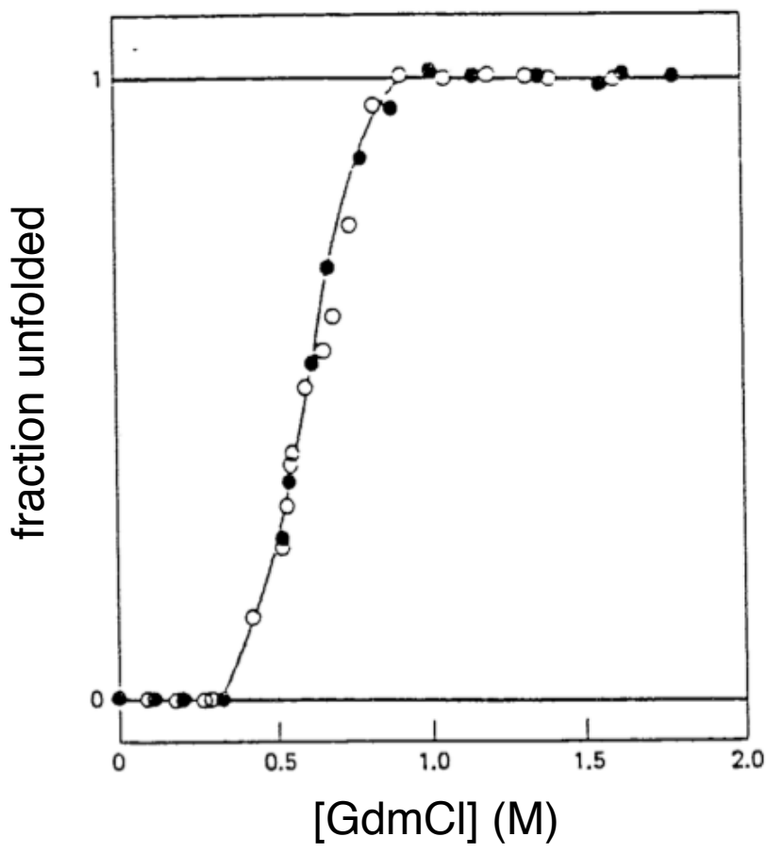
Staphlococcal nuclease A



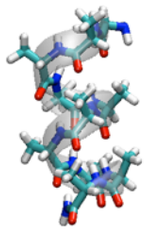
Bovine ribonuclease A



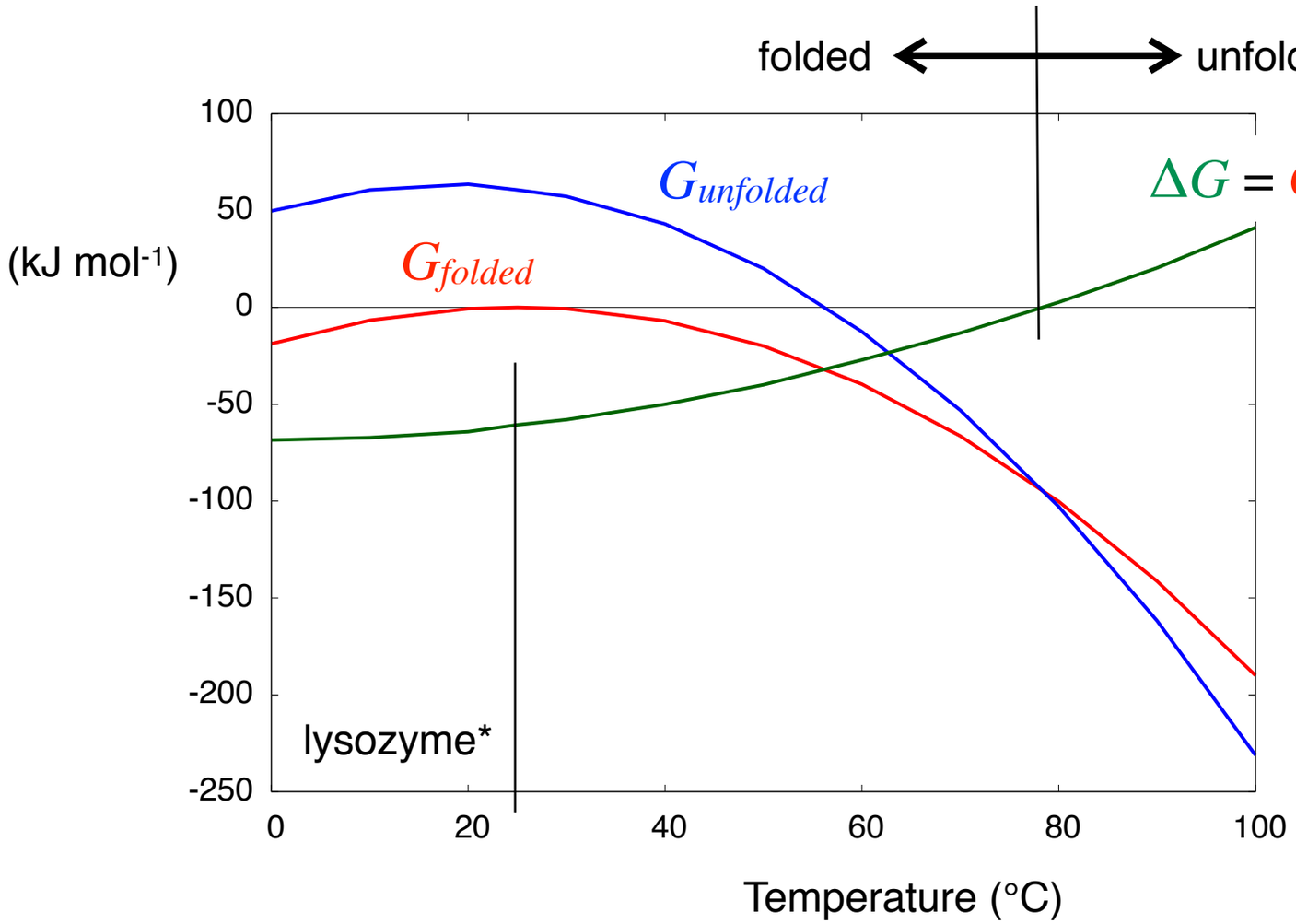
Phosphoglycerate kinase

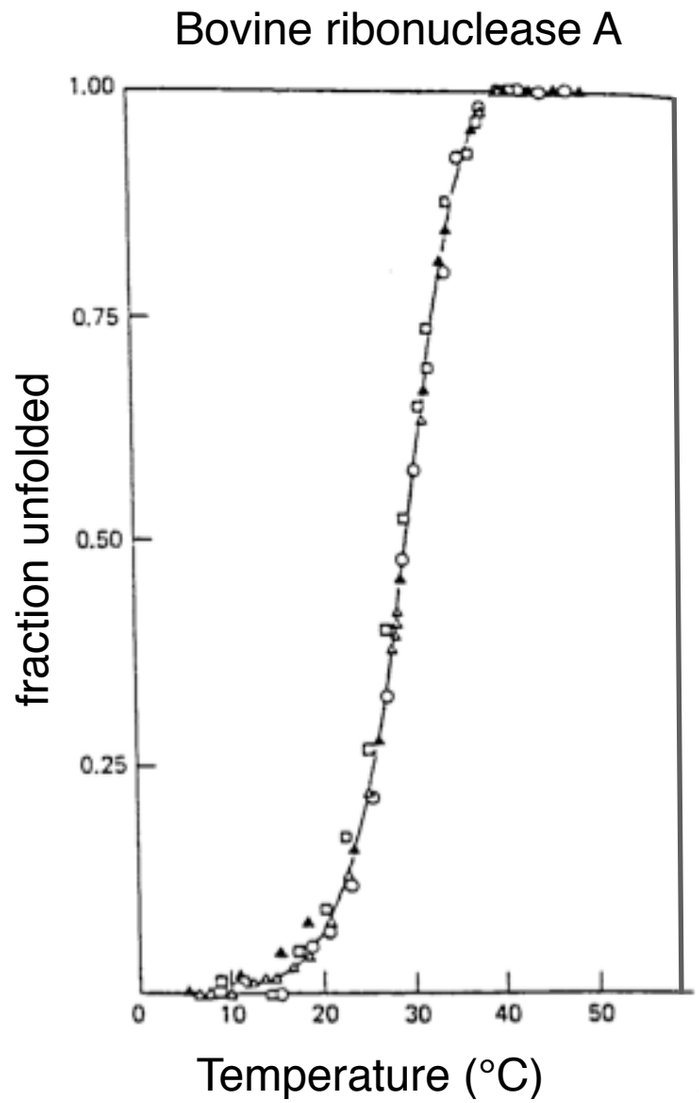


Creighton (2nd ed) Figure 7.11, pg 288

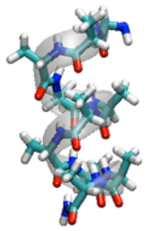


Most proteins unfold at temperatures > 60 °C and < 0 °C





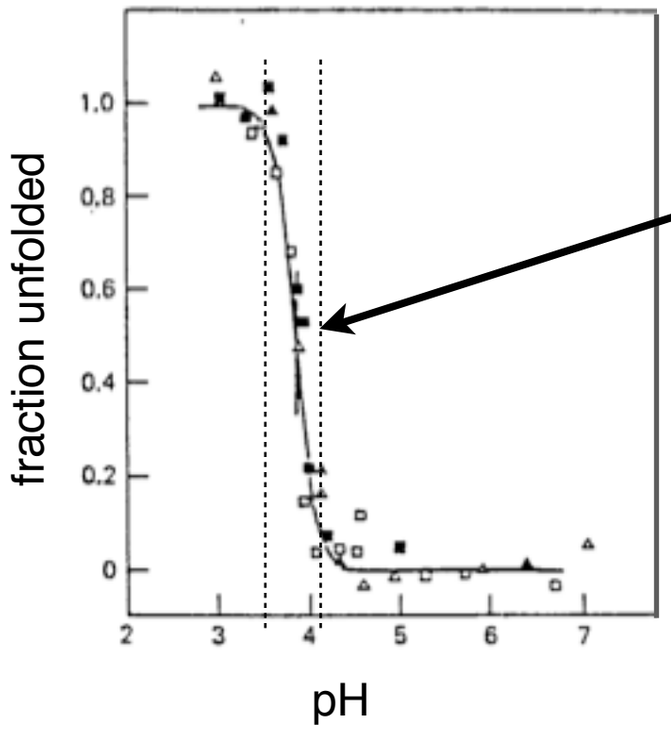
*data taken from Biophys Chem (1976) **4** 41



Most proteins unfold if pH < 5 and pH > 10

Buried groups becomes ionised at these pHs e.g. His and Tyr → electrostatic repulsion

Staphlococcal nuclease A



Proteins unfolds over only 0.3 pH units

yet 2 pH units are required to go from
9% ionisation of a group to 91%

illustrates the co-operative nature of protein folding

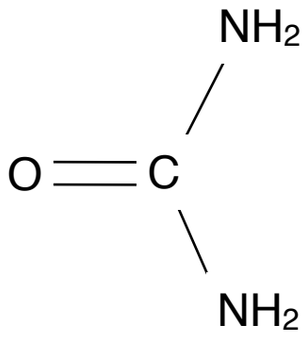
Salt bridges can increase the stability of proteins
(see Glutamate dehydrogenase example*)

*Structure (1995) **3** 1147

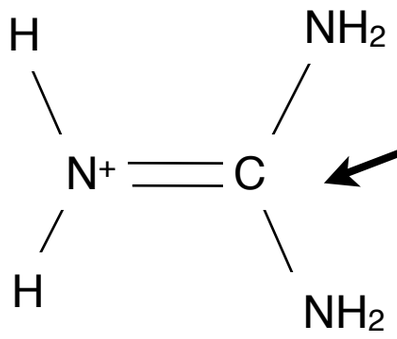
Denaturants

The two most common denaturants are urea and GdmCl

They are **chaotropic agents**



Urea



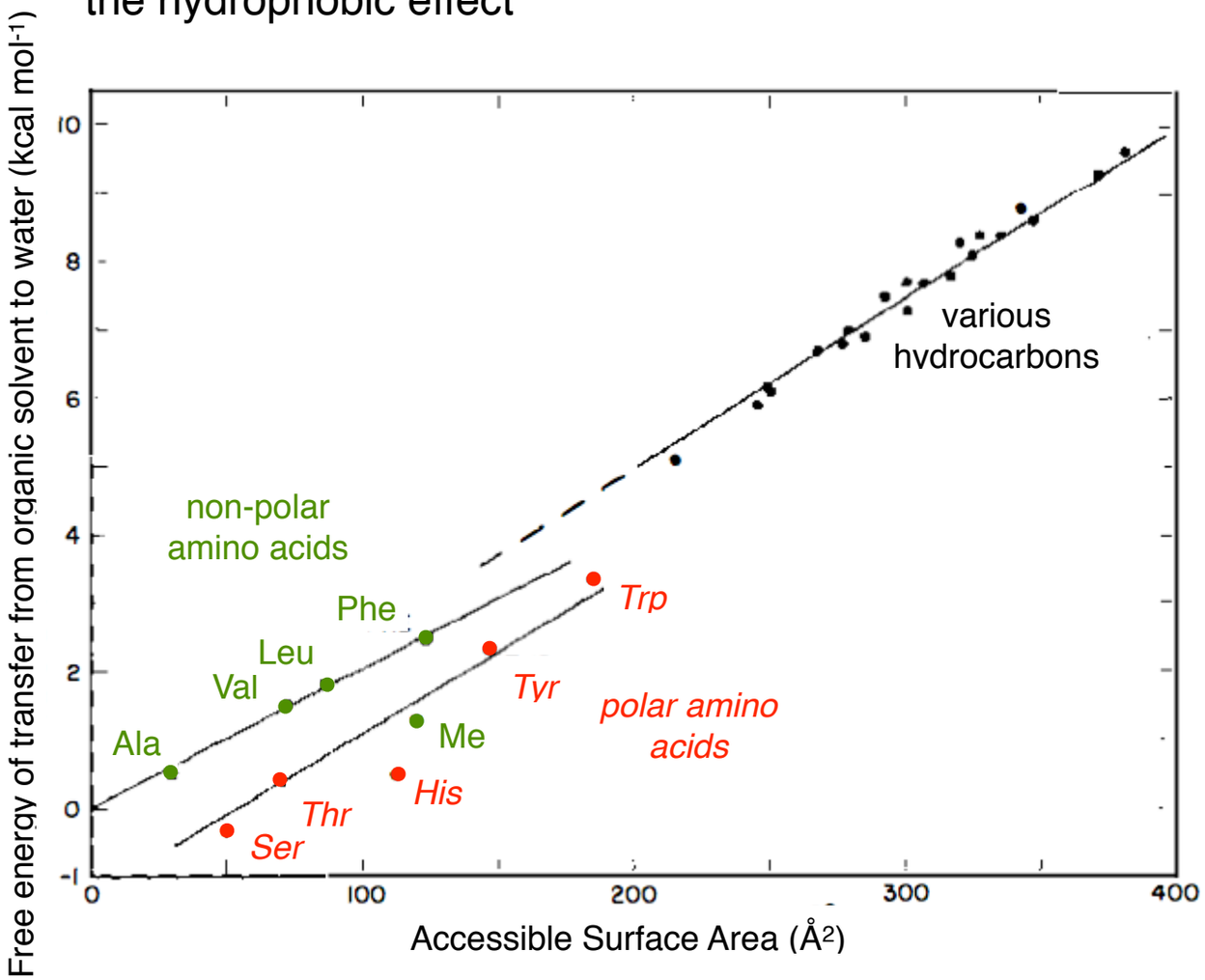
Guanidinium ion

from GdmCl

increase the solubility of nonpolar substances in water

They are usually used at high concentrations (5-10 M)

Accepted that they increase the stability of the unfolded, extended polypeptide chain by reducing the strength of the hydrophobic effect



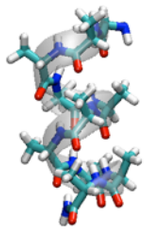
The precise mechanism is not well understood but we do know that it is favourable to transfer non-polar amino acids from water to an organic solvent.

Experimental evidence that protein in urea is unfolded

1. increase in viscosity from 5 to 50 $cm^3\ g^{-1}$

2. increased rates of hydrogen exchange on the N-H groups (because no secondary structure)

Voet (3rd ed) pg 159



Using denaturants to determine K_f

$$\Delta G_f = -RT \ln K_f$$

under normal conditions
i.e. no denaturant

presence of denaturant alters the equilibrium and therefore ΔG is a function of [denaturant]

but we can extrapolate back to find ΔG_f i.e. $\Delta G_f = \lim_{[\text{denaturant}] \to 0} \Delta G$

if the relationship is linear then we can write $\Delta G = m[\text{denaturant}] + \Delta G_f$

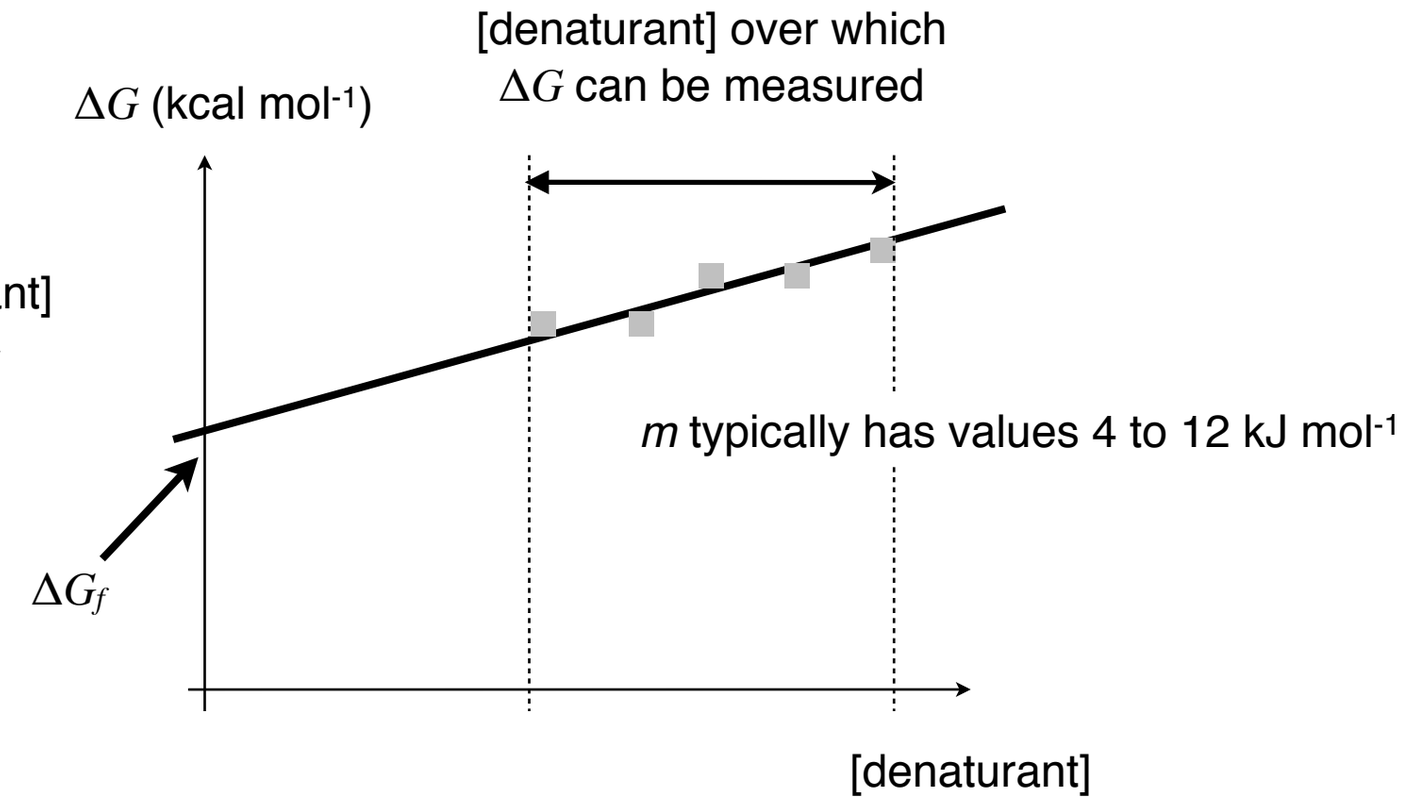
this resembles y = mx + c
(see Maths course)

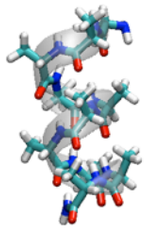
free energy of folding measured at [denaturant]

gradient

free energy of folding under normal conditions
i.e. no denaturant

plot ΔG against [denaturant]
and the y-intercept is ΔG_f





Anfinsen (1957) showed that proteins fold reversibly

1. denature the protein and reduce the disulphide bonds

+ 8M urea

+ mercaptoethanol

Ribonuclease A
(124 aa)

- 8M urea

2a. allow the protein to refold

- mercaptoethanol

+ O_2

3a. oxidise the cysteine residues

~ 100% active Ribonuclease A

2b. oxidise the cysteine residues when the protein is denatured

- mercaptoethanol

+ O_2

3b. allow the protein to fold

- 8M urea

~ 1% active Ribonuclease A

What percentage of these structures are native?

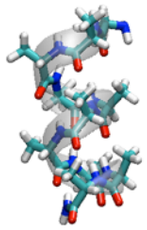
probability of forming the native state = $\frac{1}{7} \times \frac{1}{5} \times \frac{1}{3} \times \frac{1}{1} = \frac{1}{105}$

if we choose a cysteine, only one of the remaining 7 is its correct partner

$= \frac{1}{105}$

$= 0.95\ \%$

Science (1973) **181** 223 Voet (3rd ed) pg 159-160



The Levinthal paradox

Anfinsen's dogma / thermodynamic hypothesis: *"The native conformation is determined by the **totality of interatomic interactions** and hence by the amino acid sequence, in a given environment."*

How does a protein find its native state? → Assume it samples all conformations and chooses the one with the lowest (free) energy

Calculation

If we have a protein with n residues and

...each residue has 2 torsion angles ϕ and ψ

...assume ϕ and ψ can each exist in 1 of in 3 stable conformations

Then the number of conformations is

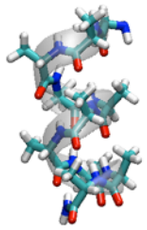
$$3^{2n} \approx 10^{n}$$

If we have a 100 residue protein and assume it takes 100 fs to convert between each conformation then the protein will fold in

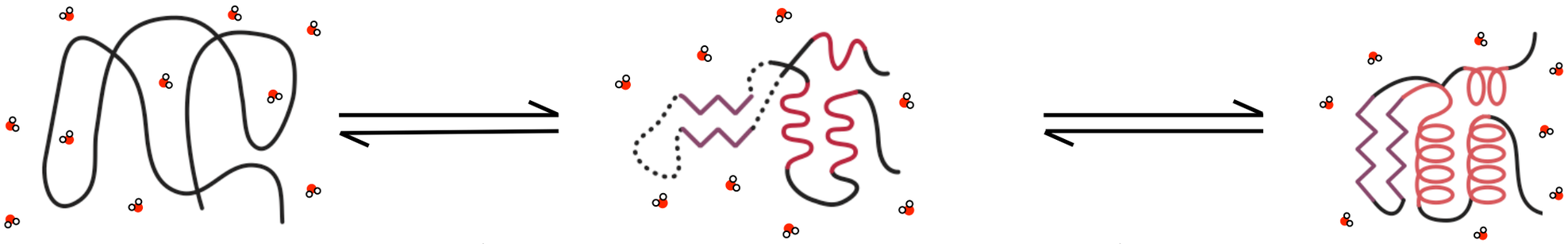
$$10^{100} \times 10^{-13}\mathrm{s} = 10^{87}\mathrm{s} = 10^{79}\mathrm{years}$$

This is not possible (most proteins fold in ms - s)

Voet (3rd ed) pg 161-162



Hence proteins do not sample every conformation as they fold



unfolded, extended state → compact molten globule → folded, compact native state

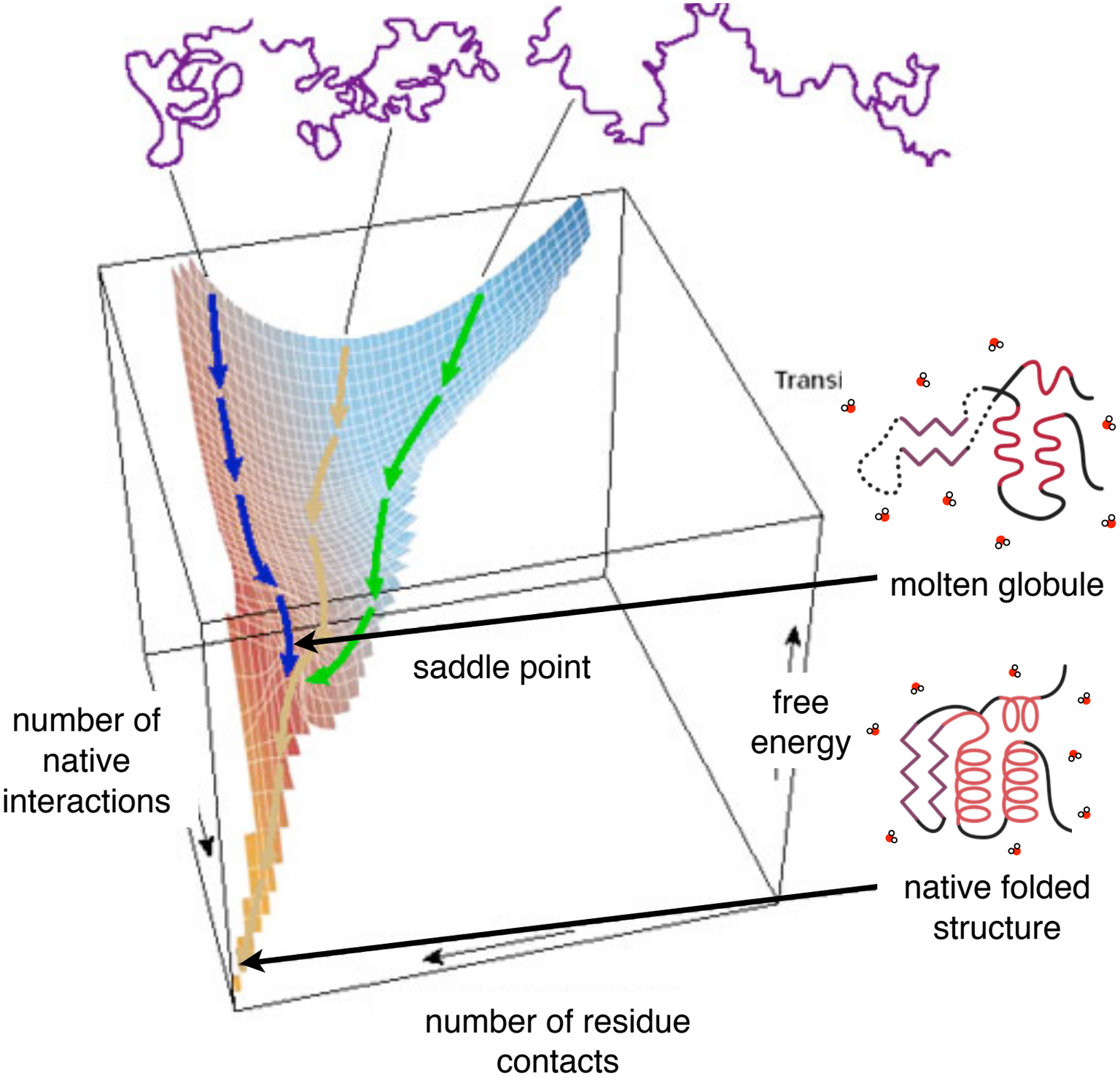
the hydrophobic effect drives **all** the unfolded extended conformations towards the molten globule

the cooperative formation of non-covalent interactions drives the molten globule towards the folded, native state

Combined together there is always a ‘direction’ i.e. the protein is always moving ‘downhill’

→ The concept of the **protein folding funnel**

The protein folding funnel



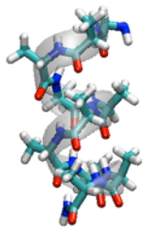
A funnel cannot be too **rugged** / bumpy otherwise the protein will get stuck in **metastable states**

1. Not all proteins can fold unaided *in vivo.* They need the help of **chaperones**

2. Some proteins can get stuck in such states; in these states then encourage the unfolding and aggregation of other proteins this can lead to **disease**.

Proteins are **optimised** by evolution not only to perform their function but also to fold *in vivo*

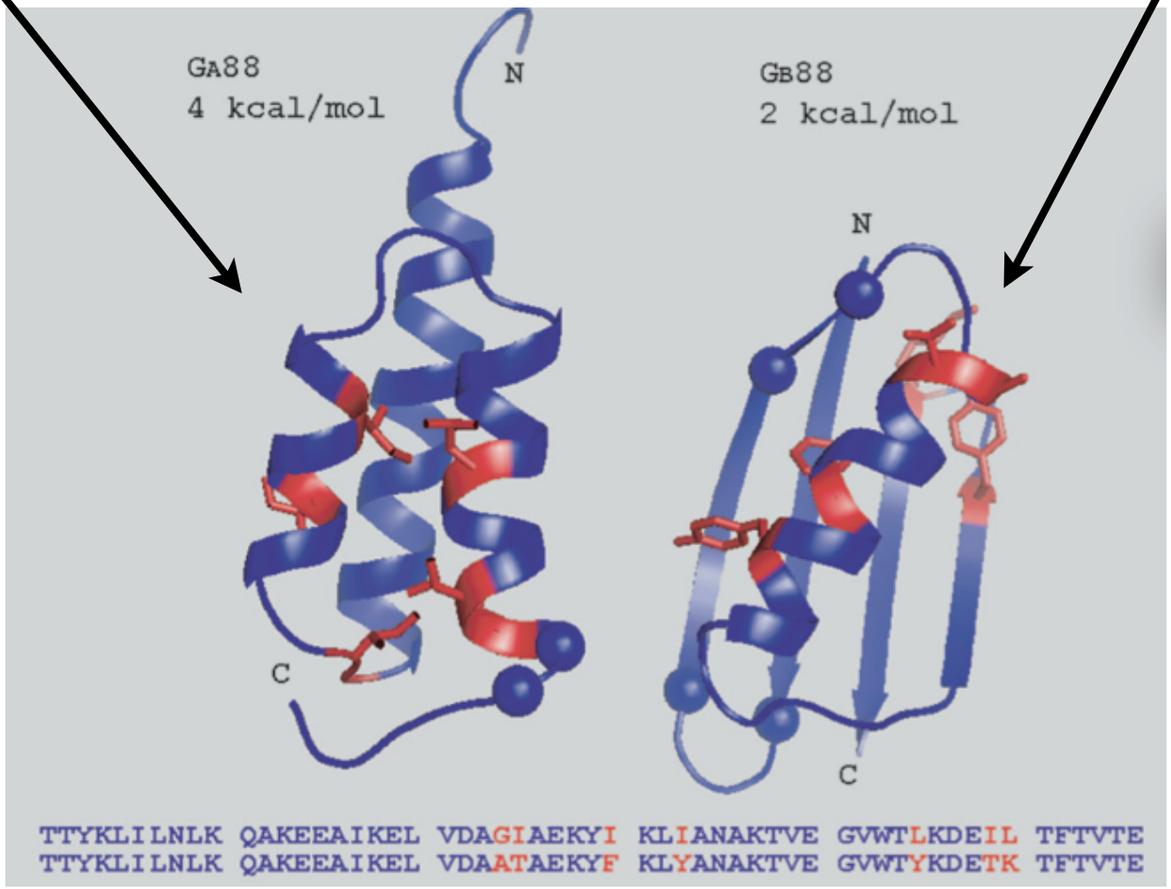
Nature (2003) **426** 884 Angew. Chem. Int. Ed. (1998) **37** 868



Small differences in sequence can lead to large differences in structure and function

3-α helix fold with
albumin-binding epitope

α/β fold with IgG epitope



both have 56 residues:

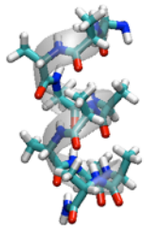
TTYKLILNLK QAKEEAIKEL VDAGIAEKYI KLIANAKTVE GVWTLKDEIL TFTVTE
TTYKLILNLK QAKEEAIKEL VDAATAEKYF KLYANAKTVE GVWTYKDETK TFTVTE

PNAS (2007) **104** 11963

49 residues (88 %) are identical

7 residues (88 %) are different

...the reverse can also be true. Large differences in sequence may only lead to small differences in structure and function.

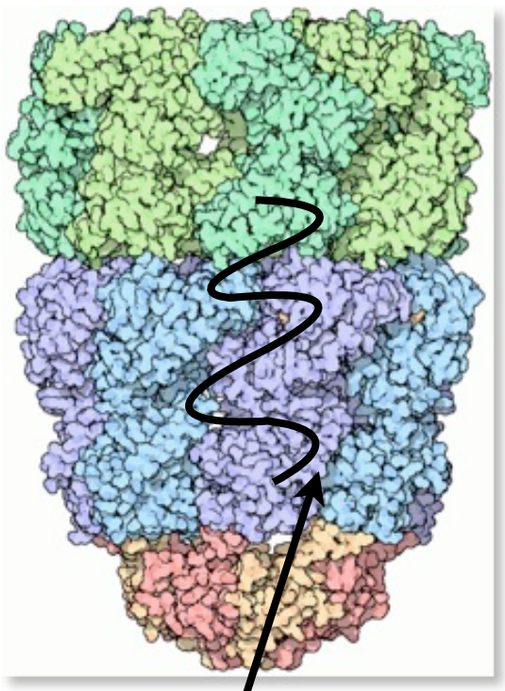


Chaperones

These are proteins that aid other proteins to fold to their correct native structure

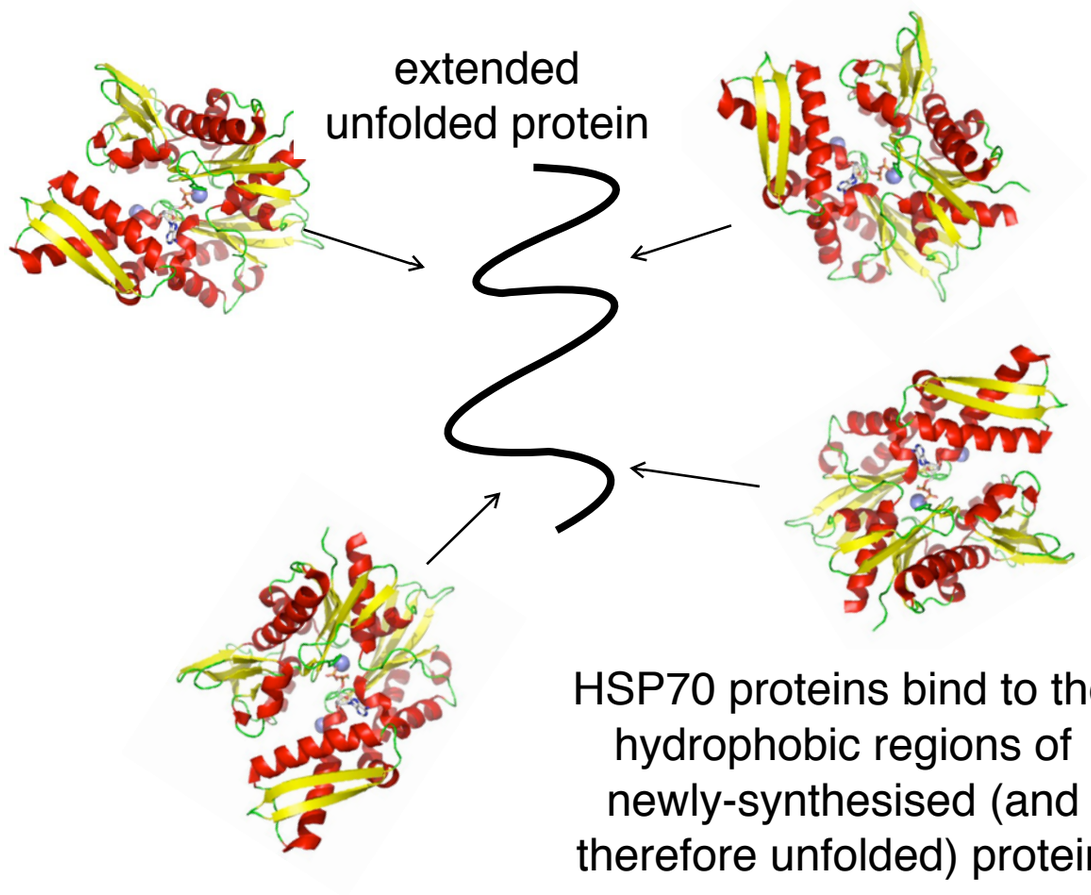
They can do this in more than one way. Two examples are

1. GroEL/GroES



These chaperones provide an enclosed cavity in which proteins can fold without interacting with the rest of the cell

2. HSP70 (Heat Shock Protein 70kDa)

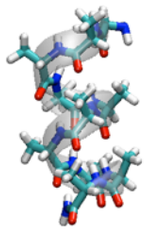


HSP70 proteins bind to the hydrophobic regions of newly-synthesised (and therefore unfolded) protein

This prevents the premature folding of the protein

The conformational changes and binding of both these proteins are powered by the hydrolysis of ATP

Voet (3rd ed) pg 165

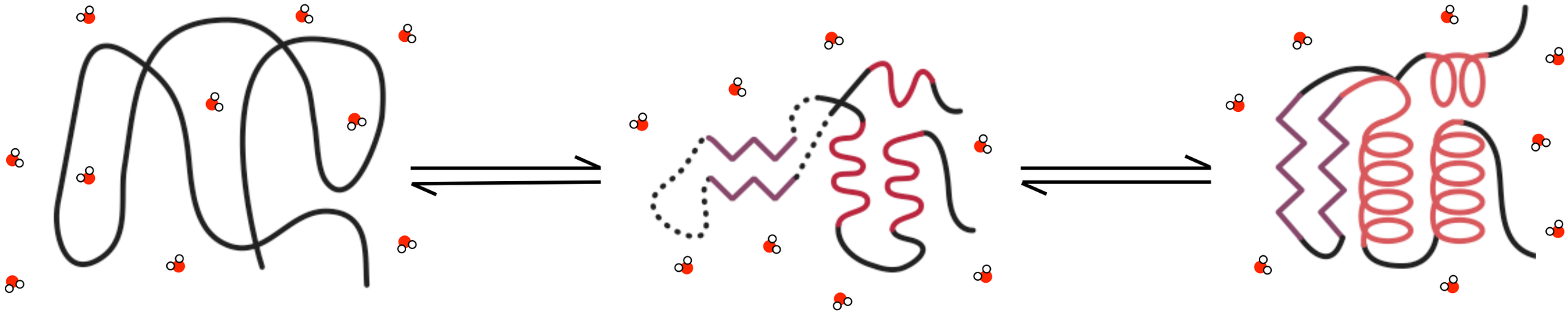


Protein misfolding* and disease

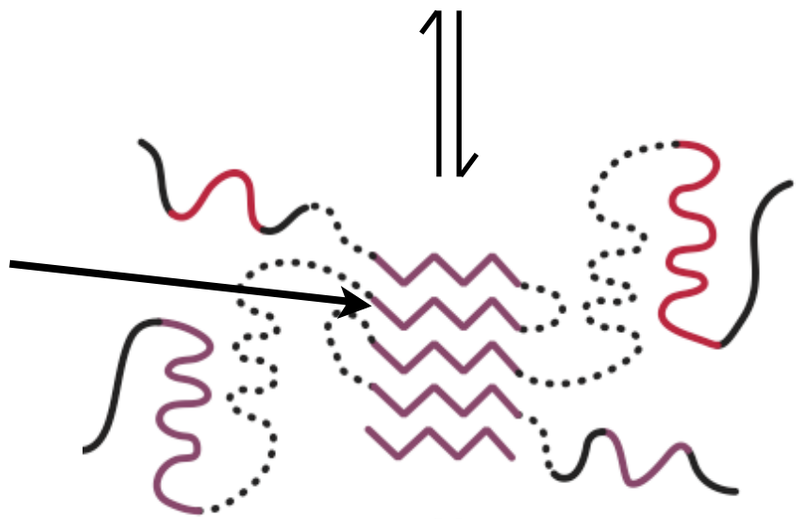
unfolded, extended state

compact **molten globule**

folded, compact native state

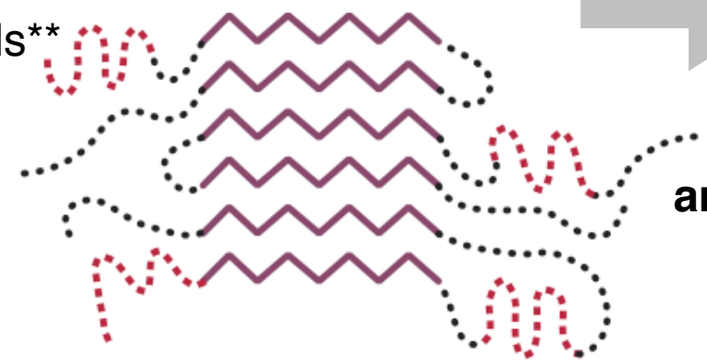


unfolded proteins form many inter-molecular hydrogen bonds forming a β-sheet like structure



amyloid precursor

Two naturally occuring variants of human lysozyme (I56T and D67H) form amyloid fibrils**

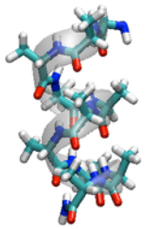


amyloid fibril

It is possible that many proteins can form this metastable aggregated state

Nature (1997) **385 787 Nature (2003) **426** 900

*Note we are ignoring the more obvious mechanisms of e.g. mutation that destabilises the native structure



Amyloid fibrils can become very large

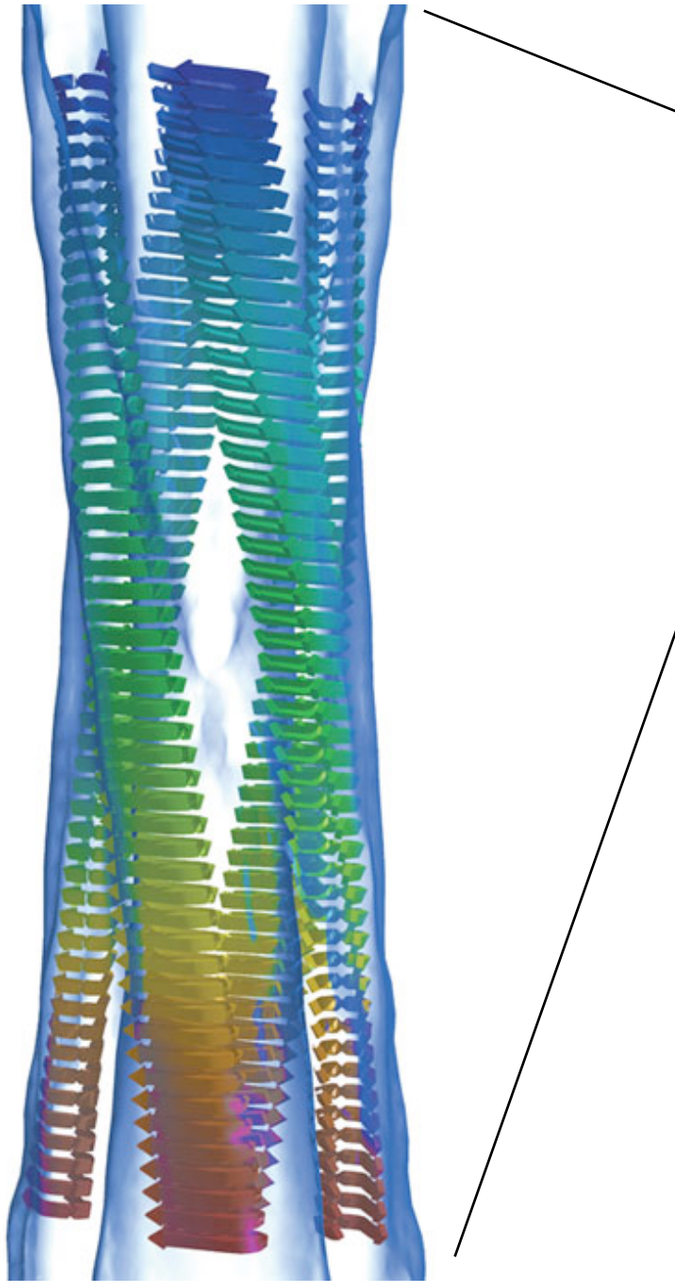
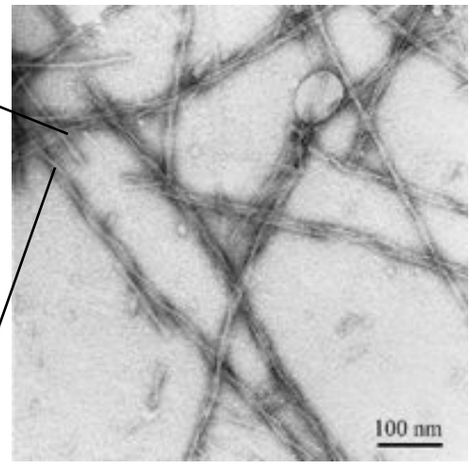
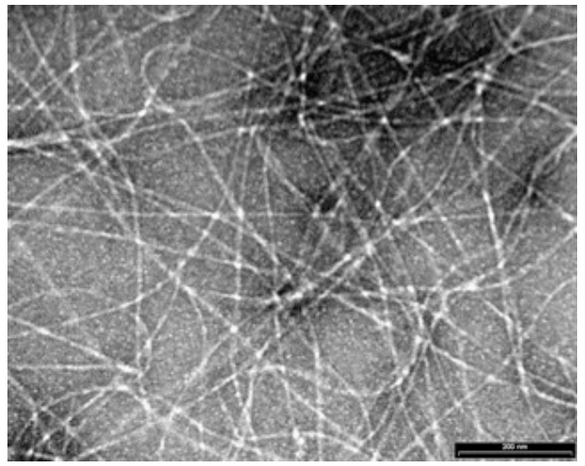


Illustration of how β-sheet-like stacking by individual misfolded proteins could lead to an **amyloid fibril** (note how the association between the monomers depends on **backbone** hydrogen bonding)

The fibrils can be large enough to be seen **pathologically**





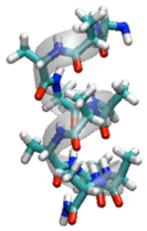
Many different diseases are *associated* with such fibrils

Clinical syndrome	Fibril subunit
Primary systemic amyloidosis	immunoglobulin light chains*
Type II diabetes	islet amyloid polypeptide
Familial Mediterranean fever	serum amyloid A protein*
Alzheimer's disease	amyloid β-peptide
Transmissible spongiform encephalopathies	prion

There is no clear picture whether the fibrils cause the disease or, for example, whether it is the smaller soluble precursors that are toxic

Nature (2003) **426** 900

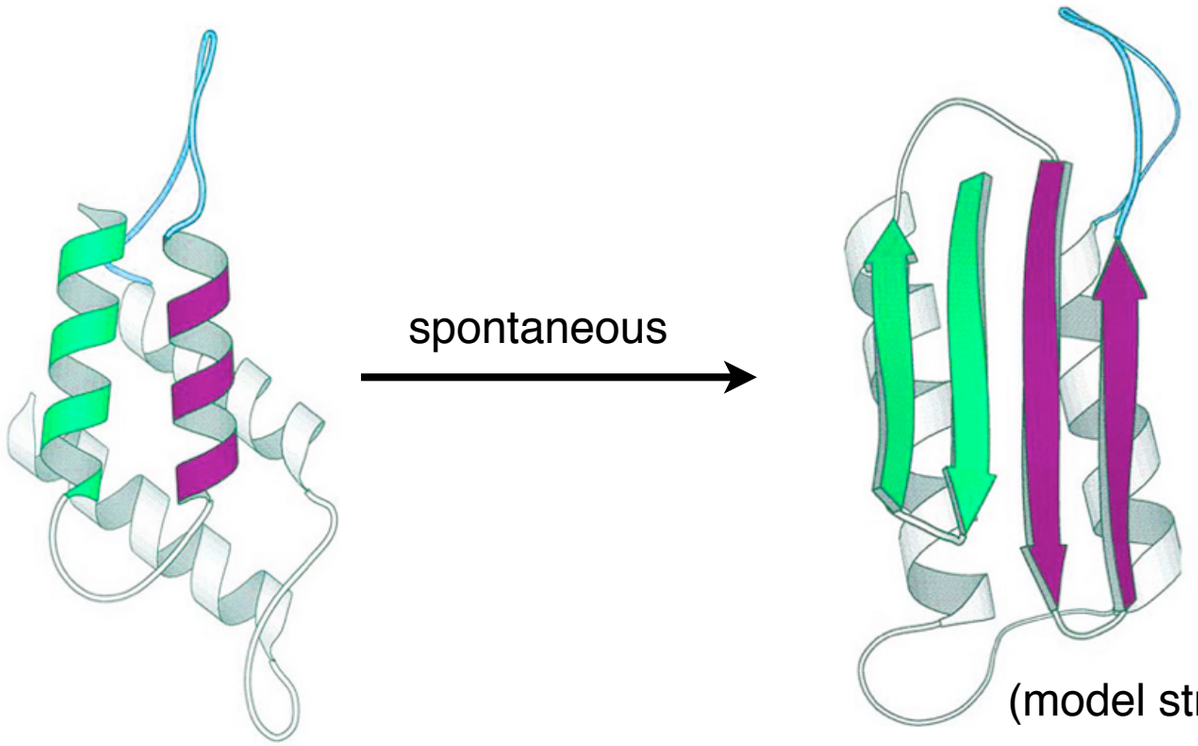
*fragments of



Prions are infectious proteins

PrP^{C}

PrP^{Sc}



spontaneous

(model structure)

Innoculation of mice with the prion protein PrPSc from scrapie infected mice induces normal cellular PrP^{C} to convert to PrP^{Sc} in an auto-catalytic reaction

PrP^{Sc} can form fibrils

The infectious agent of transmissible spongiform encephalopathies (BSE, Scrapie, vCJD*) is a protein (the **prion**)

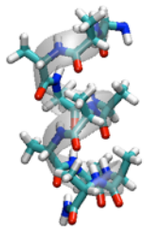
Prion are harder to destroy than viruses and bacteria. For example they are resistant to ionising and UV radiation but are susceptible to substances that destroy proteins such as proteases and phenol

Not all prions cause diseases: e.g. [PSI+] in yeast**.

This is the prion form of SUP35, a protein release factor (*E. coli* RF3) and is not transmitted from cell to cell but is inherited by daughter cells in a non-Mendelian fashion. [PSI+] enhances the suppression of nonsense codons.

** Cell (1997) **89** 495

* mad cow's disease. See Science (1997) **278** 245



IBM Blue Gene

In December 1999 IBM announced it was developing a new series of supercomputers - the **BlueGene** series

TOP 10 Systems - 11/2009

1	Jaguar - Cray XT5-HE Opteron Six Core 2.6 GHz
2	Roadrunner - BladeCenter QS22/LS21 Cluster, PowerXCell 8i 3.2 Ghz / Opteron DC 1.8 GHz, Voltaire Infiniband
3	Kraken XT5 - Cray XT5-HE Opteron Six Core 2.6 GHz
4	JUGENE - Blue Gene/P Solution
5	Tianhe-1 - NUDT TH-1 Cluster, Xeon E5540/E5450, ATI Radeon HD 4870, Infiniband
6	Pleiades - SGI Altix ICE 8200EX, Xeon QC 3.0 GHz/Nehalem EP 2.93 Ghz
7	BlueGene/L - eServer Blue Gene Solution
8	Blue Gene/P Solution
9	Ranger - SunBlade x6420, Opteron QC 2.3 Ghz, Infiniband
10	Red Sky - Sun Blade x6275, Xeon X55xx 2.93 Ghz, Infiniband

Second generation: BlueGene/P at Jülich, Germany

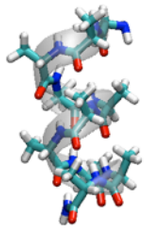
294 912 CPUs

0.8 / 1.0 PFlops (x 10^{15} operations per second)

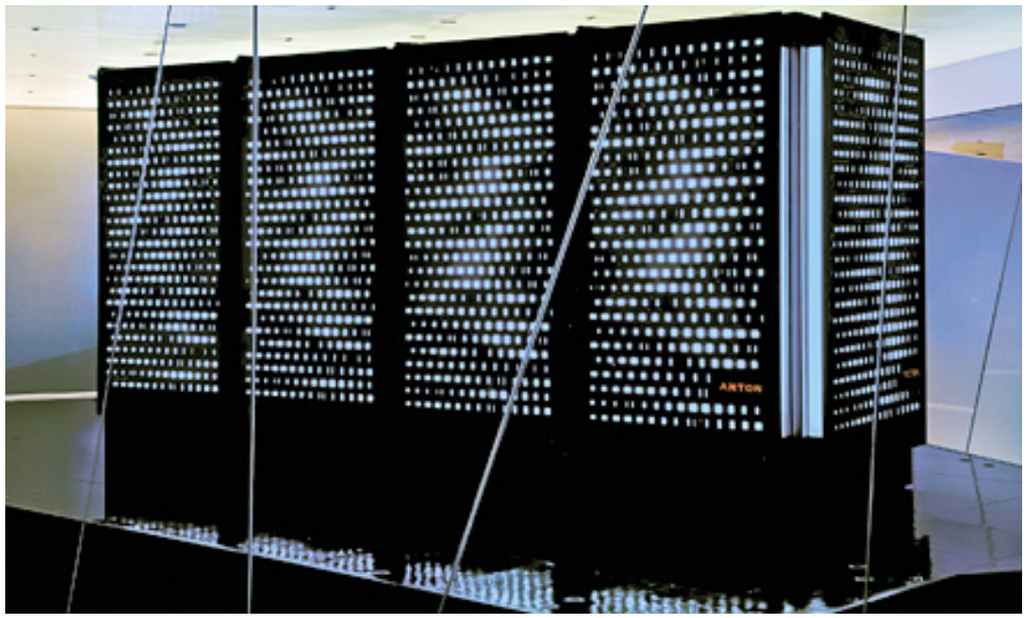
The original goal was a BlueGene machine would be able to fold **one protein per year** (i.e. simulate ms - s of dynamics of a single protein)

This gives you an idea of how hard the protein structure prediction problem is

http://www.top500.org/



DE Shaw Research



one of 11 ANTONs

David Shaw, a hedge fund manager, set up his own research institute in New York, DE Shaw Research

In 2007, DEShaw Research designed and manufactured a supercomputer specifically designed to run molecular dynamics of proteins (ANTON)

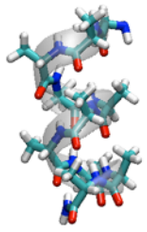
They also wrote their own MD code to run on the machine

As a result it is far faster than any other single computer

In 2010 they folded a fragment of the villin headpiece (68 μs) and a fast folding WW domain (38 μs) and their results compared well with experimental structures*.

They have since examined the dynamics of GPCRs, voltage-gated ion channels and a receptor tyrosine kinase (RTK).

* Science (2010) **330**:341

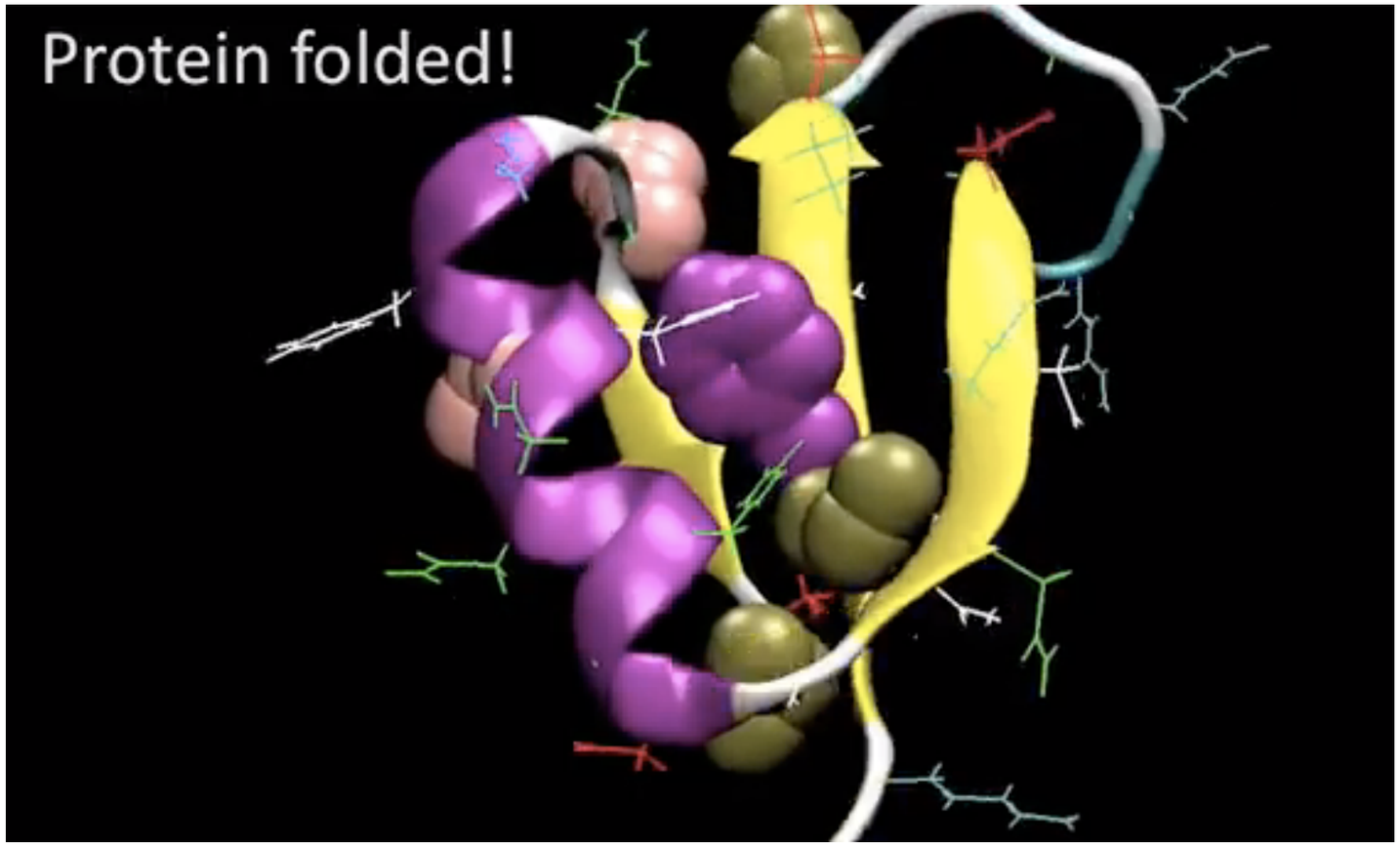


folding@home

UNIVERSITY OF OXFORD

http://folding.stanford.edu

Simulations of protein folding...

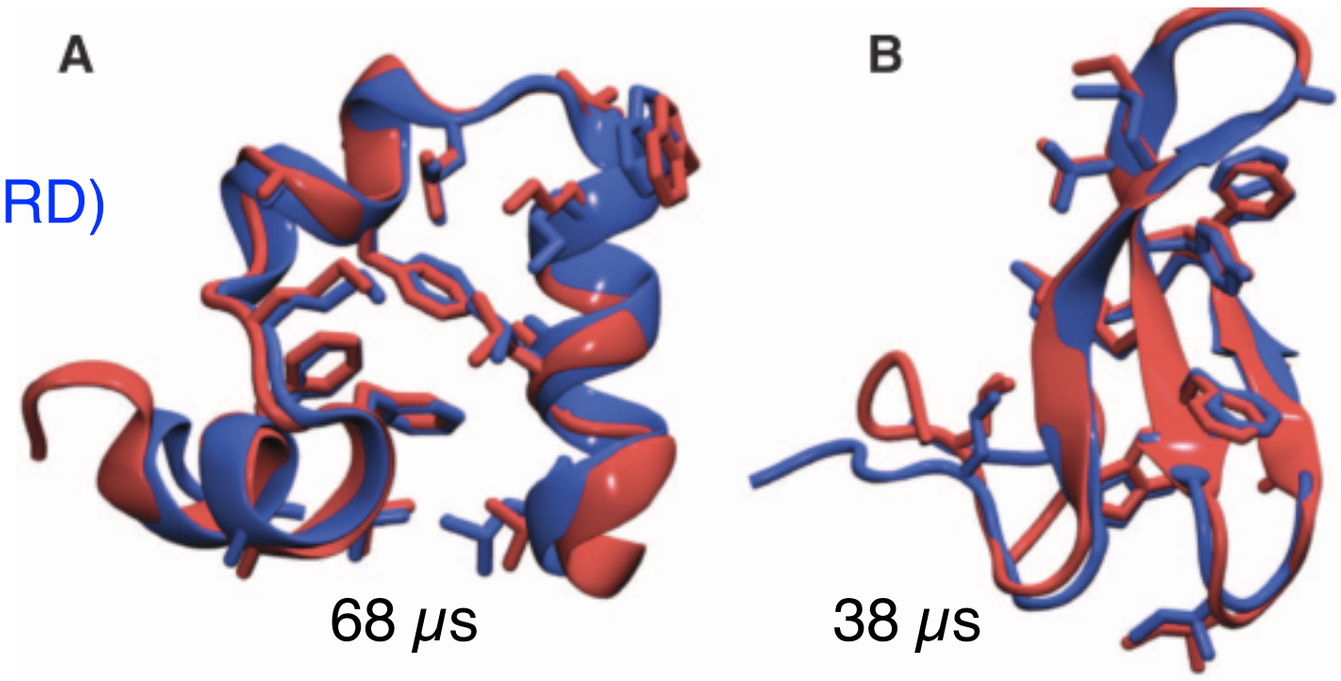


from folding@home

http://www.youtube.com/watch?v=gFcp2Xpd29I

expt (XRD)

MD



from DEShaw (Shaw2010)