DNA/Protein structure-function analysis and prediction

- Protein Folding and energetics:
  - Introduction to folding
  - Folding and flexibility (Ch. 6)
  - Energetics and Thermodynamics
Active protein conformation

- Active conformation of protein is the native state
- unfolded, denatured state
  - high temperature
  - high pressure
  - high concentrations urea (8 M)
- Equilibrium between two forms

![Diagram showing denatured and native states of a protein. The denatured state is on the left, and the native state is on the right.](image)
Anfinsen’s Theorem (1950’s)

- Primary structure determines tertiary structure.
  
  In the mid 1950’s Anfinsen began to concentrate on the problem of the relationship between structure and function in enzymes. [...] He proposed that the information determining the tertiary structure of a protein resides in the chemistry of its amino acid sequence. [...] It was demonstrated that, after cleavage of disulfide bonds and disruption of tertiary structure, many proteins could spontaneously refold to their native forms. This work resulted in general acceptance of the ‘thermodynamic hypothesis’ (Nobel Prize Chemistry 1972)."  
  
  www.nobel.se/chemistry/laureates/1972/anfinsen-bio.html

- Anfinsen performed un-folding/re-folding experiments!
Dimensions: Sequence Space

- How many sequences of length $n$ are possible?
  \[ N(\text{seq}) = 20 \cdot 20 \cdot 20 \cdot \ldots = 20^n \]
  e.g. for $n = 100$, $N = 20^{100} \approx 10^{130}$, is nearly infinite.

- The probability $p$ of finding twice the same sequence is
  \[ p = 1/N, \text{ e.g. } 1/10^{130} \]
  is nearly zero.

- Evolution: divergent or convergent
  - sequences are dissimilar,
    even in convergent evolution.
Dimensions: Fold Space

• How many folds exist?
  – Sequences cluster into sequence families and fold families
  – some have many members, some few or only one:

• Using Zipf’s law:
  \[ n(r) = \frac{a}{r^b} \]

• For sequence families:
  \[ b \approx 0.64 \quad \rightarrow \quad n \approx 60000 \]

• For fold families:
  \[ b \approx 0.8 \quad \rightarrow \quad n \approx 14000 \]
Levinthal’s paradox (1969)

- Denatured protein re-folds in ~ 0.1 – 1000 seconds.

- Protein with e.g. 100 amino acids each with 2 torsions (φ en ψ).
  Each can assume 3 conformations (1 trans, 2 gauche).
  \( 3^{100} \times 2 \approx 10^{95} \) possible conformations!

- Or:
  100 amino acids with 3 possibilities in Ramachandran plot (α, β, L):
  \( 3^{100} \approx 10^{47} \) conformations.

- If the protein can visit one conformation in one ps (10^{-12} s),
  exhaustive search costs \( 10^{47} \times 10^{-12} \) s = \( 10^{35} \) s \( \approx \) 10^{27} years!
  (the lifetime of the universe \( \approx \) 10^{10} years...)
Levinthal’s paradox

Protein folding problem:
- Predict the 3D structure from sequence
- Understand the folding process
What to fold?
...fastest folders

Rates: predicted vs experiment

Experiments:
- villin: Raleigh, et al, SUNY, Stony Brook
- BBAW: Gruebele, et al, UIUC
- beta hairpin: Eaton, et al, NIH
- alpha helix: Eaton, et al, NIH
- PPA: Gruebele, et al, UIUC

Predictions:
- Pande, et al, Stanford
Molten globule

- First step: hydrophobic collapse
- Molten globule: globular structure, not yet correct folded
- Local minimum on the free energy surface
Folded state

- **Native state** = lowest point on the free energy landscape

- Many possible routes
- Many possible local minima (misfolded structures)
DNA/Protein structure-function analysis and prediction

- Protein Folding and energetics:
  - Introduction to folding
  - Folding and flexibility (Ch. 6)
  - Energetics and Thermodynamics
Helper proteins

• Forming and breaking disulfide bridges
  – Disulfide bridge forming enzymes: Dsb
  – protein disulfide isomerase: PDI

• “Isomerization” of proline residues
  – Peptidyl prolyl isomerasers

• Chaperones
  – Heat shock proteins
  – GroEL/GroES complex
  – Preventing or breaking ‘undesirable interactions’…
Disulfide bridges

- Equilibriums during the folding process
Proline: two conformations

- Peptide bond nearly always *trans* (1000:1)

- For proline *cis* conformation also possible (4:1)

- Isomerization is bottleneck, cyclophilin catalyses
Chaperones

• During folding process hydrophobic parts outside?
  – Risk for aggregation of proteins
• Chaperones offer protection
  – Are mainly formed at high temperatures (when needed)
  – Heat-shock proteins: Hsp70, Hsp60 (GroEL), Hsp10 (GroES)
GroEL/GroES complex

- GroEL:
  - 2 x seven subunits in a ring
  - Each subunit has equatorial, intermediate and apical domain
  - ATP hydrolyse, ATP/ADP diffuse through intermediate domain

- GroES:
  - Also seven subunits
  - Closes cavity of GroEL
GroEL/GroES mechanism

• GroES binding changes both sides of GroEL
  – closed cavity
  – open cavity

• cycle
  – protein binds side 1
  – GroES covers, ATP binds
    – ATP → ADP + Pi
  – ATP binds side 2
  – ATP → ADP + Pi
    • GroES opens
      • folded protein exits
      • ADP exits
  – New protein binds
Alternative folding: prions

- Prion proteins are found in the brains
- Function unknown
- Two forms
  - normal alpha-structure
  - harmful beta-structure
- beta-structure can aggregate and form ‘plaques’
  - Blocks certain tissues and functions in the brains
Protein flexibility

• Also a correctly folded protein is dynamic
  – Crystal structure yields average position of the atoms
  – ‘Breathing’ overall motion possible
B-factors

- The average motion of an atom around the average position
Conformational changes

- Often conformational changes play an important role for the function of the protein
- Estrogen receptor
  - With activator (agonist) bound: active
  - With inactivator (antagonist) bound: not active
Allosteric control

- Often two conformations possible
  - active T(ense) en inactive R(elaxed)

- Modulators change the conformation in the active form (or the inactive form)

- Not bound to active site: allosteric control
DNA/Protein structure-function analysis and prediction

• Protein Folding and energetics:
  – Introduction to folding
  – Folding and flexibility (Ch. 6)
  – Energetics and Thermodynamics
Folding energy

- Each protein conformation has a certain energy and a certain flexibility (entropy)
- Corresponds to a point on a multidimensional free energy surface

Three coordinates per atom
3N-6 dimensions possible

\[ \Delta G = \Delta H - T \Delta S \]

may have higher energy but lower free energy than coordinate x
Peptide folding from simulation

- A small (beta-)peptide forms helical structure according to NMR

\[
\begin{array}{c}
\text{H}_2\text{N} & \text{C} \text{H} & \text{O} & \text{N} \\
\text{N} & \text{H} & \text{O} & \text{H}_2\text{N} \\
\text{H} & \text{O} & \text{N} & \text{H} \\
\text{O} & \text{N} & \text{H} & \text{O} \\
\text{O} & \text{N} & \text{H} & \text{O} \\
\text{N} & \text{H} & \text{O} & \text{H} \\
\end{array}
\]

- Computer simulations of the atomic motions: molecular dynamics
Folding and un-folding in 200 ns

Unfolded structures

Folded structures

all different?

how different?

$3^{21} \approx 10^{10}$ possibilities!

all the same
Temperature dependence

The graph shows the temperature dependence of folding equilibrium. The folding equilibrium depends on temperature, as indicated by the chart which displays RMSD over time at various temperatures: 360 K, 350 K, 340 K, 320 K, and 298 K. The chart uses different colors to distinguish between unfolded and folded states.
Pressure dependence

![Graph showing RMSD over time at different pressures](image)

- **2000 atm**
- **1000 atm**
- **1 atm**

- **folding equilibrium depends on pressure**
Surprising result

- Number of relevant non-folded structures is very much smaller than the number of possible non-folded structures.

<table>
<thead>
<tr>
<th></th>
<th>Number of aminoacids in protein chain</th>
<th>Folding time (exp/sim) (seconds)</th>
<th>Number possible structures</th>
<th>Number relevant structures</th>
</tr>
</thead>
<tbody>
<tr>
<td>peptide</td>
<td>10</td>
<td>$10^{-8}$</td>
<td>$3^{20} \approx 10^9$</td>
<td>$10^3$</td>
</tr>
<tr>
<td>protein</td>
<td>100</td>
<td>$10^{-2}$</td>
<td>$3^{200} \approx 10^{90}$</td>
<td>$10^9$</td>
</tr>
</tbody>
</table>

- If the number of relevant non-folded structures increases proportionally with the folding time, only $10^9$ protein structures need to be simulated in stead of $10^{90}$ structures.
- Folding-mechanism perhaps simpler after all…
Main points

- Anfinsen: proteins fold reversibly!
- Levinthal: too many conformations for fast folding?
  - First hydrophobic collapse, then local rearrangement
    - Protein folding funnel
  - Assistance with protein folding
    - Sulphur bridge formation
    - Proline isomerization
    - Chaperonins
- Intrinsic flexibility: Breathing / Conformational change
  - Conformational changes for
    - Activation / Deactivation
    - Allosteric modulation
- Dynamics:
  - Simulations of reversible folding of a peptide