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/775-11

COLLEGE IN BIOPHYSICS: EXPERIMENTAL AND THEORETICAL ASPECTS OF BIOMOLECULES

26 September - 14 October 1994

Miramare - Trieste, Italy

Protein Folding: An Introduction

Nicholas D. Socci
University of California at San Diego
La Jolla - CA, USA

An Introduction and overview of Protein Structure and Folding

A Theoretical Physicist's Perspective on Proteins

N. Socci

Overview of Lectures (5)

1 Intro to basic protein biochem.

- Structure & Folding
- Folding Problem & other possibly interesting Questions

2 Proteins from a physics viewpoint

- Simple Protein Like models
- Generic Defin of 2° Structure
- Orgins of 2°

3,4,5 Simple Lattice Models

Folding Dynamics: Kinetics vs. Thermodynamics

References

- Std. Biochem Text [Zubay, Stryer, Lewisher]
  Just the First Few chapters
  (stop when you hit glycolysis)

- "Principles of Protein Structure"
  G. E. Schulz

- "Protein Folding" Creighton { Actually he has many nice books on protein folding

- Lattice Work: Chan & Dill
  Annu. Reviews of Biophys. Biophys Chem
  20: 447 (1991)

- Reprints: References therein

- Annual Review of Biophysics & Biomolecular Struct.

Important Distinction Between Biology Physics (Philosophical)

Biology celebrates Diversity

(Biology is full of annoying exceptions'

Physics is often the search for universal behavior.

Luckily Biology is complex enough that both approaches are need to understand it.

Enormous Diversity of Organisms built from

A small (relatively) common set of molecular building blocks.

HOW?

1 ans. POLYMERS

4 Key Biopolymers: Fatty Acids
Polysaccharides
Nucleic Acids
Polypeptides (AKA. Proteins)

* Commonality: not an accident → protocell
- Important for higher Organisms
- Less important for Bacteria ∴ can exploit different Antibiotics ← or be exploited by them.

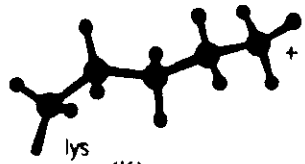
EXTERNAL

ACIDIC

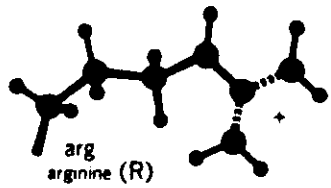
asp
aspartic acid (D)

glu
glutamic acid (E)

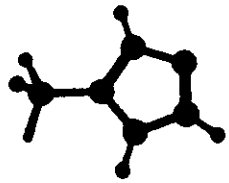
BASIC



lys
lysine (K)



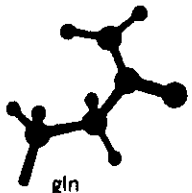
arg
arginine (R)



his
histidine (H)

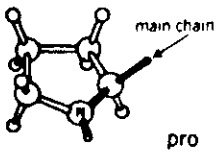
NEUTRAL

asn
asparagine (N)

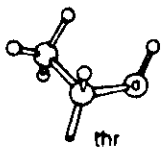


gln
glutamine (Q)

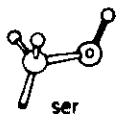
AMBIVALENT



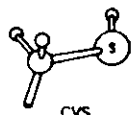
pro
proline (P)



thr
threonine (T)



ser
serine (S)

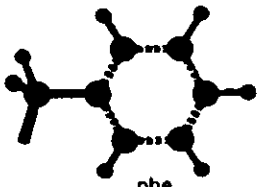


cys
cysteine (C)

ala
alanine (A)

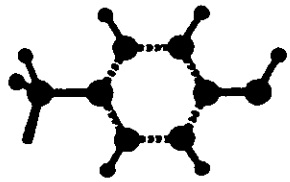
gly
glycine (G)

INTERNAL

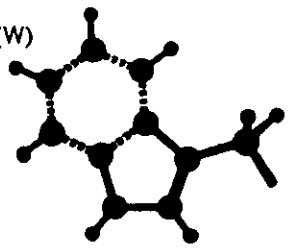


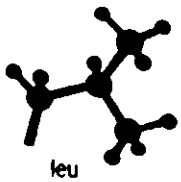
phe
phenylalanine (F)

tyr
tyrosine (Y)

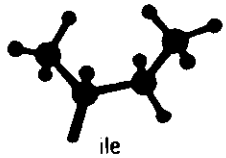


trp
tryptophan (W)



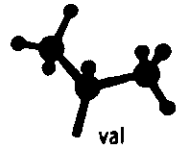


leu
leucine (L)



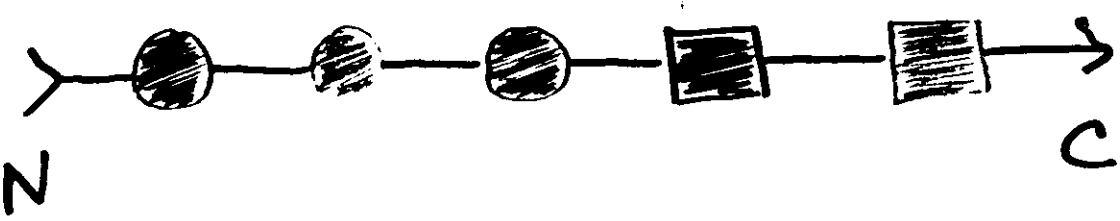
ile
isoleucine (I)

met
methionine (M)



val
valine (V)

Proteins are linear, heteropolymers

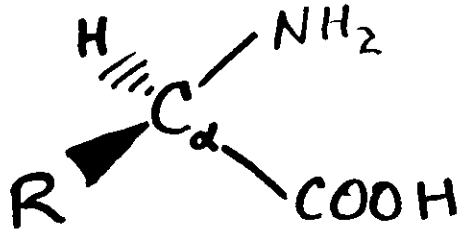


(Not branched)

[exep. cys Pro

Most Proteins are made from a common set of 20 Amino Acids [exep. Collagen Bacterial Wall ↳

Basic Structure



20 different R's

Chiral

All L-stereoisomers

exc some bacteria use D

Some belive Weak asymet explains why L

20 is a pretty large number of monomers (DNA, 4) → Diversity $20^{100} = 10^{130}$

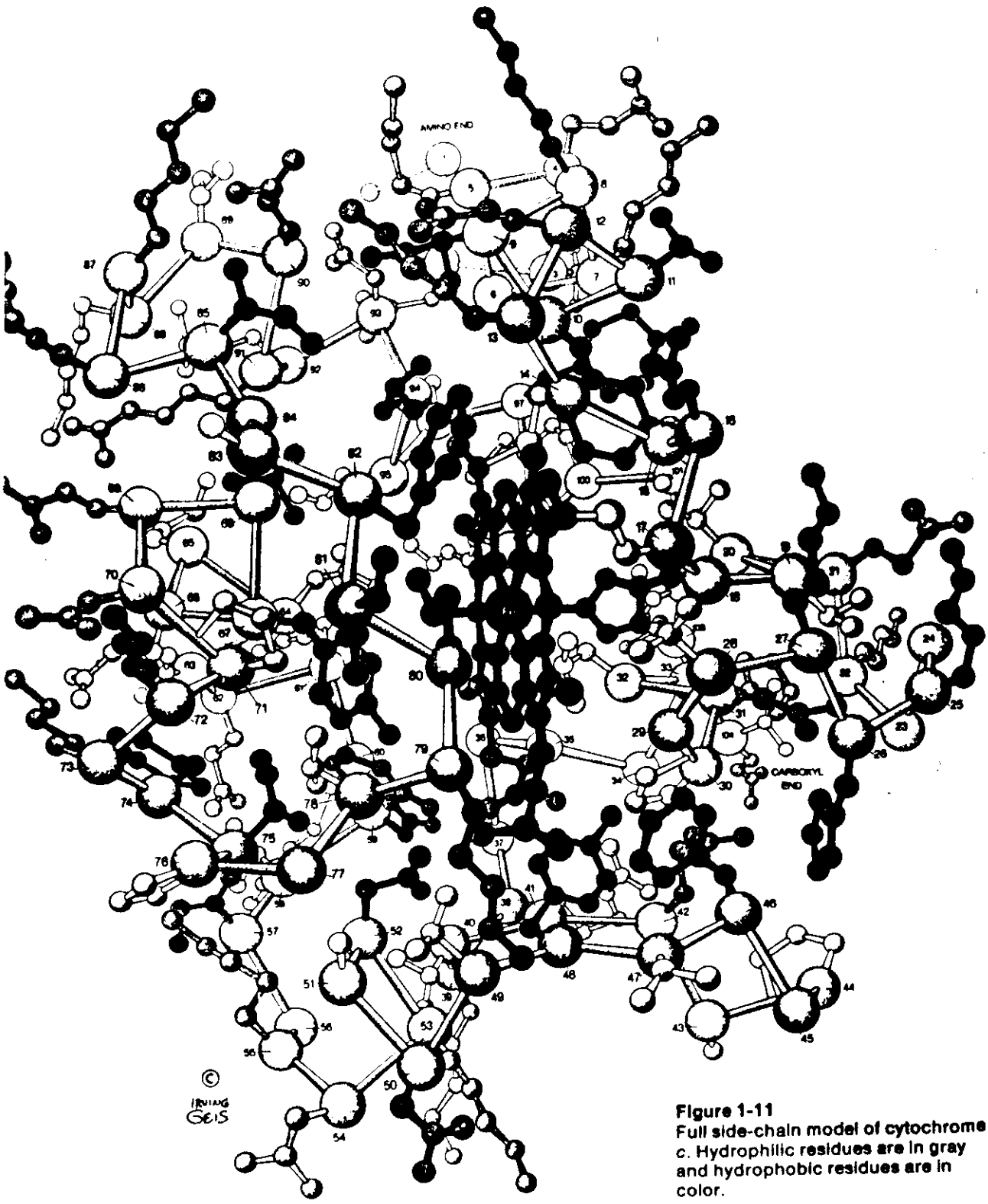
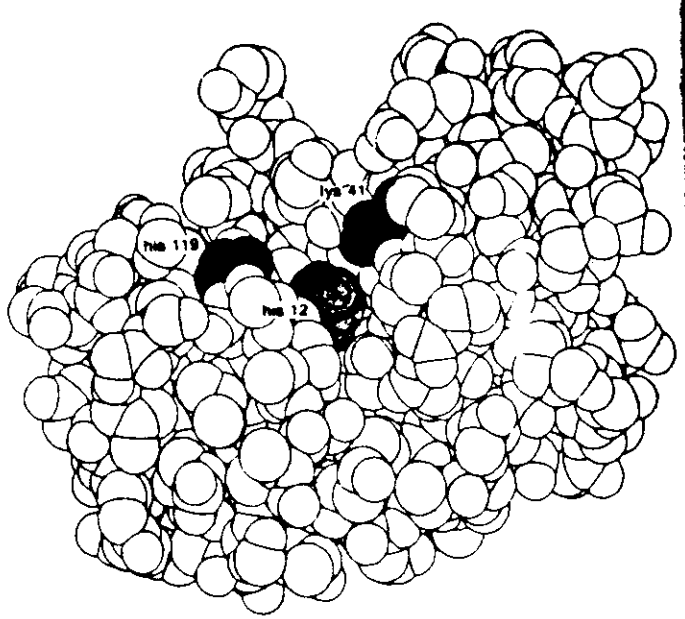
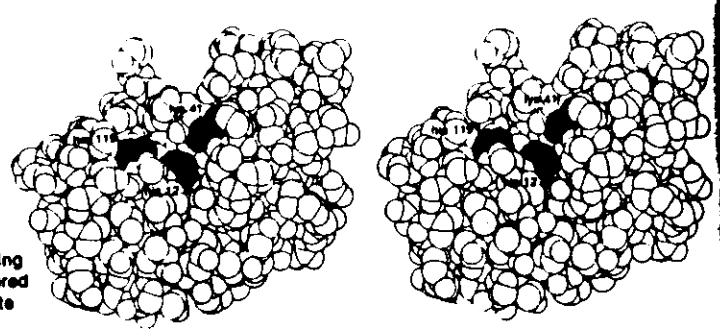


Figure 1-11
Full side-chain model of cytochrome *c*. Hydrophilic residues are in gray and hydrophobic residues are in color.



1-30
filling model of ribonuclease S
ree critical side chains, His
s 12, and Lys 41, shown in
Ribonuclease S has a cleav-
tween residues 20 and 21 (see
r 4 for a discussion of the rela-
p between this enzyme and
urally occurring enzyme).

The 3-D illusion can be seen without stereo glasses by the following method. With your eyes about 10 inches from the page, stare at the drawings below as if you were looking straight ahead at a far-away object. A double image will form and the central pair should drift together and fuse. Then the illusion becomes apparent. Adjust the page, if necessary, so that a horizontal line is perfectly parallel with the eyes. As an aid to seeing one image with each eye, use two cardboard tubes from toilet-paper rolls. Close the right eye and focus the left eye on the left image. Do the same for the other eye. Now with both eyes together, the 3-D illusion should appear.



1-31
views of the same space-filling
shown above. The three colored
hains are part of the active site
e.

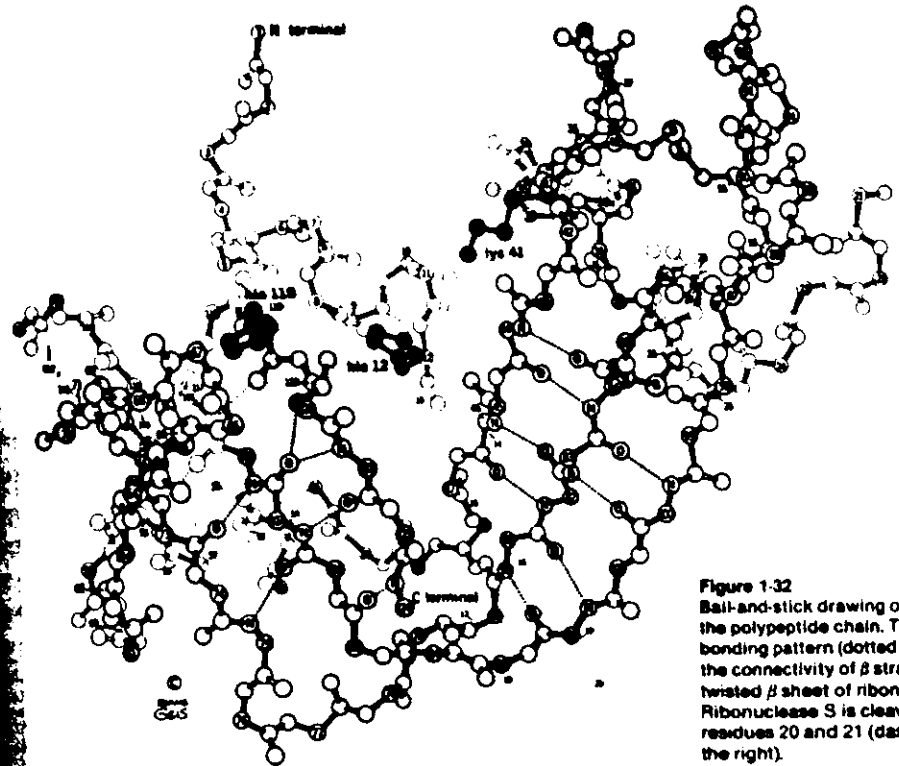


Figure 1-32
Ball-and-stick drawing of the cours
the polypeptide chain. The hydroge
bonding pattern (dotted lines) show
the connectivity of β strands in the
twisted β sheet of ribonuclease S.
Ribonuclease S is cleaved betwee
residues 20 and 21 (dashed line to
the right).

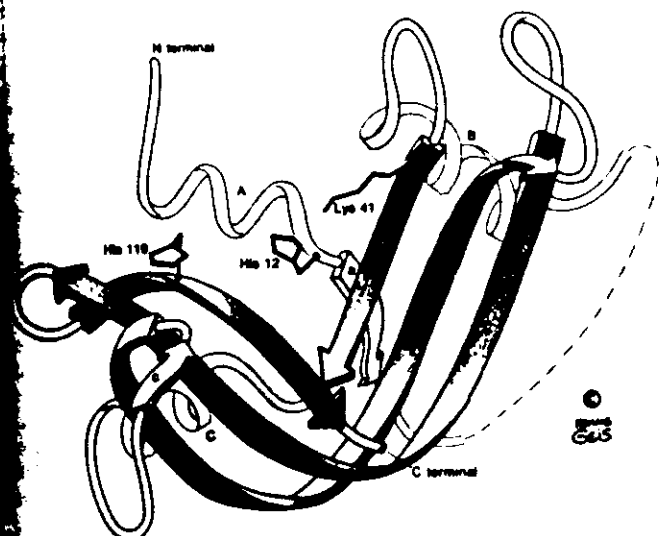


Figure 1-33
Ribbon drawing of ribonuclease
The three significant side chains
in black and a tint of color defin
the β strands. Helices are design
sequentially by capital letters an
strands by lowercase letters.

Proteins are not as complex as they First Appear.

① Amino Acids can be grouped in various ways.
one important grouping is

hydrophobic vs hydrophillic
(apolar) (Polar)

② Heterogeneity is All in side chains
Backbone is Reglular (like DNA, RNA)

(Fig)

③ Special Nature of Peptide Bond
freezes <u>one</u> angle [almost always in trans

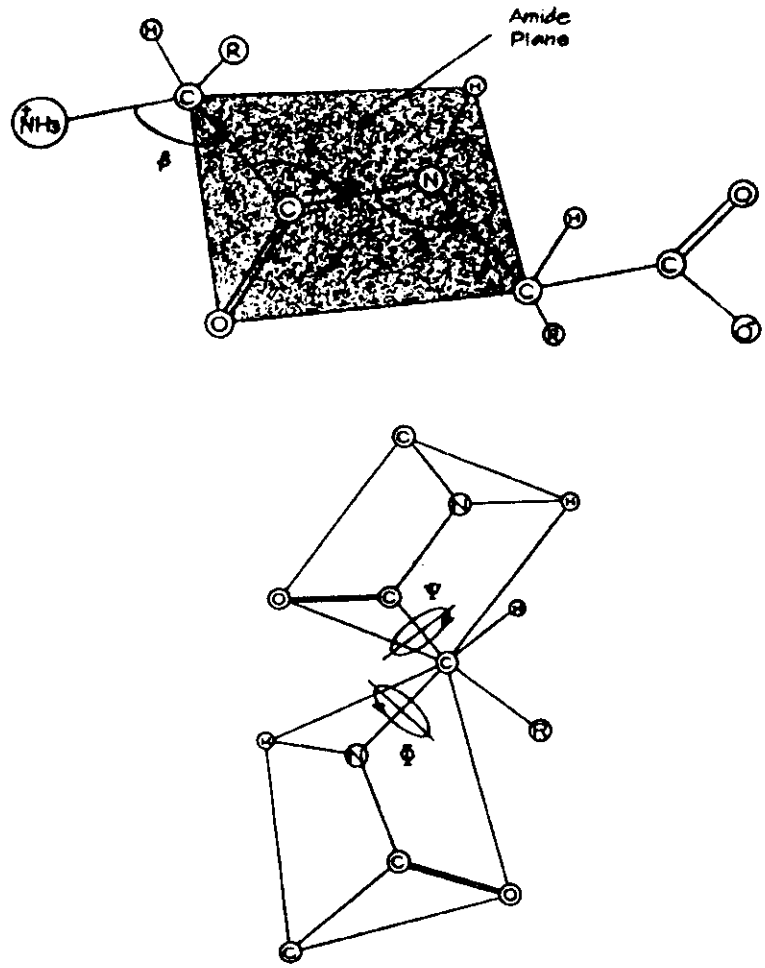
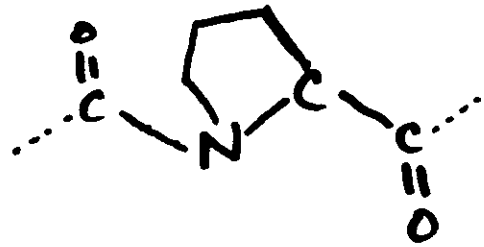


Figure *.2: The figure on top is of a dipeptide. Note the peptide bond between the carbonyl carbon of one amino acid and the amino nitrogen of the other. The formation of the peptide binds results in loss of one oxygen which is removed in the form of water. The left end of the chain is the *N-terminus* and the right is the *C-terminus*. The atoms that at the corner of the shaded rectangle remain planar due to the special nature of the peptide bond. The figure on the bottom shows a two amino acid segment of a longer chain with the torsion angles Φ and Ψ labeled.

Linear "Exceptions" Cystein & Proline

Proline



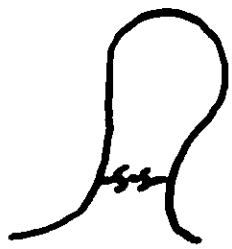
- "Locks" Φ angle
- But Frees Peptide Bond { cis-trans
  ω angle

Cysteine

Can Cross Link

$\}-SH \quad HS-\{ \Rightarrow \}-S-S-\{$

Disulfide Bond



- Important for extracellular proteins [BPTI 3]
- Also for multichain molecules [Insulin]

Proteins have many interesting propeties { catalys ET

Here interested in folding property.

In the proper conditions { pH T salt } proteins will fold spontaneously to stable, compact, specific structures,

- May have a few conformations
- Thermal motion

stable → 0° — 40°, 60° C

compact → ~.75 ≳ .74 spheres

Specific 3D structure is determined by the sequence of Amino Acids

2nd Genetic Code

(DNA, RNA) —1st→ AA seq —2nd→ 

Solved (MOSTLY) — Not Solved

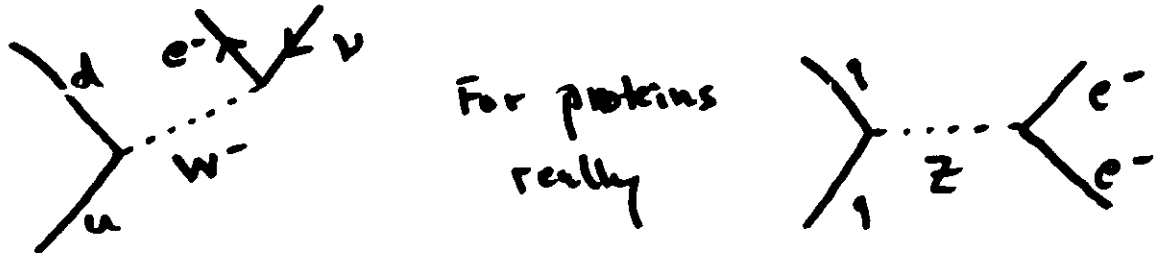
* Protein Folding Problem

Protein Folding: An Introduction
Part II

Nick S.

Loose Ends:

Weak ≡ ElectroWeak interaction
(as opposed to not-powerful)



(mirror)

Key Point: Chirality breaks parity symmetry
: So does the Weak force

More Loose Ends

Mapping Sequence ⟶ Structure

"Homomorphic" "Many-to-one"

Different Seq map to very simmilar Struct

- Sequence Homology.

  Serine Proteases 40% seq Homology
  - very similar structures {Identical Active sites}
  - Divergent evolution

- Mutations

  Proteins are stable against A wide range of mutations
  - Structure remains mostly the same
  - Stability can be affected.

  Lysozyme Mutants

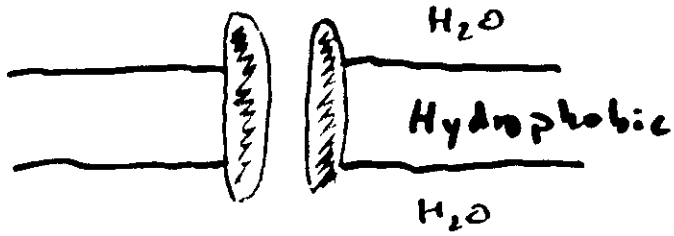
Important Restriction:

Most proteins studied (folding / structure) are Globular, Water Soluble, Crystallizable Proteins

↓

May have important structural consequences

- Only one (?z?) membrane protein has been crystallized: Photosynthetic RC
- Membrane proteins live in both water and oil like environments



- Not much known about how membrane proteins fold or are inserted in membrane

- Also neglect short peptides
  Most do not have specific structures, except when bound to something
  - neurotransmitters
  - insulin
  - immune recognition fragments

Forces involved in protein folding

4 Forces

- Gravity
- * EM *
- Strong (Color)
- Weak

EM force is the only relevant force for folding.

However; relatively useless statement too reductionist.

- Laplace → Mechanics ⇒ Predict Future
- Feynman → 95% photons & electrons (EM)

Feynman is right but does not really help in understanding of proteins

Complex System so must reduce <u>but</u> also must take care not to go too far

Energy Conversions

	Kcal/mol	KJ/mol	J x 10^{-21}	eV	Wave #	T °K
	1	4.19	6.95	0.043	350	504
	23	96.5	160.2	1	8065	11616
	.578	2.42	4.02	0.025	202	291
C—C	88	368	611	3.82	30800	44000
S—S	51	214	355	2.21	18000	25700
H-bond						
vacuum	6	25.1	41.7	0.260	2100	3024
sol'n	~ 2	8.4	14	0.09	700	1007
van der Waal	0.2	0.84	1.39	0.009	70	101
ree Energy Fold	10	42	70	.43	3500	5035
hoton						
blue	71	300	500	3.1	25000	36000
red	41	171	283	1.8	14285	21000

Try Again: NOTE: ENERGY SCALES

Forces involved in Protein Folding

(various Faces of electrostatics)

① Covalent (Bonds) ~100 Kcal/M

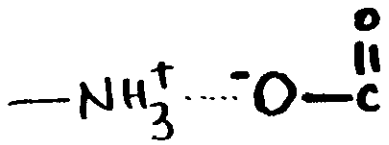
Do not play a direct role in folding: simply holds chain together. Exception: Cystein -S-S-

---

5-.5 KCAL

② Electrostatic (charge-charge | dipole-dipole)

SALT BRIDGES



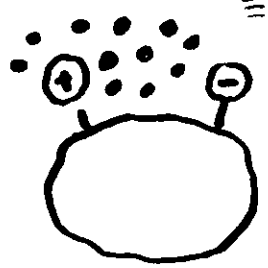
Charged Side Chains (pH dependent)

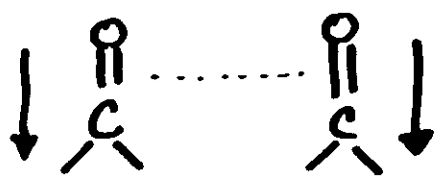
$$F = \frac{q_1 q_2}{\epsilon R^2}$$

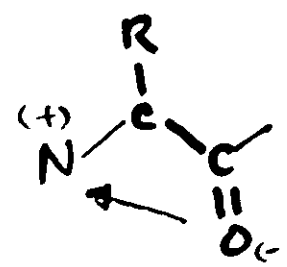
$\epsilon = 80$ Water

$= 4,5$?? proteins

Since most charge groups on surface Significant Screening







Dipole-Dipole

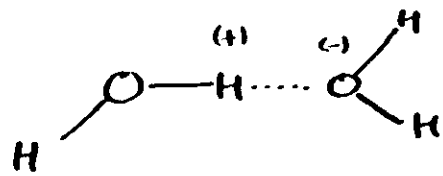
$E \propto \frac{1}{\epsilon r^3}$

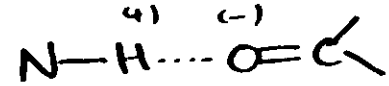
③ Hydrogen Bonds

$3 \sim 4$ KCAL/MO

Very Important

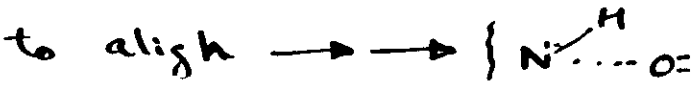
- Large ϵ
- Much more common than Salt Bridges





Donor Acceptor

Proteins have Lots of H-Bond D, A

Like dipole-dipole want to align → → 

However MUCH MORE <u>DIRECTIONAL</u>

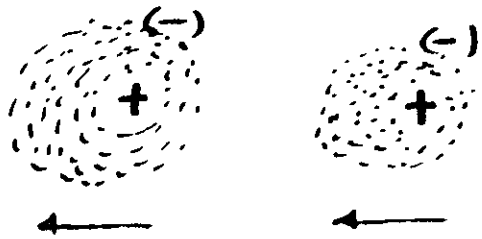
This Directionality gives rise to much of the specific character of protein Structure (ie why α-helix 3.6 periodicity)

---

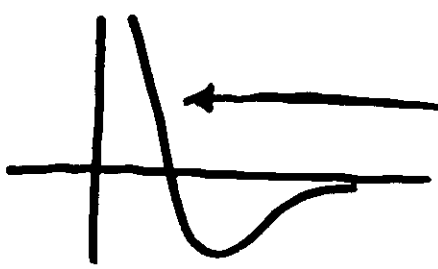
④ Van der Waal, Dispersion, Induce dipole

Caused by distortion of electrons

$\sim .1 - .01$ Kcal/



$$E = \epsilon\left[\underbrace{\left(\frac{\sigma}{r}\right)^{12}}_{\text{Phenom}} - \underbrace{\left(\frac{\sigma}{r}\right)^{6}}_{\text{Induce Dipole}}\right]$$



Weakest But Key Role

Excluded Volume interaction

⑤ Hydrophobic Effect

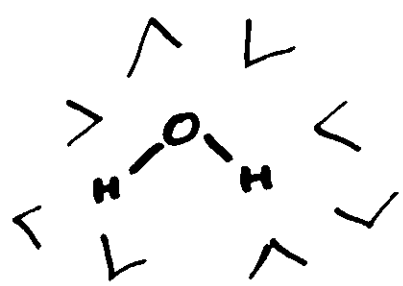
[$P\Delta V \cong 0$ so $G = F$

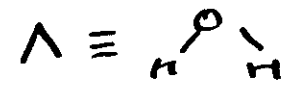
First: Must think Free Energy

$$F = E - TS \quad (\Delta F = \Delta E - T\Delta S)$$

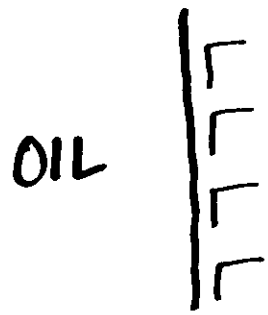
↑ NOTE SIGN want ΔS inc

- A molecule of water in water has lots of rotational free since it is surrounded by H-Bond Donors & Acceptors





- But at oil interface water will locally order to maximize H-bonds

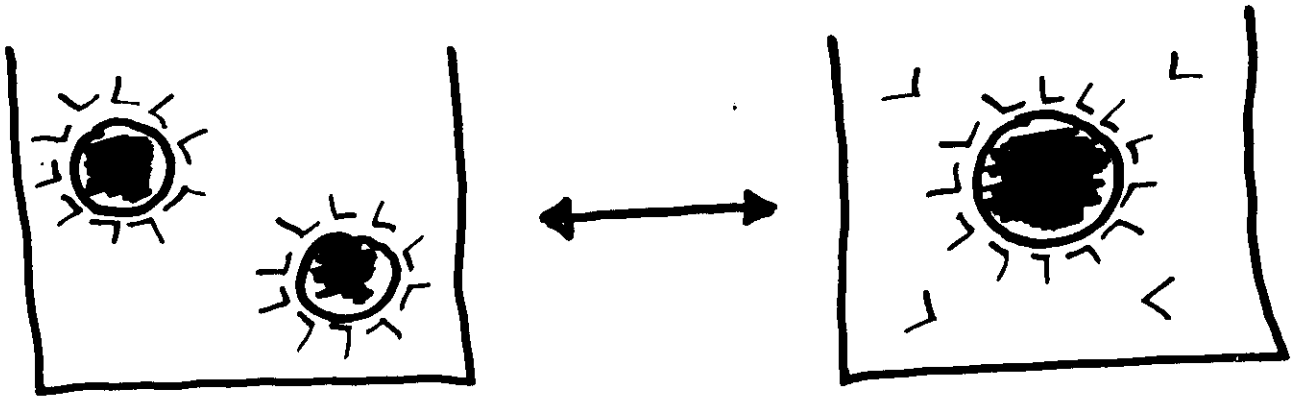


lose entropy $\Delta S < 0$

but compensated by ΔE of H bonds

However do not want to lose too much entropy

Look at two oil bubbles



In joining the two bubbles Reduce there surface area ($\Delta V = 0$). Since the amount or ordered (surface) water $\propto$ to surface Area $\underline{\Delta S > 0}$ for joining

so $\Delta F < 0$ oil droplets want to merge

Hydrophobic effect is to minimize exposed surface of non-polar material

Effective Attraction [Many Body Effect]

- Not really this simple, but gives basic idea
- Question: Why does salt (NaCl) dissolve??

Dominat Force which Drives Protein Folding is the Hydrophobic effect:

The folding of the chain creates a "core" of hydrophobic residues effectively removing them from the water. The surface generally contains the polar (hydrophillic) residues.

Similar to the phase seperation of oil and water. Micelles & Cell membranes exploit the same effect.

[Cold Denaturation]

Note: Although the hydrophobic effect is the dominant force it is not the only force involved in folding:

- H-bonds
- Packing (Steric) Excluded Vol
- HE is fairly non-specific, difficult to imagine getting unique structures with it alone

Many proteins can fold and unfold (denature) reversibly and do so <u>in vitro</u> {as opposed to in vivo

First Demonstrated by Anfinsen on Ribonuclease

⇒ Thermodynamic Hypothesis: The folded state is the global Free Energy minimum of the protein – water system.

Lots of Data to support this [Total Rev Synthesis of Ribonuclease]

Some feel there is a problem though

- Leventhal Paradox

- # of conformations γ^N $\gamma \sim 4$ N = Length
- Explore 1 conformation every picosec (10^{-12})
- For N = 100 10^{48} secs to fold **IF** proteins explore conformation space randomly (Universe ~ 10^{17})

Conc. ⇒ Pathway(s) must exist to direct protein to native state.

Talk about later ⇒ Eugene will also