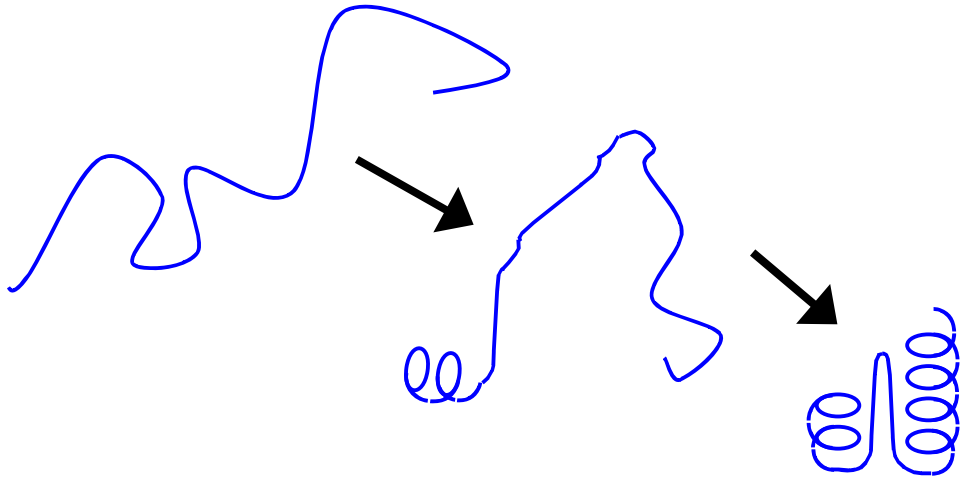
Protein Folding

Andrew Torda July 2008 67.912 Struktur & Simulation

Question..

- how does a protein do this ?



Background / stories

- from biochemistry Übungen (protein folding – easy)
  - take lysozyme / ribonuclease...
  - put in 8M urea (unfolds)
  - remove urea (refolds)
- conclusion ?
  - the protein sequence is all you need to fold a protein
  - is this true ? Not always

Alternative (logical reasoning)

- protein folding should be impossible…

Protein folding should be difficult

from simple theory – Levinthal's paradox

- each amino acid has 2 or 3 or N conformations
- for a protein of M residues, it should visit N^M
- what if it spends 10^{-15}s at each conformation ?
- time to find one conformation for N=3 and 100 residues
- 3^{100} x 10^{-15} s about 1.6 x 10^{25} years
- is this serious ? useful ?
  - not correct approach to problem
  - proteins cannot be exploring space randomly

Who cares about protein folding ?

Religious belief

- if we could understand folding we could
  - predict structure
  - design proteins that fold better (more stable)
  - identify essential residues for folding (not suitable for mutagenesis)

Issues / questions

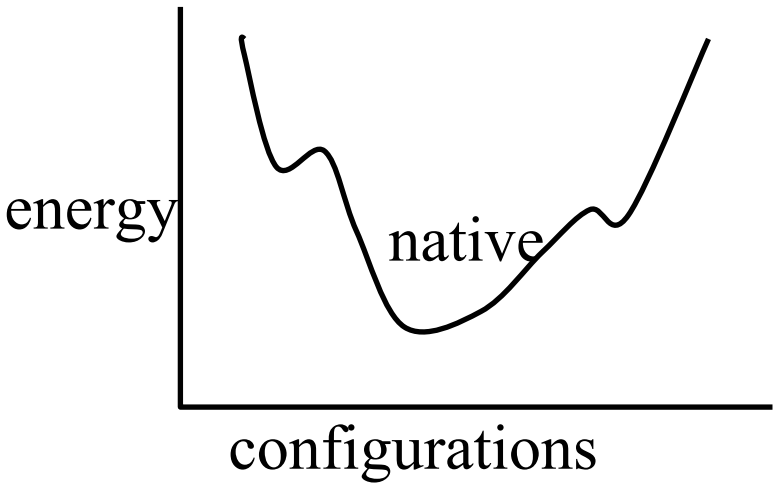
- Kinetic versus thermodynamic
- What order do events happen in ? (collapse vs secondary struct)
- Is unfolding the same as folding ?
- Is folding in a test tube the same as nature ?
- Is unfolding in a test tube the same ?

Are proteins in energy minima ?

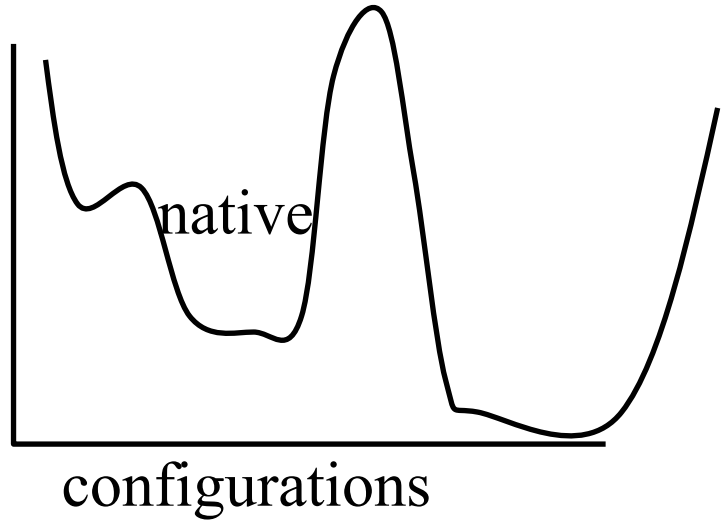
- Anfinsen story..
  - proteins can be unfolded and refolded alone
  - all the information is in the sequence
  - native conformations are the (free) energy minimum
  - thermodynamic belief
- more modern
  - many many proteins cannot be refolded in the lab
    - consequence .. maybe they need something else
    - maybe they are not always in free energy minimum
- kinetically trapped proteins..

Kinetic versus thermodynamic

- If proteins fold spontaneously and remain folded, they are thermodynamically determined
- If you leave a protein long enough and it unfolds, it was not in an energy minimum



thermodynamic



kinetically
trapped

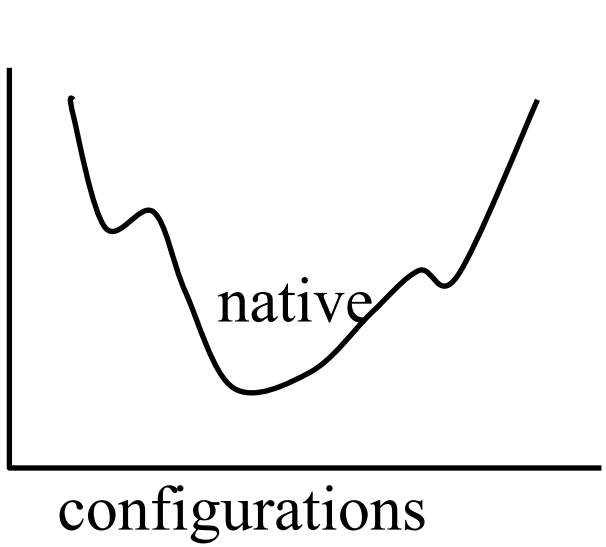
- is this purely academic ?

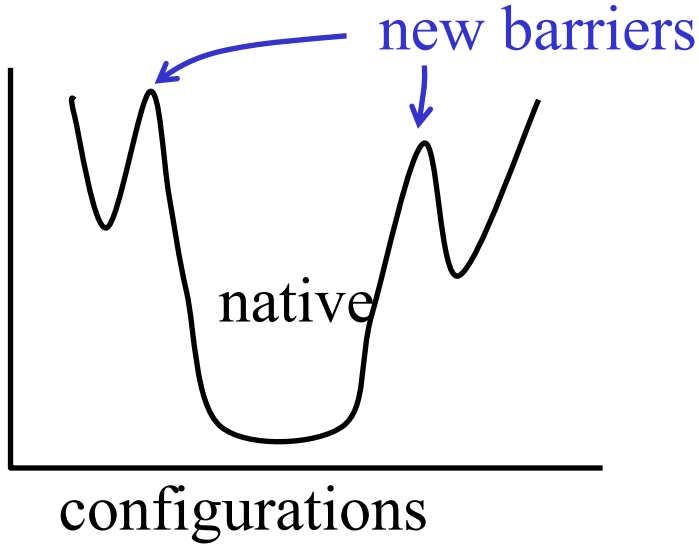
Consequences

- Thermodynamic
  - conforms to classic view
  - protein structure prediction
    - just a matter of modelling the real world
- Kinetically trapped
  - we cannot predict structure from sequence just by energies !
- Can we see which is the case ?
  - leave a protein for 10 minutes and see if it finds another state
  - for 10 years ? No.
  - depends on barriers
- Some stories
  - some evidence of kinetic trapping
  - some proteins do have other states
    - fibrils, Alzheimers, mad cow disease

Evolution / design consequences

- imagine I model physics perfectly and can predict structure
- I design a better more stable protein





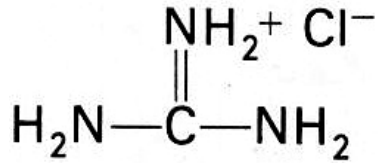
- my new protein may be more stable
- it may never be found
- evolutionary implications
  - protein sequences may evolve for folding (+ structure and function)

Change of direction

- enough background on folding
- brief overview of experiments
  - no exam questions on this

Experiments

- What timescales do we think of ?
  - maybe 10^{-6} for folding
  - maybe orders of magnitude slower (sometimes faster)
- Experimental approaches
  - force protein to unfold
    - chemistry (guanidinium HCL, 8M urea)
    - temperature (heat, cold)
    - change conditions and watch
  - try to measure very fast
  - try to change timescale
  - try to measure unfolding



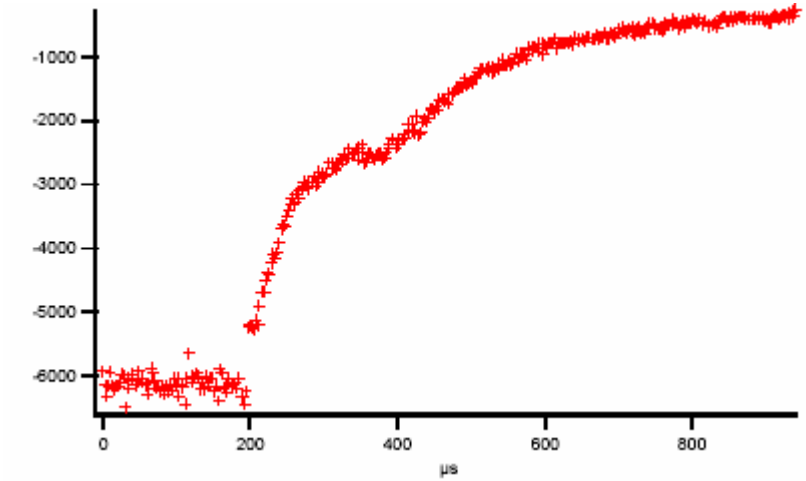
Some real time experiments

- temperature jump
- stop - flow
- fluorescence
- NMR
- circular dichroism (CD)

Temperature jump

Background..

- protein is cold (2˚)
- bang with infra-red laser
- follow with trp fluorescence
- shortest time ?
  - about 250 ns
- main information ?
  - kinetics
  - not much specific structure

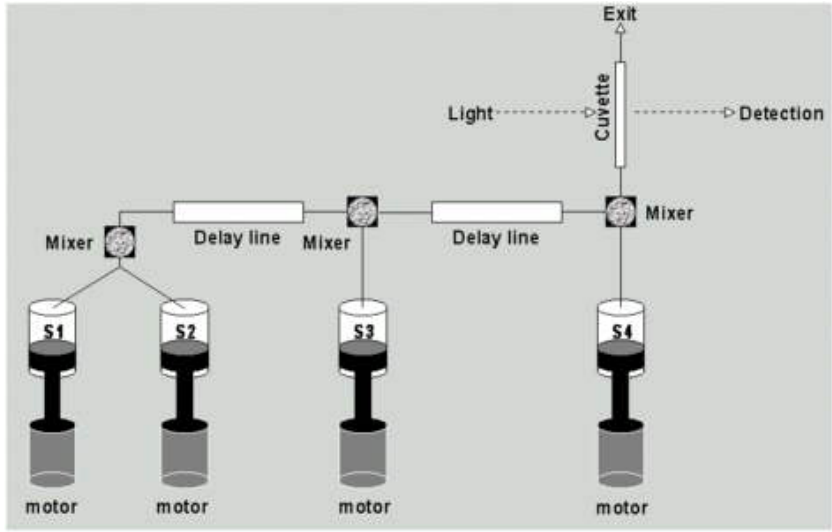


picture from Dimitriadis, G, Drysdale, A, Radford, S and Smith, A, "Investigating early events in protein folding by laser induced temperature jump"

Stop flow

- Start from chemically unfolded protein
- Use quick mixing / change of conditions to refold

- example
  - lysozyme + guanadinium HCL
  - suddenly dilute in buffer
  - protein refolds

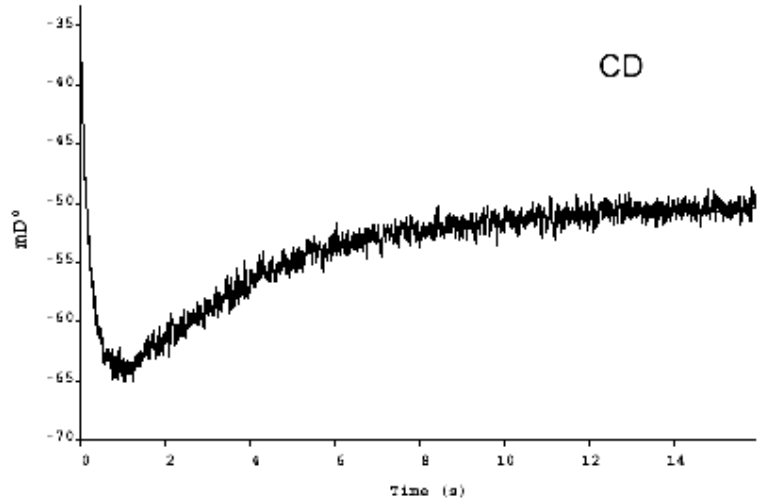


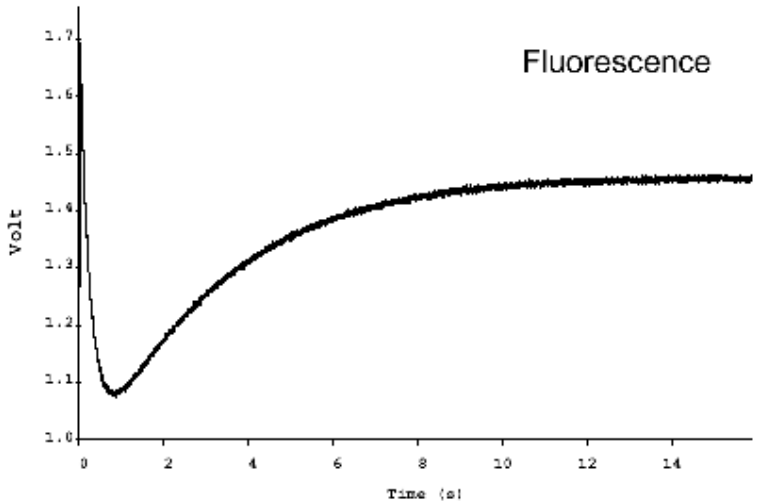
pictures from http://www.molecularkinetics.com/sf/SFM.html

Stop flow and spectroscopy

watch with

- circular dichroism (follow secondary structure)
- fluorescence
  - (absorption, re-emission, polarisation)
- time scale ?
  - $> 10^{-3}$ s
- difficulty ?
  - fast mixing / dilution





pictures from http://www.molecularkinetics.com/sf/SFM.html

NMR

- most detailed structural information
- problem ?
  - not a sensitive spectroscopy
    - needs more sample
    - slower (signal averaging)
- advantages
  - most detailed structure information
- timescale
  - minutes to hours for details
  - maybe some seconds for 1-D spectra
  - fastest for limited kinds of information
    - put in a few labels (^{19}F) – peaks can be recorded quickly

What can one do with kinetic data

- information may not be very specific (structurally)
- combine with chemistry / molecular biology
- put labels on sites and repeat folding
  - fluorescence labels
  - NMR labels (^{19}F, other techniques)
- molecular biology
  - mutate residues and find out
    - which residues change kinetics

Equilibrium / steady state approaches

- Time resolved approaches are very hard
- give up and use an equilibrium approach
- examples – temperature or chemistry
- temperature
  - at 27˚, protein is folded, at 70˚, unfolded
  - record NMR spectra at 25˚, 35˚, 45˚
  - no problem with time
  - should be able to follow which parts of spectrum disappear
- chemistry
  - protein is folded in PO_4 buffer, acetate buffer, ...
  - unfolded in 8M urea, guanidinium HCL
  - record spectra at various concentrations

What do we know

Some possibilities

- proteins form secondary structure first
  - helices and sheets then arrange themselves

OR

- hydrophobic collapse
  - hydrophobic residues find each other
  - backbone rearranges and secondary structure is fixed

OR

- some key residues interact
- then comes secondary structure and hydrophobic core

OR

- some mixture

OR

- different proteins behave differently / there are no rules

Why is experiment difficult

- Technical difficulties (obvious)
- Tradeoff
  - fast methods – less information
  - more information – too slow
- How real is it ?
- Imaginary technique :
  - I can take any protein in denaturant
  - suddenly bring back native conditions
  - follow every detail
  - is this what happens in nature ?...

How real is experiment ?

Our bodies – about 150 mM salt, regulated pH, temperature, ...

If I denature a protein with high salt

- is the partially folded state natural ?
  - it comes from disrupting a very special set of ionic interactions

If I heat the protein

- guaranteed to visit high energy states which are not natural

Hope !

- the strongest interactions are formed first – last to be broken

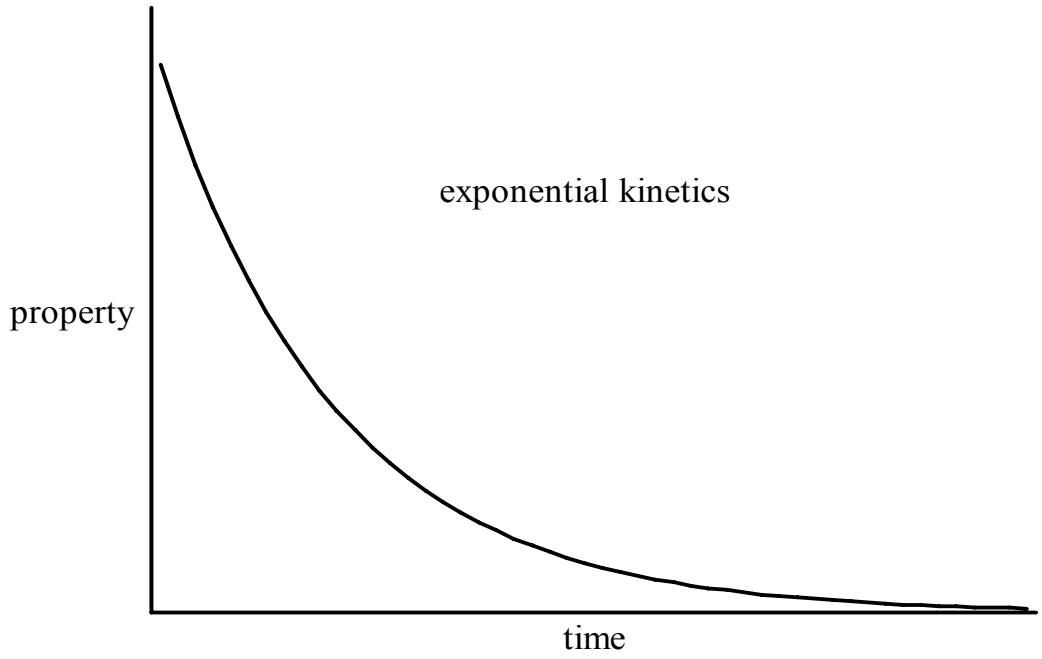
Do proteins fold like this in nature ?

- proteins made from N to C terminus
- N terminus gets a chance to find structure, before rest of protein is there
- would permit very specific paths / kinetic trapping

Next ... simulation and theory...

General kinetics

- What have we seen so far ?
  - most properties have something like exponential decay
  - property = $a\ e^{\alpha\ t}$
  - rate of change proportional to quantity present
- whatever model should explain at least this



Monster Simulations

- Brute force long MD examples
  - 36 residues - months x 256 processors 10^{-6} s
    - does not fold to native state
    - reasons ?
      - force field no good ?
      - simulation time not long enough ?
  - IBM blue gene (Pitera and Swop)
    - 20 residues – implicit solvent 92×10^{-9}
    - quite a good answer
  - Distributed – folding @ home
    - some impressive results – bounds on folding times

More feasible

- What can be simulated with MD
  - peptides (not here)
  - unfolding

Simulating unfolding

- Atomistic simulation of real protein too slow
  - take native structure at 300 K
  - gradually heat up
  - watch it fall apart
  - what breaks first ?
    - secondary structure ?
    - overall fold ?
    - everything ?
- Reasons to believe
  - the last interactions to form (folding) may be first to break (unfolding)

Problems simulating unfolding

Problems

- the system is visiting high energy states which may not really exist
- force fields are parameterised for 300 K
- property of unfolded state(s)
  - statistics may be dominated by huge number of partially folded states (more later)

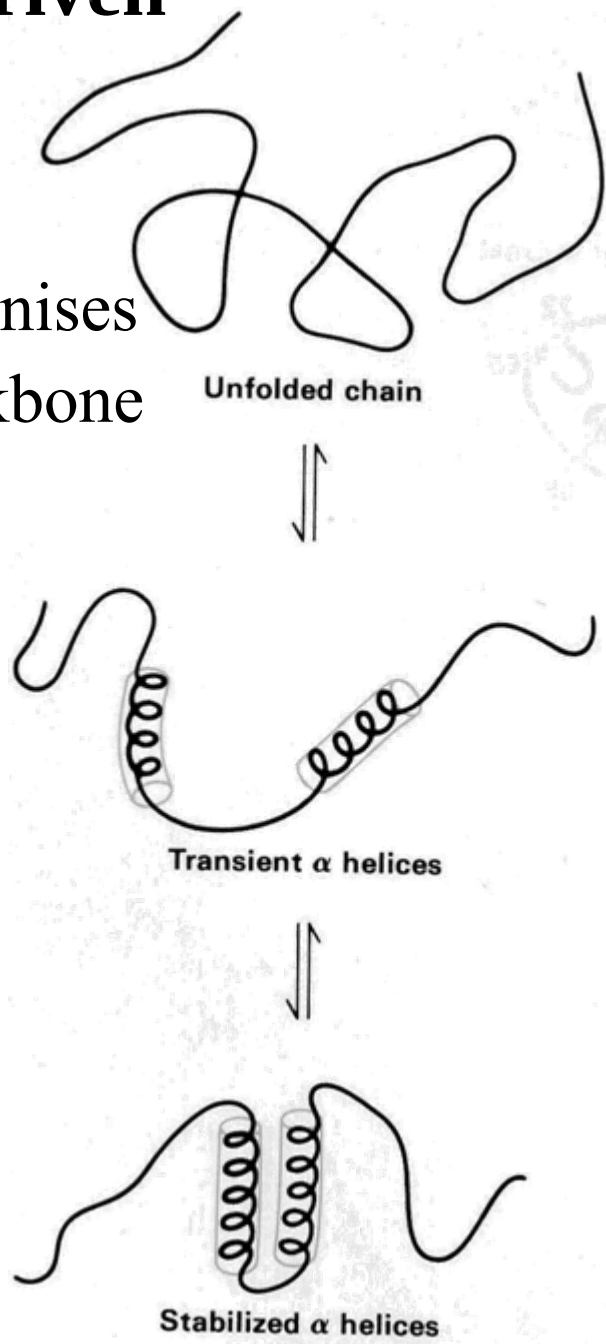
Forgetting atomistic detail

- What are questions we can ask ?
- What can we guess without any calculations ?
- Questions
  - is there such a thing as a folding pathway ?
  - how should we look at folding ?
    - secondary structure forms first and is rearranged
    - hydrophobic residues come together and then secondary structure forms ?
    - a few important contacts are formed, then structure forms

Side chain vs backbone driven

Textbook

- local secondary structure forms, then reorganises
- secondary structure depends largely on backbone
- Alternative
  - sidechains are very important



Stryer, L. Biochemistry, 1981, WH Freeman, "Biochemistry", page 36

Sidechains might be important

- backbone view does not predict collapse of protein
- α-helix and β-strand propensity is weak
  - examples from last semester
  - isolated peptides are not stable
  - β-strands often depend on long-range H bonds
- helix / strand formation depends on environment / solvent and is not known in open structure
- fold is largely predictable / characterised by pattern of HP (sidechains)
- proteins are most sensitive to mutations in core (they are important for stability)

taken from Dill, K.A., Protein Sci., 8, 1166-1180, 1999, Polymer principles and protein folding

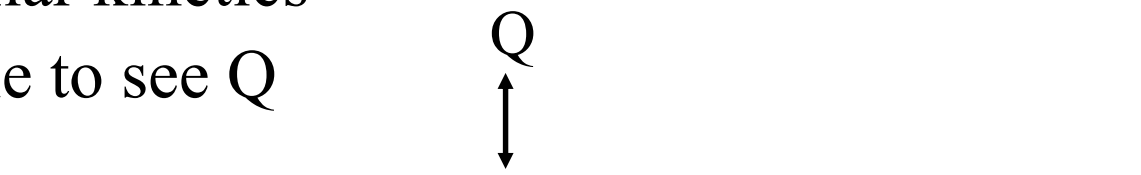
Types of pathway

From classical chemistry we would like a path

unfolded -> A-> B-> C-> folded (native)

Slightly more complicated

- could still give us similar kinetics
- would expect to be able to see Q



Basic idea

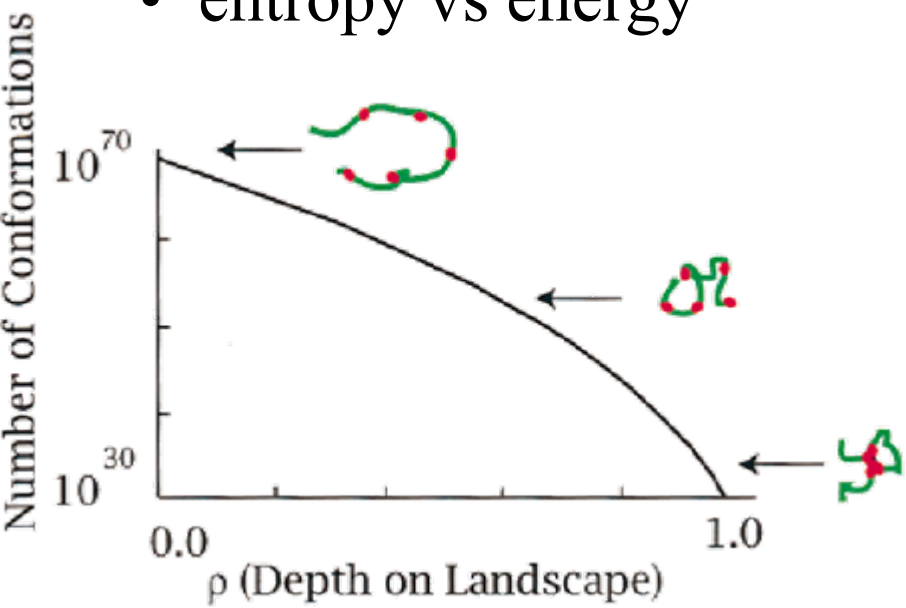
- molecules may get sidetracked, but
  - every molecule sees A, B, C...
- where does it come from ?
  - Levinthal's answer to paradox
    - there must be a preferred pathway
- old view / microstate pathway approach

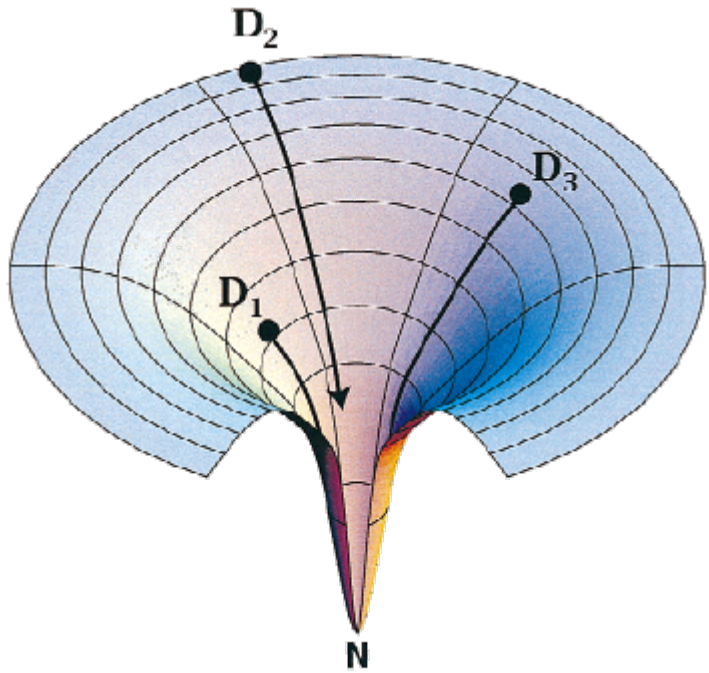
Consequence of simple pathway

- does not really disagree with two state kinetics
  - A or B or C might be part of transition barrier
- pathway with detours explains multi-state kinetics
- does it sound intuitively reasonable ?
  - what if a mutation perturbs A or B or C
    - whole pathway might break
    - maybe OK (this is why some mutants do not fold)

Ensemble view

- conformation space is huge
- will a protein be able to find a neat path through it ?
- should we even look for paths ?
- consider a multitude of paths...
- is this merely a cute picture ?
- first implication...
  - entropy vs energy





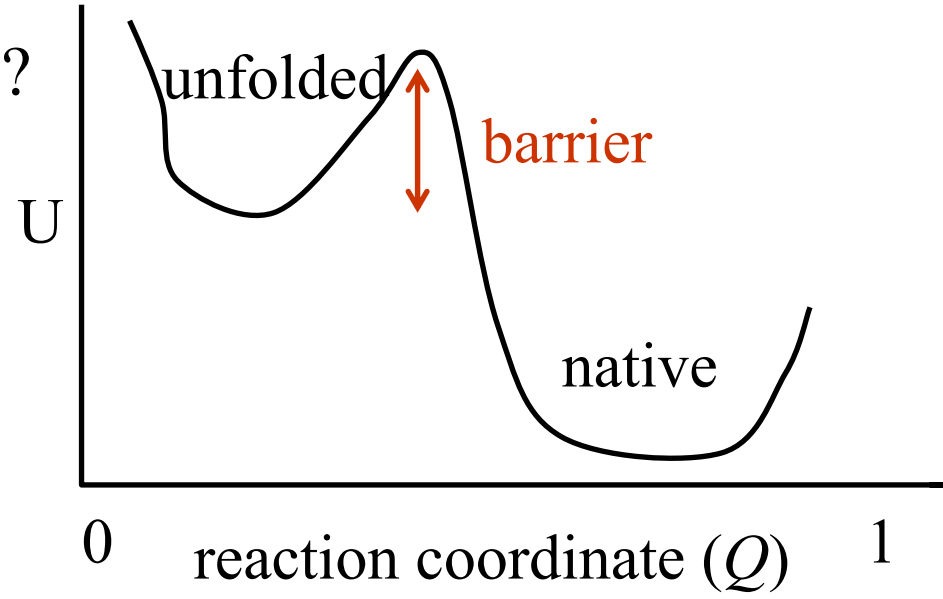
taken from Dill, K.A., Protein Sci., 8, 1166-1180, 1999, Polymer principles and protein folding

Consequence of ensemble view

- As a protein folds
  - potential energy goes down (happy)
  - number of possible states goes down
  - entropy goes down (unhappy)
- interpretation in chemical terms
  - do we have a off-pathway intermediate ? (Q)
- do we have a reaction coordinate ?
  - not a classic one
  - can we invent one ? yes
    - if two atoms are in contact in the final structure
      - native contact
    - Q = number of contacts which are correct

Reaction path

- is this like a chemical reaction ?
  - no
  - many molecules have same Q, but different conformations
- we want at least two state kinetics
  - where does barrier come from ?



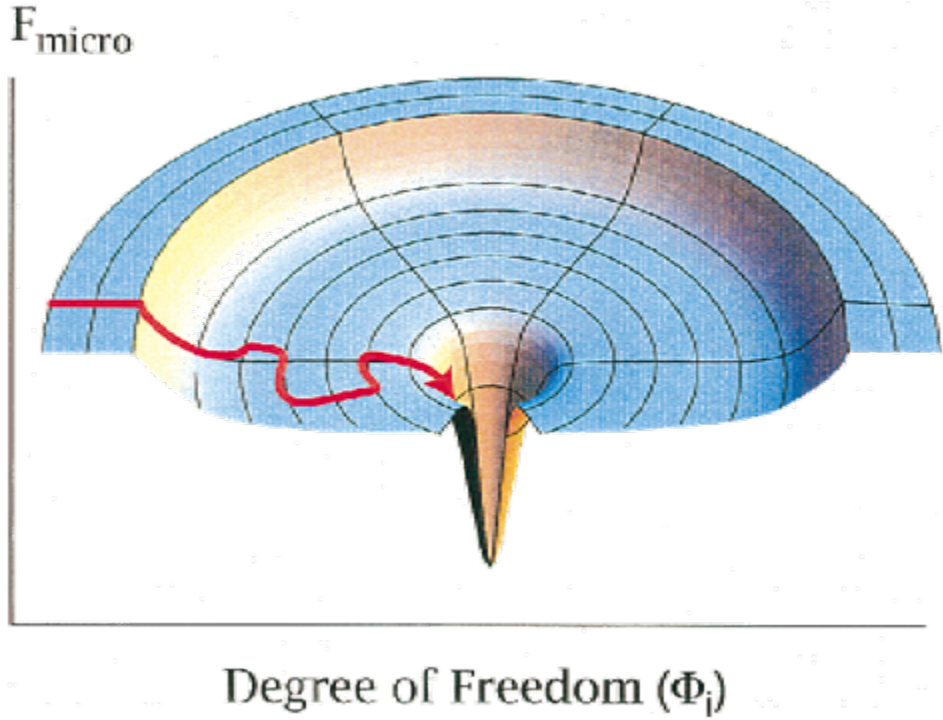
Entropic barriers

Nature cares about free energies

- $\Delta G = U - T\,\Delta S$

If a molecule walks around

- it takes a long time
- looks the same
  as an energy barrier



- Are these pictures useful ?
- Do they agree with calculation ?

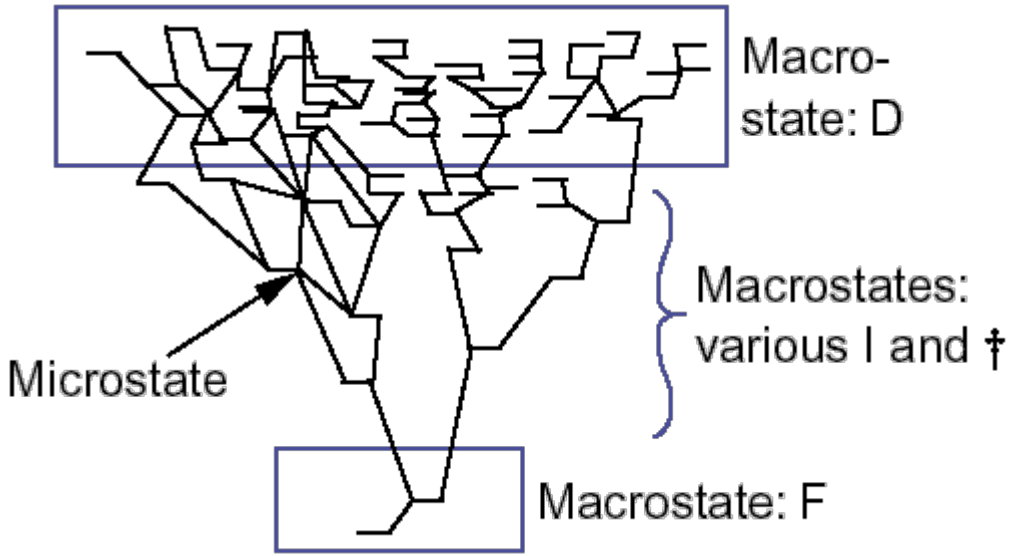
taken from Dill, K.A., Protein Sci., 8, 1166-1180, 1999, Polymer principles and protein folding

Interpretation of landscape

Does this disagree with conventional pathway ?

A->B->C ?

- mostly at early stages
- there is a multiplicity of "A"
- when near native, there are relatively few conformations, so there may be something more like a pathway



Gruebele, M, Curr. Opin. Struct. Biol. 12, 161-168, 2002, Protein folding: the free energy surface

Agreement with other ideas

Agreement with experiment ?

- experiment says most about average properties
  - these are the same in landscape picture
  - should we expect to find well defined, early intermediates ?

Agreement with MD simulation ?

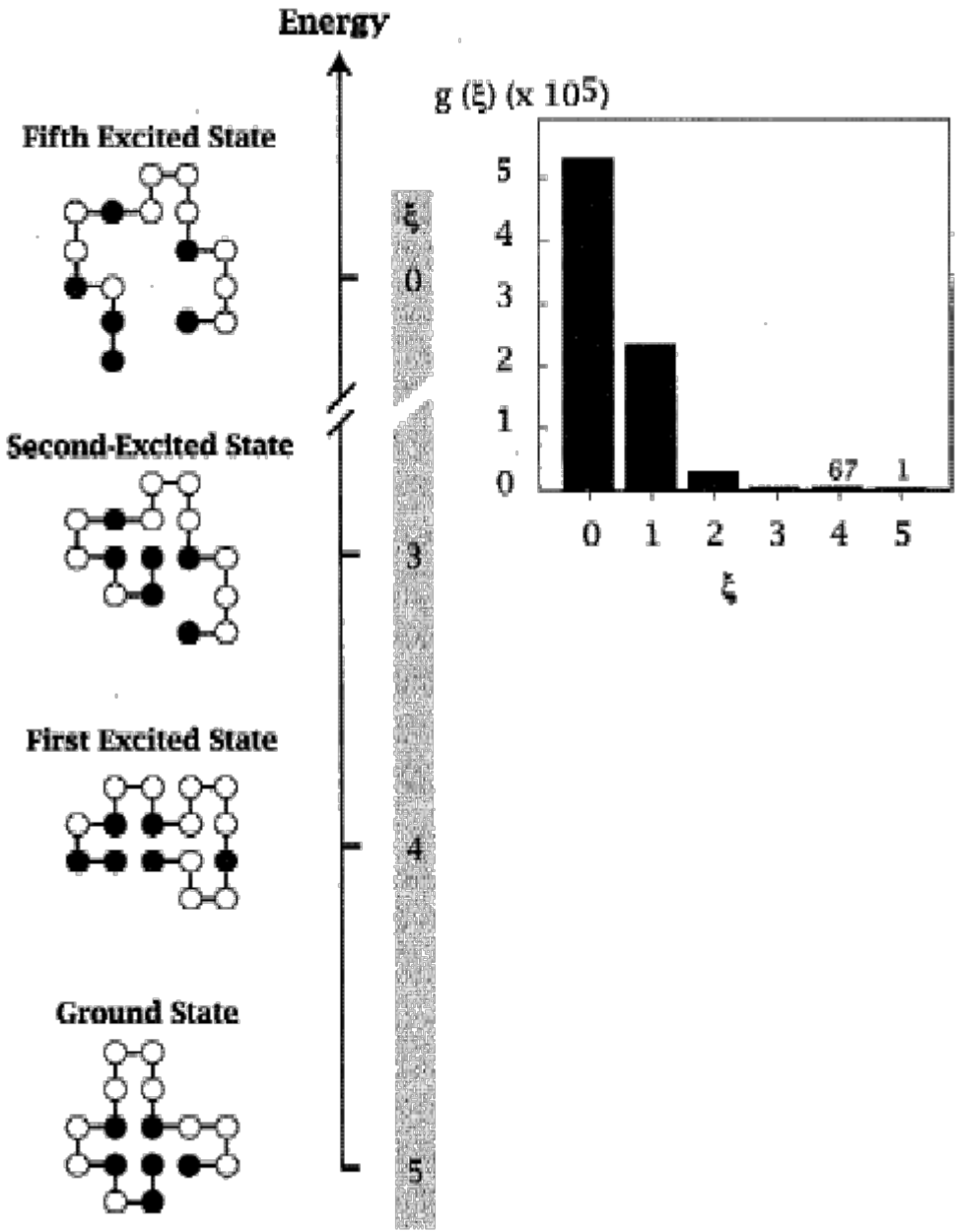
- peptide results – argue that they correspond to near native view

Simple models

Return of HP models

- g(ξ) is density of states
- how many conformations have ξ correct contacts

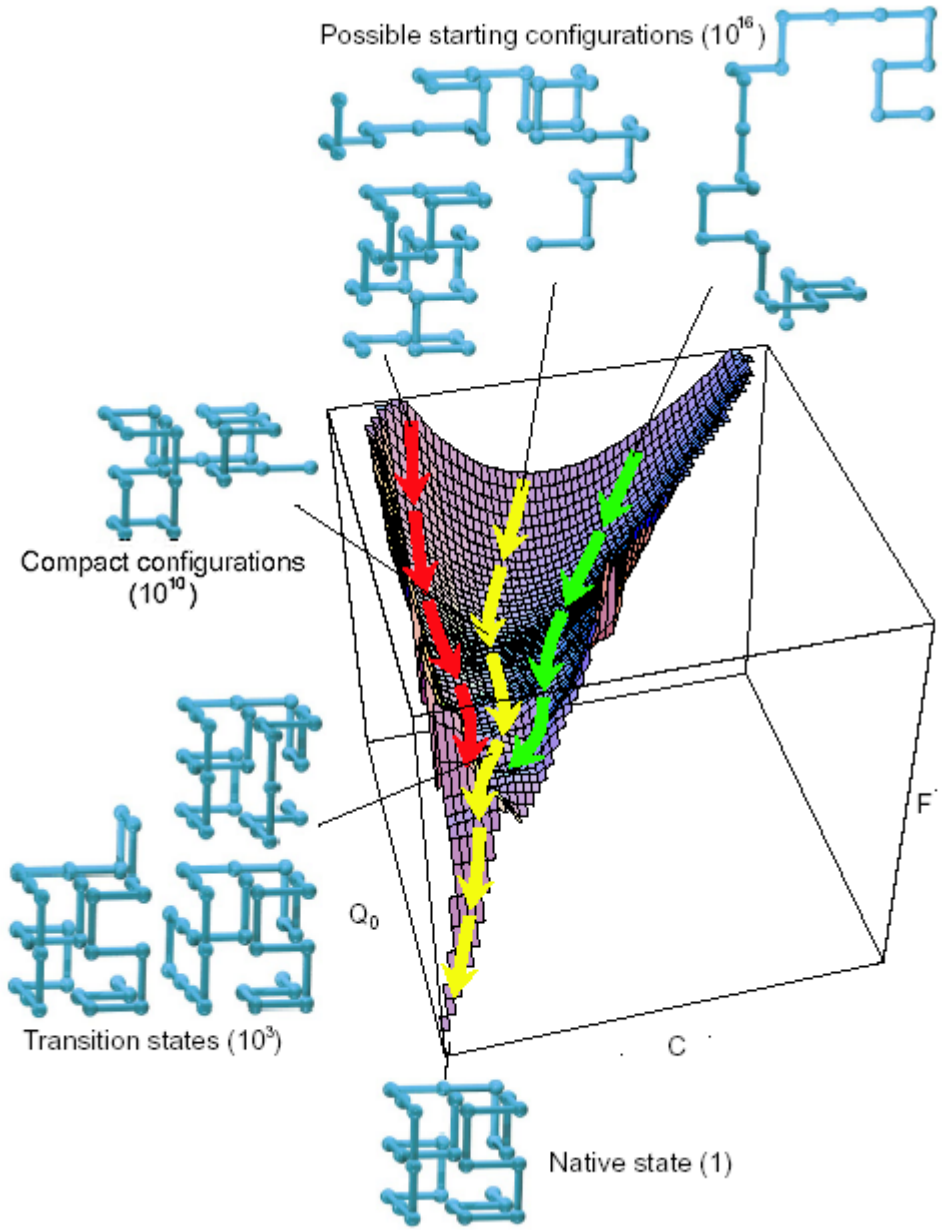
A bigger calculation



taken from Dill, K.A., Protein Sci., 8, 1166-1180, 1999, Polymer principles and protein folding

A larger calculation

- 27 residue
- simple lattice model
- estimations by sampling
  - not exhaustive
  - Q_0 correct contacts
  - C total contacts
  - F free energy



from Dinner, A.R., Šali, A., Smith, L.J. Dobson, C.M., Karplus, M, Trends Biochem Sci. 25, 331-339, 2000, Understanding protein folding ...

(anti) Summary

What we have not covered

- more MD results, especially unfolding
- other views of protein folding (centred about vital residues)
- less physical views
  - kinetic models
  - models based on distances without coordinates

Summary

- Gap between experiment and theory
  - experiment usually gives us averages
  - most calculations look at details
- Very different views on folding may be hard to distinguish
- Folding may not be hierarchical
- Folding may be guided by sidechains
- Early folding may be best modelled by either
  - very crude models (so space can be sampled)
  - forgetting particles and thinking in terms of populations
- Any useful model should agree with exponential kinetics and more complicated variations
- Even an ensemble view should explain results like critical residues