INTERNATIONAL ATOMIC ENERGY AGENCY
UNITED NATIONS EDUCATIONAL, SCIENTIFIC AND CULTURAL ORGANIZATION
INTERNATIONAL CENTRE FOR THEORETICAL PHYSICS
I.C.T.P., P.O. BOX 586, 34100 TRIESTE, ITALY, CABLE: CENTRATOM TRIESTE

H4.SMR/642 - 15

College on Methods and Experimental Techniques in Biophysics

28 September - 23 October 1992

Protein Folding
Transparencies (set no.2)

J.N. ONUCHIC
University of California, U.S.A.

These are preliminary lecture notes, intended only for distribution to participants.

From Last Lecture

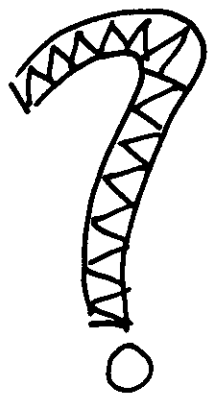
- Anfinsen's experiment '60
- Thermodynamic Hypothesis
- Levinthal's paradox '68

Idea of "Folding Pathway" '83

AS THE ONLY WAY

OF FINDING

THE FOLDED (UNIQUE) STATE

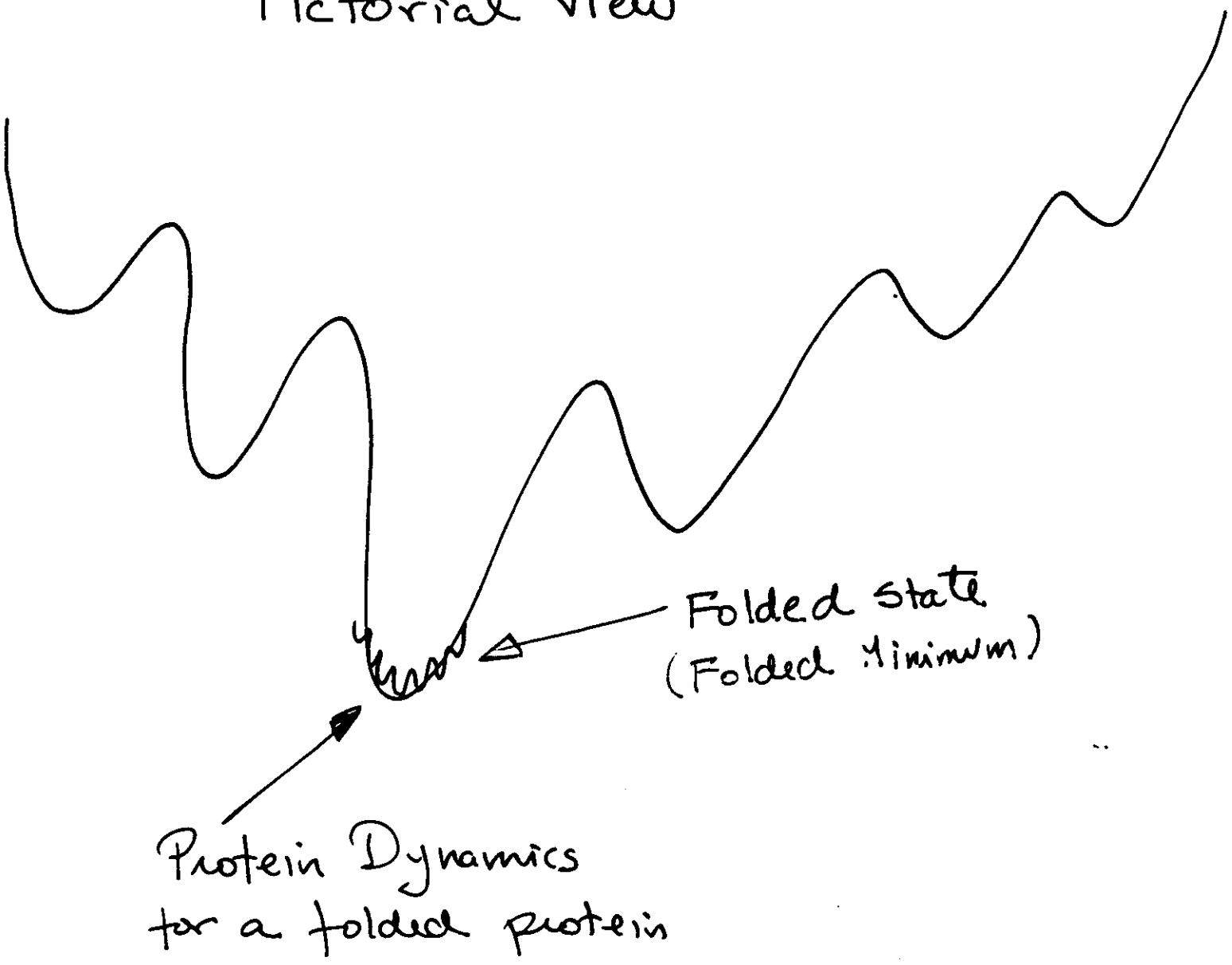


*How the sequence determines the structure?

What does one means by a unique state ?

Proteins fold in a very complex and multidimentional Potential

Pictorial view



A THERMODYNAMIC APPROACH

Experimental Evidence

Assuming only 2 states $\begin{cases} N - \text{Native} \\ U - \text{Unfolded} \end{cases}$

$$K(T) = \frac{[N]}{[U]} = \frac{1-\alpha}{\alpha}$$

α = % unfolded

$$\Delta G^\circ = -RT \ln K \sim -5 - 15 \ \frac{\text{kcal}}{\text{mol}}$$

1) * K can be measured by several techniques
spectral measurements (C.D.), Fluorescence quenching, gradient electrophoresis

2) * depends on T, pH, denaturants (urea) etc

But (FROM THERMODYNAMICAL RELATIONS)

$$\Delta H_{VH} = RT^2 \left(\frac{\Delta \ln K}{\Delta T} \right) \Big|_{\text{around } T_c}$$

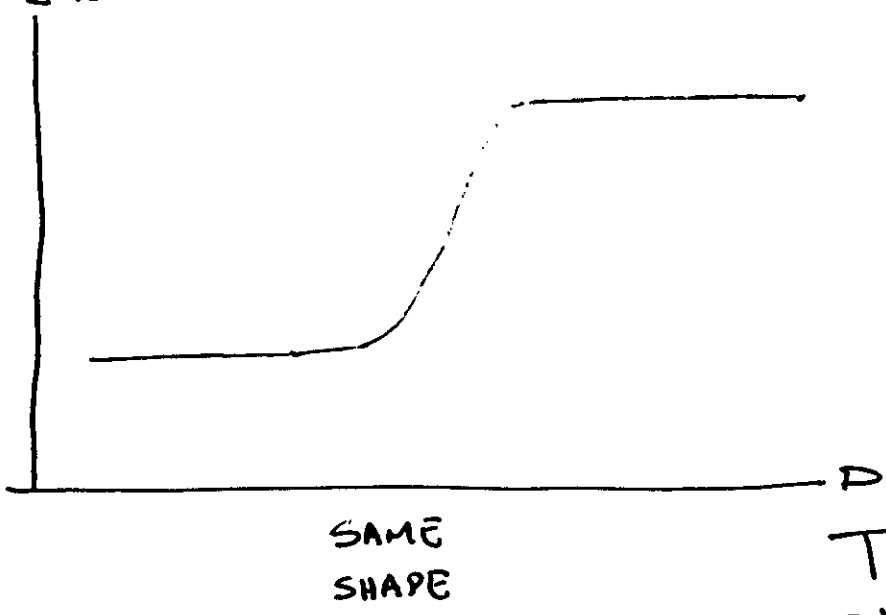
This can be measured by calorimetry

This can be measured by (A)

PERFECT AGREEMENT

ALSO

- % U = α



T_c ~ 40-80°C

pH_c 2-3

INTERESTING POINT

- single amino acid (AROUND ITS PK) – varie pH 2-3 pH units to get 90% change

Protein – 0.3 pH units

ALL IN FAVOR OF

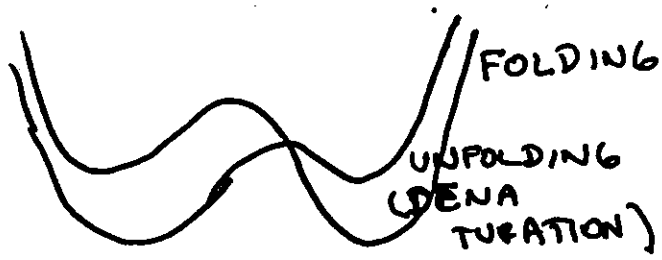
ALL OR NONE MODEL } How to explain this ??

COOPERATIVITY

THE ENTIRE PROTEIN TAKES PART IN THE TRANSITION

Theoretical Models — Thermodynamics

What is the origin of the "First ORDER" barrier?



Two available models:

① Alonso and Dill

- FIRST HYDROPHOBIC COLLAPSE
  INTERMEDIATE STATE ON THE [redacted] CHOSEN PATHWAY IS A RANDOMLY COLLAPSED STATE (FREE ENERGY BARRIER)
- SECOND — REARRANGEMENT OF THIS STATE

PROBLEMS:
- All pathways involve a barrier
- Free energy does not depend on the sequence

$$\Delta G(\rho, \theta, x) = \Delta G_1(\rho) + \Delta G_2(\theta, x)$$

ρ - density of the protein

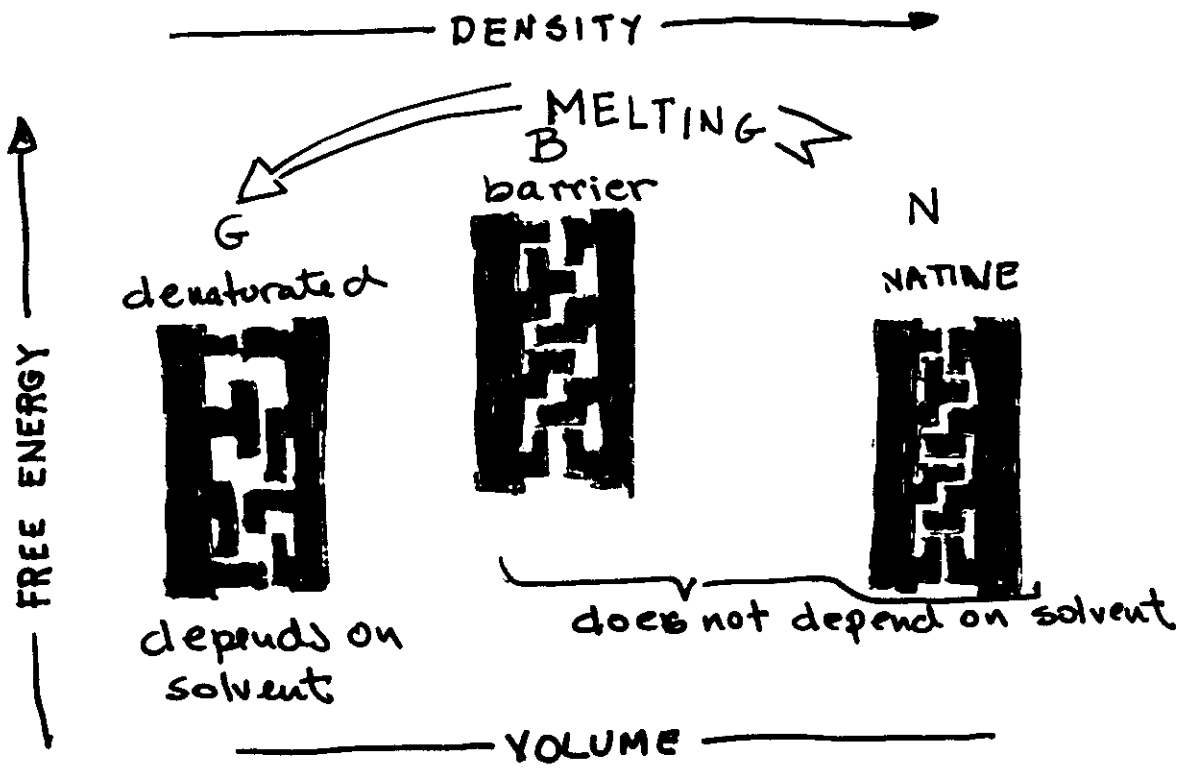
θ - fraction of surface sites occupied by hydrophobic residues

x - fraction of core sites occupied by hydrophobic residues

* NO DIRECT DEPENDENCE ON THE SEQUENCE

(2) Shakhnovich and Finkelstein

Protein denaturation as intramolecular melting



- Denaturation due to "side chain" unfreezing is a first-order phase transition with a compact transition state
- The barrier for unfolding is energetic but for folding is entropic
- The protein in its denatured state is compact with a volume about 30% larger than the native state

↑ This is very controversial!!

Dynamics

ON THE SPIRIT OF THERMO DYNAMICS

MODELS FOR FOLDING
(HYPOTHESIS)

1. Collapse - Reconfiguration Dill
   - A - Hydrophobic Collapse
   - B - Secondary Structure rearrangement

2. Diffusion - Collision Karplus
   - A. Secondary Structure formation
   - B - Tertiary Structure formation

No details given here — sub microsecond simulation with a mean Force Potential.

3. Framework Model Baldwin
   - A - Particular Stable State of secondary Structure is formed
   - B - It acts as a nucleation site

HOW TO TEST THEM ?

Statistical Physics - Bryngelson-Wolynes

USING
- Mean Field Approximation
- Spin-glasses - Random Energy Model

Developed an Analitical Model for Folding

BASED ON

"PRINCIPLE OF THE MINIMAL FRUSTRATION"

TO BEAT LEVINTHAL

Questions:

1 - Who Knows what are spin glasses?

2 - Who khows what is frustration?

3 - Why is any of that relevant to Protein Folding?

How to check these ideas!

How to include sequence-dependence (+) dynamics?

Lattice Simulations

(1) Simple Model — Full enumeration

DILL
- Simple 2D Lattices
- Simple 2-Letter code

WHAT IS A N-LETTER CODE ?

Results — How many non-degenerate minima?

N=10	1024 sequences	6 have non-degenerate minimum
N=17	3404 sequences	no-degenerate minimum

Conclusion ⇒ Compactness induces formation of "secondary structure"

VERY CONTROVERSIAL BUT SUPPORTS Collapse-reconfiguration

Why? Is 2D a problem?

FULL PROTEIN SIMULATIONS - FAST VIEW

Skolnick and Kolinski

Ann. Rev. Phys. Chem. 40 : 207-235 (89)
Science 280 : 1121 (96)

Lattices
- initially tetrahedral lattice
- more complex lattice (α carbons connected by a vector $(\pm 2, \pm 1, 0)$

Model for the Amino Acids
- α-carbon representation of AA's
- Rigid "β carbon" side position added

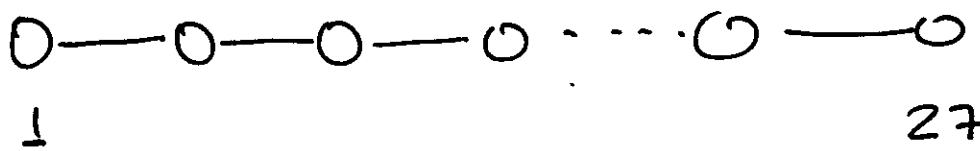
Potentials
- spatial contacts terms
- local propensities (very controversial)

How to compute them

- Full enumaration is impossible
- Monte Carlo Dynamics

- With their potential they were able to fold the protein
- multiplicity of folding pathway

THE 3x3x3 CUBE



27 AA's moving on a cubic lattice

Shakhnovich et al

$$E = B_0 \sum_{ij}^{N} \Delta(r_i - r_0) + \sum_{ij}^{N} B_{ij}\, \Delta(r_i - r_0)$$

$$P(B_{ij}) = (2\pi B^2)^{-1/2}\, e^{-\frac{B_{ij}^2}{2B^2}}$$

$B_0 = -2\ K_B T$
$B = 1$
} unique ground state below T_θ transition
Most of runs at $T = 1$

$5 \times 10^5 \longleftrightarrow 5 \times 10^8$ steps of simulation

① ONLY 1/30 folded

No Real Conclusion →
- Rapid non-cooperative compactization
- followed by a slower transition over a free energy barrier

② NO INFORMATION ON SEQUENCE DEPENDENCE