-

Class #4, Molecular Biophysics (Physics 621/321), W.F. Reed, Fall 2003

Biochemical overview I: Proteins and Nucleic Acids

Unifying notions of biochemistry

*** All proteins are made of 20 main amino acids.**

*** All nucleic acids are made of 4 nucleotides. Virtually unlimited combinations are possible.**

*** Adenosine triphosphate (ATP) is a universal source of free energy.**

*** The Genetic Code is carried by DNA in all organisms (except some viruses), and mediates the accumulation and transmission of biological information.**

*** All biomolecules and processes are the result of evolution.**

*** Biochemical processes and structures are defined by very precise compositions, conformations and interactions. Minor changes can have drastic biological effects. e.g. D vs. L chirality. Cellulose vs. carbohydrate.**

*** Organisms are compartmentalized by membranes that mediate matter and energy exchange with the environment.**

*** The biological milieu is aqueous.**

*** Biological molecules are built of light atoms: C, H, O, N, P, S**

ions Na^+, K^+, Ca^{++}, Mg^{++}

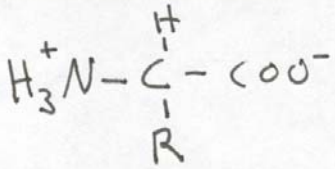
some Fe, Co and Mn

Proteins

-

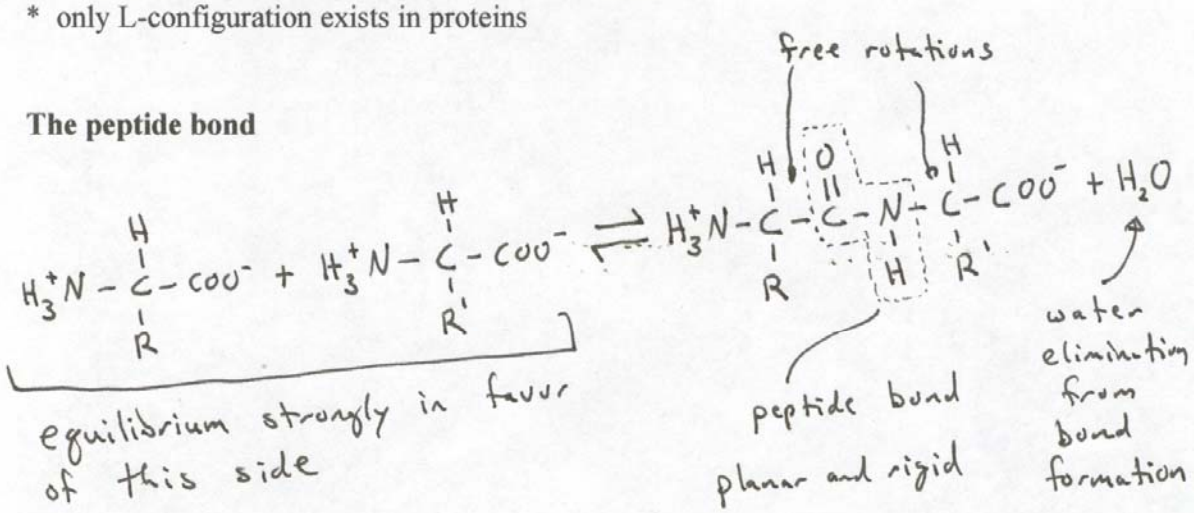
Proteins

There are 20 main amino acids:



*zwitterionic at physiological pH (for non-ionizable R)
* only L-configuration exists in proteins

The peptide bond



Only L- symmetry exists

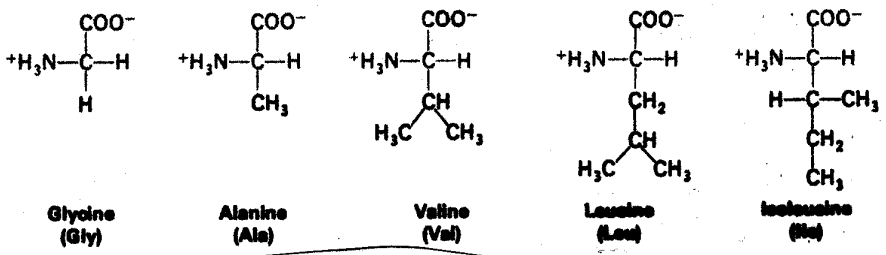
Order convention

start @ amino ⟶ carboxyl

H_3^+N - - - - - - - - - - COO^-

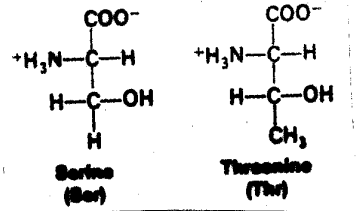
The 20 common amino acids

aliphatic side chains



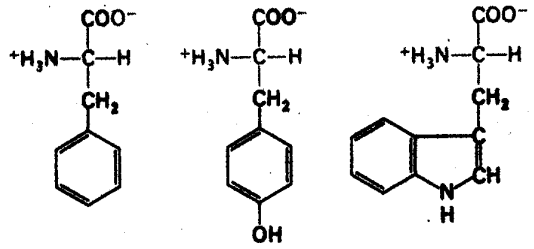
Glycine (Gly) | Alanine (Ala) | Valine (Val) | Leucine (Leu) | Isoleucine (Ile)

aliphatic hydroxyl sidechains



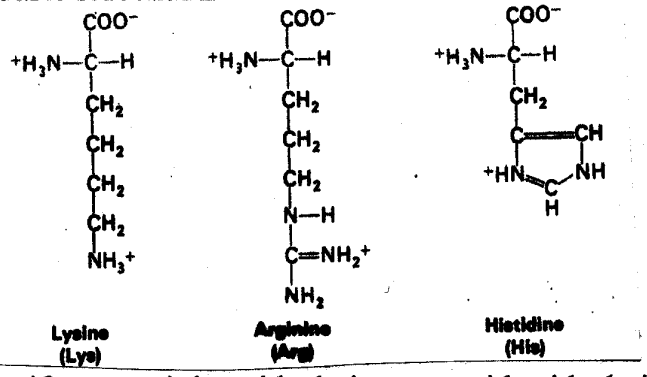
Serine (Ser) | Threonine (Thr)

aromatic sidechains



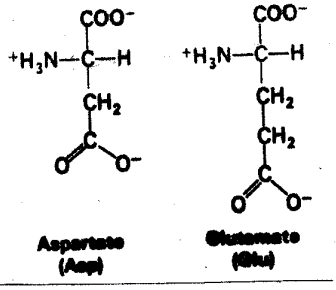
Phenylalanine (Phe) | Tyrosine (Tyr) | Tryptophan (Trp)

basic sidechains



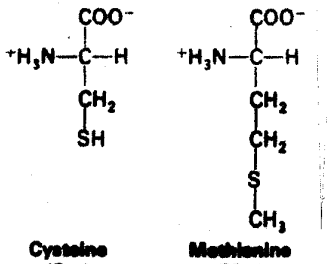
Lysine (Lys) | Arginine (Arg) | Histidine (His)

acidic sidechains



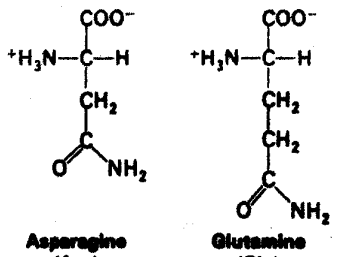
Aspartate (Asp) | Glutamate (Glu)

sulfur-containing sidechains



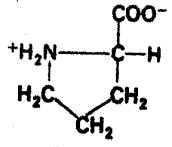
Cysteine (Cys) | Methionine (Met)

amide sidechains



Asparagine (Asn) | Glutamine (Gln)

proline



Proline (Pro)

-

-

Protein Functions

*** All known enzymes are proteins. Virtually all biochemical reactions are catalyzed by enzymes. Proteins themselves exert feedback control in metabolic integration.**

*** Transport and storage. Proteins transport many small molecules and ions. e.g. Hemoglobin-oxygen**

*** Motility, muscle action**

cilia, flagella, muscle; actin/myosin

*** Structural material**

bone, tendon, skin; collagen and elastin

*** Immune system**

anti-bodies, self-identification

*** Generation and transmission of nerve impulses**

reception and transmission across synapses

*** Control elements of DNA**

Gene regulation and expression

*** Chaperones**

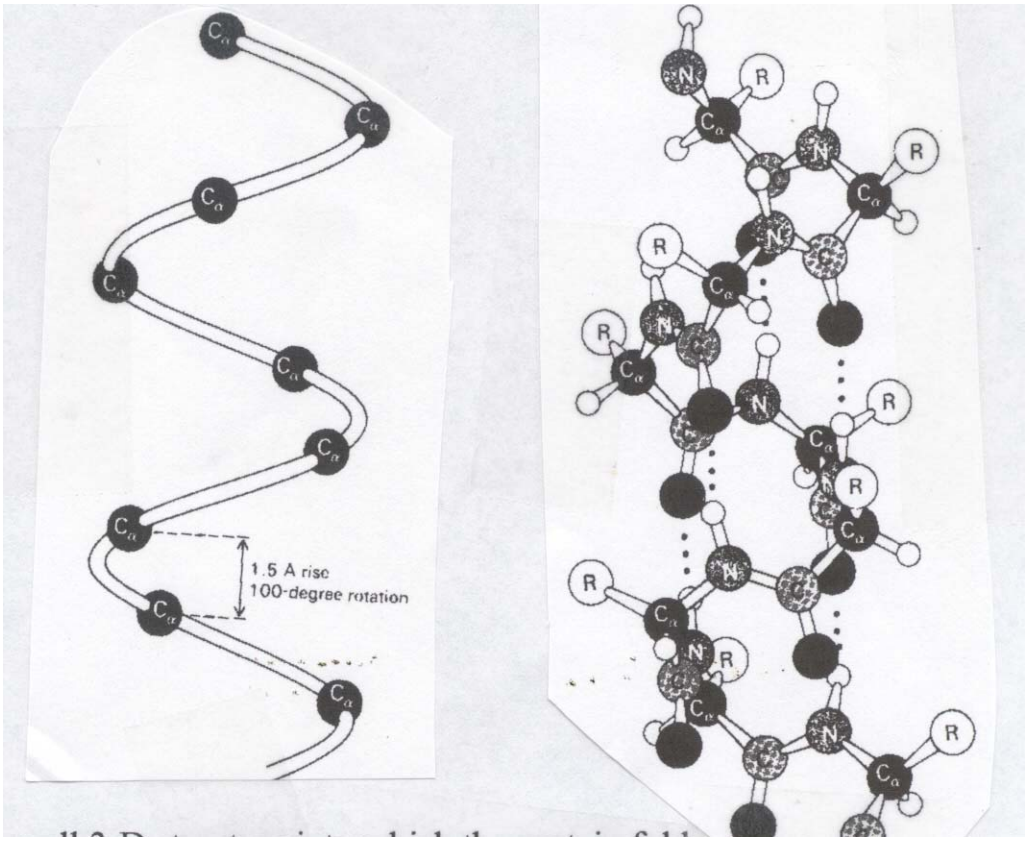
protein folding

Levels of Enzyme Organization

Primary structure: **The sequence of amino acids specifying the enzyme.**

Secondary Structure: The regularities of the three dimensional sub-structures.

α-Helix



*** α-helix backbone is the peptide backbone: functional groups stick out.**

*** All the main chain CO and NH groups are H-bonded, staggered by 4 units.**

*** There is a 1.5Ang. vertical rise and 100^0 of turn between successive a.a. 3.6 a.a./turn, or 5.4 Ang/turn**

*** All α-helices are right-handed.**

*** Two and more helices can wind together to form fibrous bundles; e.g. keratin in hair, myosin in muscle.**

Protein secondary structure, cont'd.

-

β-sheets

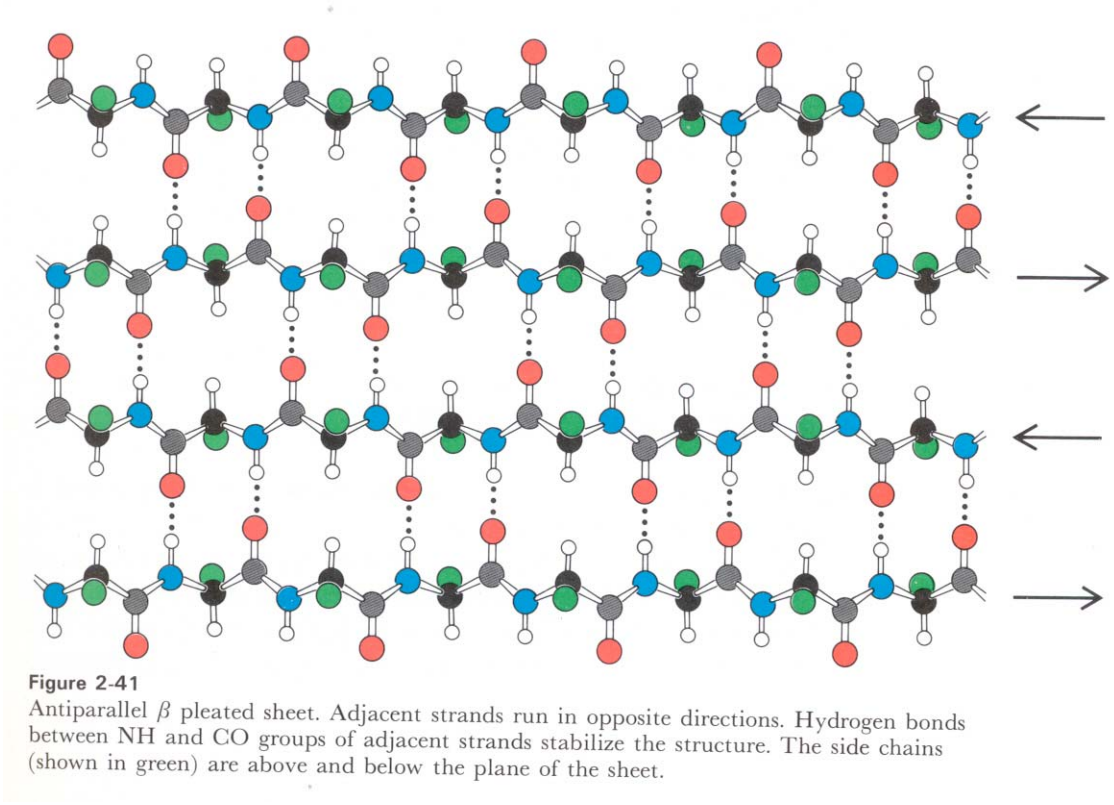


Figure 2-41
Antiparallel β pleated sheet. Adjacent strands run in opposite directions. Hydrogen bonds between NH and CO groups of adjacent strands stabilize the structure. The side chains (shown in green) are above and below the plane of the sheet.

(From Biochemistry, by L. Stryer, Freeman Pub.)

*** Protein backbone is fully extended, all-trans. 3.5 Ang. between adjacent a.a.**

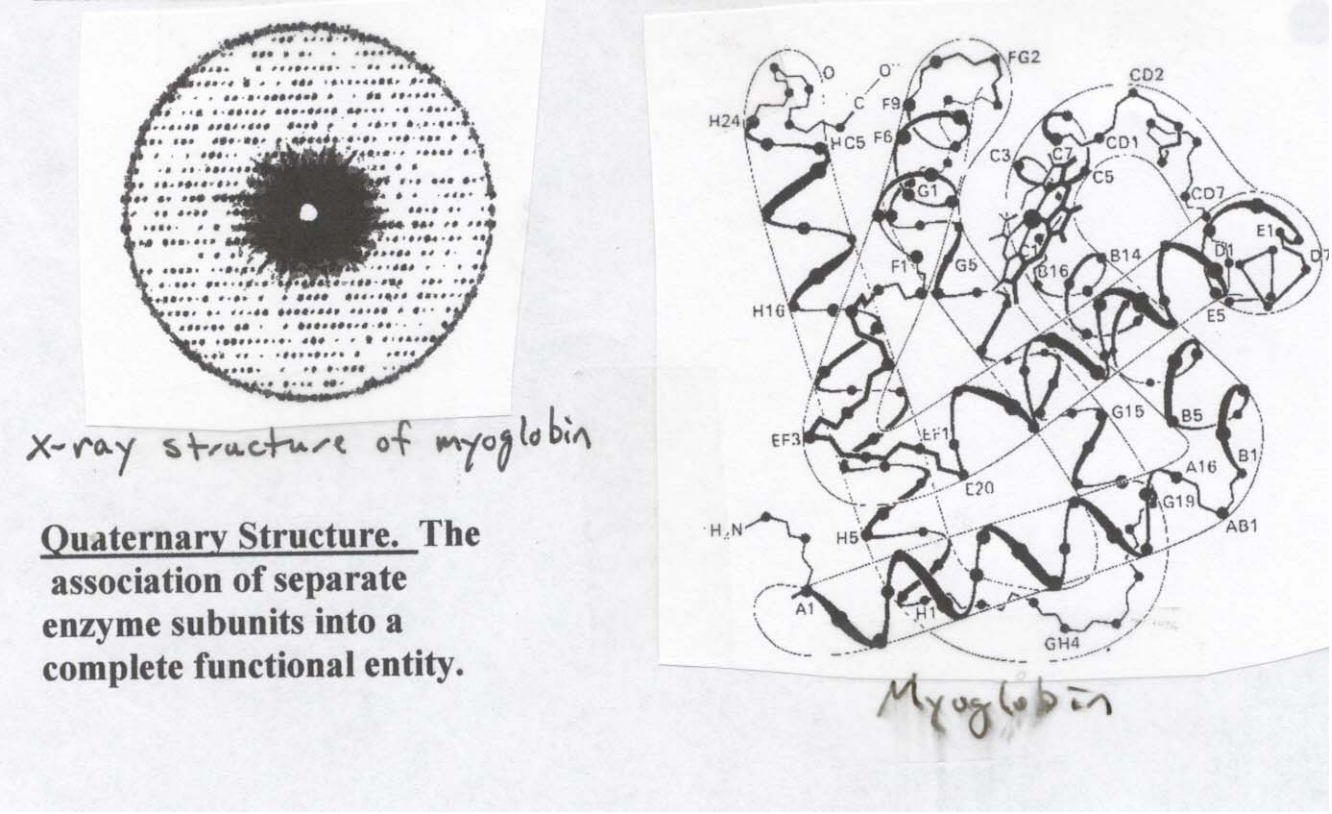
*** sheets can be parallel or anti-parallel and are between different polypeptide strands**

*** Protein strands can make sharp reverses in direction via**

β-turns

Levels of Protein Structure

Tertiary structure: The overall 3-D structure of the protein



(From Biochemistry, by L. Stryer, Freeman Pub.)

Protein folding problem: Predict complete secondary, tertiary and quaternary structure from the primary structure.

Protein chaperones - Molelules that discriminate between slowly-folding and misfolded proteins, aid in correct folding and localize with aggregates to inhibit their formation.

Protein aggregate formation occurs frequently *in vitro*, is suspected to occur *in vivo*, and are related to amyloid diseases, such as Alzheimers.

NEWS OF THE WEEK

APRIL 1, 2002 - EDITED BY JANICE LONG & LOUISA WRAY DALTON

DRUG DISCOVERY

PROMISCUITY DOESN'T PAY

Chemists find mechanism behind false positive drug screening 'hits'

Some molecules culled from compound libraries look like promising drug leads: They're good at inhibiting a particular disease-causing enzyme. But soon it becomes apparent that these molecules aren't at all selective; they inhibit everything under the sun. Chemists call them "promiscuous."

And it's a bitter pill to swallow when you discover that you've invested months, even years, in a promiscuous inhibitor, only to discover that it's biologically useless.

A new study, however, has identified what may be happening to a significant percentage of these compounds—knowledge that could save pharmaceutical companies much fruitless pursuit.

Some of the molecules, it seems, are collecting in large clumps known as aggregates that inhibit enzymes by gumming them up, rather than "docking" monogamously into an enzyme pocket, report Brian K. Shoichet, chemistry professor at Northwestern University, his graduate student Susan L. McGovern, and their colleagues [*J. Med. Chem.*, **45**, 1712 (2002)]. Fortunately, it's simple enough to test for aggregates right off the bat.

"It's important to understand the molecular basis for false positive activity, and this paper makes an important contribution," says Mark Murcko, chief technology officer at Vertex Pharmaceuticals.

The pharmaceutical research community has already pricked up its ears, says Christopher A. Lipinski, a computational chemist with Pfizer Global Research & Development. Since the paper was published online in early March, he notes, three groups at Pfizer have passed it around. "You're talking about the ability, potentially, to save months to years of time," he says.

The fact that some compounds are just bad eggs has been common knowledge among pharmaceutical chemists. But putting a mechanism to the problem has proved difficult. Scientists have had a number of suspicions. For example, it's been thought that some compounds might be acting as covalent inhibitors, chemically reacting with enzymes.

"This work represents an intriguing observation as to why certain hits in high-throughput screening don't make it beyond the screening stage," says Yvonne Martin, a computational chemist at Abbott Laboratories.

Shoichet's group made the discovery while searching for a penicillinase inhibitor. Most of the hits from their screening program were promiscuous.

Several unusual characteristics pointed them toward the problem's source. Most significantly, when they increased the enzyme's concentration, the molecules' ability to inhibit disappeared. The logical explanation, Shoichet says, was aggregates. "Everything clicked when we did that," he says. And with transmission electron microscopy, they indeed saw aggregates 10 to 100 times bigger than the enzyme.

Next, they persuaded the pharmaceutical company Pharmacia to give them 35 compounds from its libraries. Tests soon found that 20 of those compounds formed aggregates.

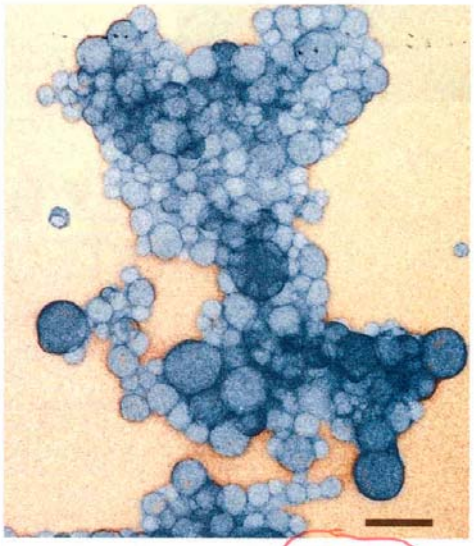
"People can begin to think more about the problem, now that there's some data," says Gerald M. Maggiora, a senior computational scientist at Pharmacia's lab.

"Both experimental and computational screening teams should be aware of this extensive set of false positives," says Irwin D. (Jack) Kuntz, director of the Molecular Design Institute of the University of California, San Francisco.

The aggregates' structures, as well as why and how often molecules form them, are still mysteries. But ultimately, Lipinski says, the work should begin to address "a fundamental problem that's been plaguing people for a long time."—ELIZABETH WILSON

PHOTO BY ELIZABETH WILSON

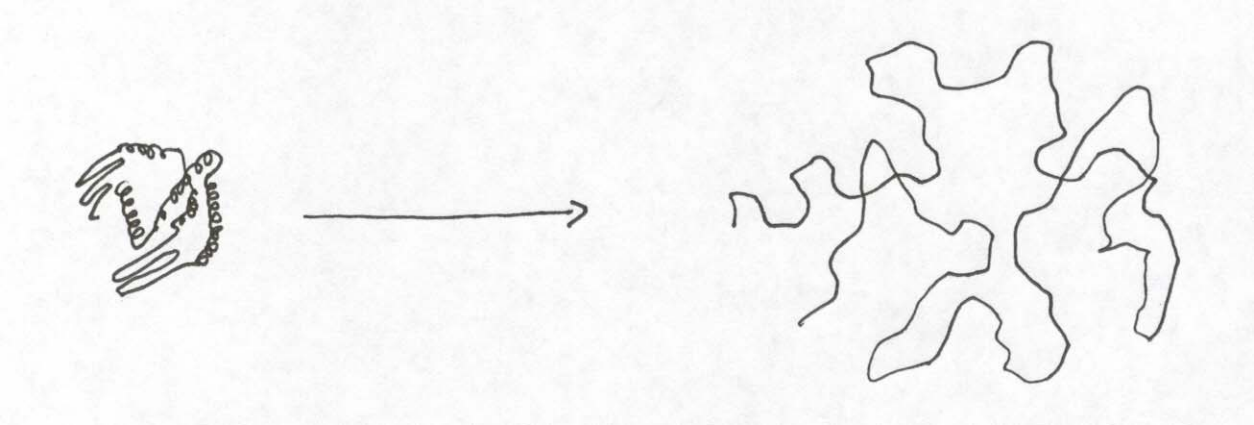
Shoichet



AGGREGATES **Some promising initial drug leads may in fact be biologically useless clumps. (Bar equals 100 nm.)**

Protein Denaturation

Loss of 3-D structure due to effects such as pH, Temperature, or denaturing agents such as urea and Guanadine-HCl.



3-D globular protein $\longrightarrow$ random coil

biologically active biologically inactive

In many cases re-naturation can proceed spontanteously

Total # of possible protein conformations $N=3^n$, n= # of a.a.

e.g. for n=100, $N=3^{100}=5x10^{47}$ different conformations!

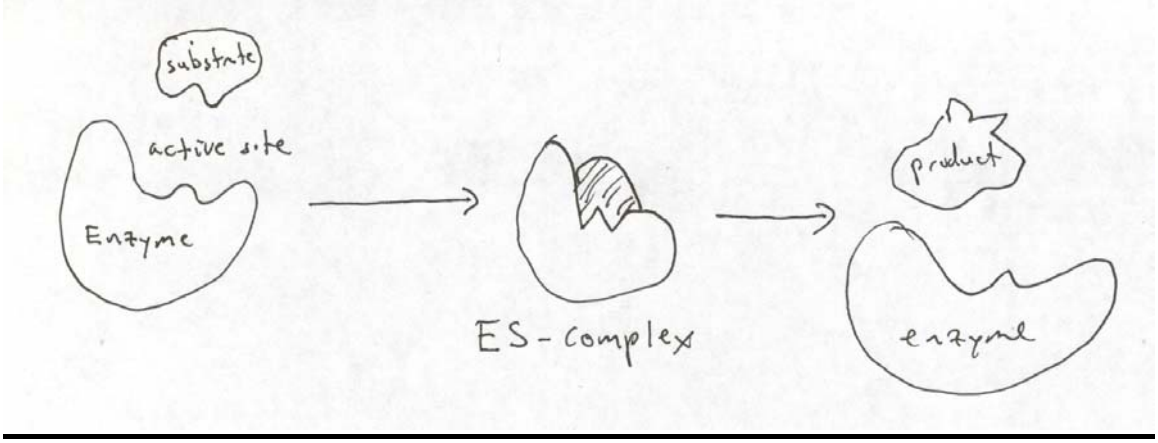
yet the protein primary sequence is normally sufficient to determine the 3-D folded state :

$I_{to\ fold}=\log_2(3^{100})=158$ bits needed to specify the native conformation

In the sequence itself there are $I_{sequence}=\log_2(20^{100})=432$ bits

i.e. enough information in the sequence to specify the structure.

Simple enzyme model: Michaelis-Menten Kinetics



$$E + S \underset{k_2}{\overset{k_1}{\rightleftarrows}} E + S \xrightarrow{k_3} E + P$$

Production rate $V = \frac{d[P]}{dt} = k_3[ES]$

@ steady state rate [ES] = constant

$$k_1[E][S] = (k_2 + k_3)[ES]$$

total enzyme conc. $[E]_T$=[E]+[ES]

solving for [ES] leads to

$$V = \frac{k_3[S][E]_T}{k_M + [S]}, \qquad k_M = \frac{k_2 + k_3}{k_1}, \quad V_{max} = k_3[E]_T$$

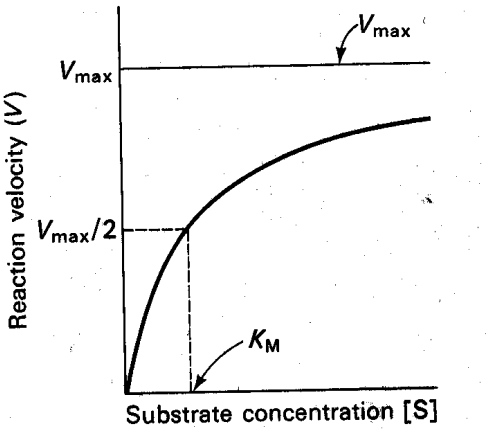
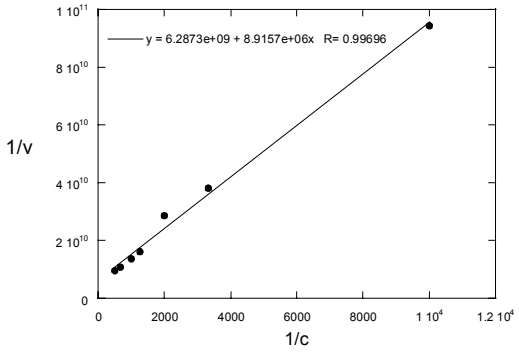
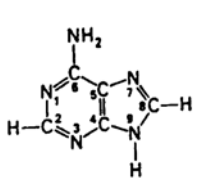


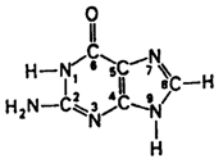
Figure 5



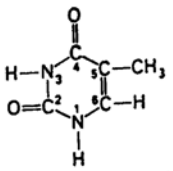
Nucleic Acids - DNA



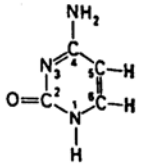
Adenine (A)



Guanine (G)

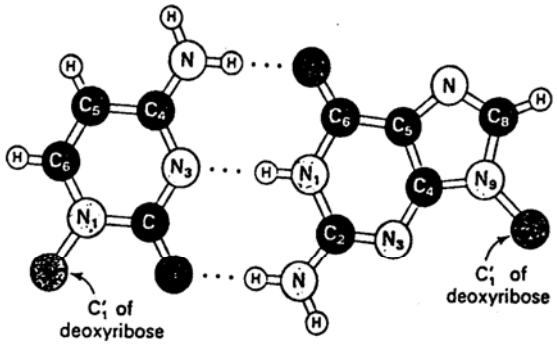


Thymine (T)

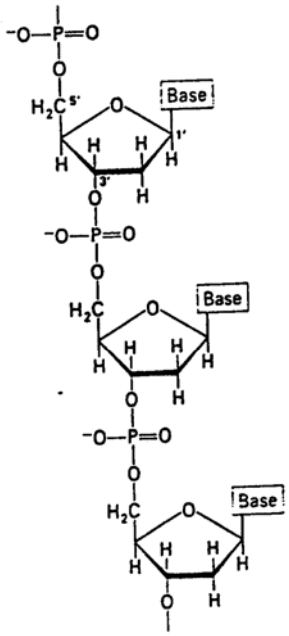


Cytosine (C)

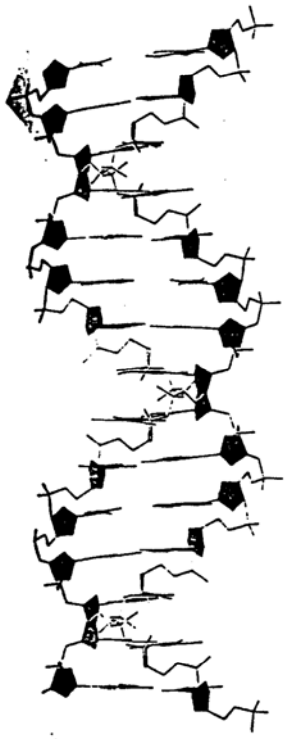
Pyrimidine and Purine Bases



Model of a guanine-cytosine base pair.



Structure of part of a DNA chain.



Skeletal model of double-helical DNA. The structure repeats at intervals of 34 Å, which corresponds to ten residues on each chain.

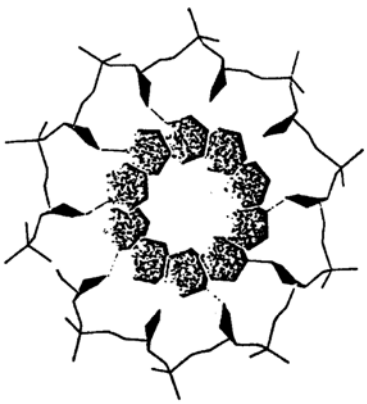


Diagram of one of the strands of a DNA double helix, viewed down the helix axis. The bases (all pyrimidines here) are inside, whereas the sugar-phosphate backbone is outside. The tenfold symmetry is evident. The bases are shown in blue and the sugars in red.

The Genetic Code

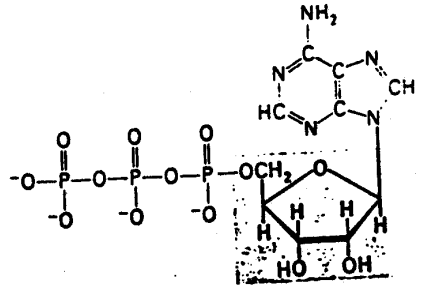
First position (5' end)	Second position				Third position (3' end)
	U	C	A	G	
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop	Stop	A
	Leu	Ser	Stop	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	U
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val	Ala	Glu	Gly	G

Adenylates

1) Energy Source : Adenosine Triphosphate (ATP)

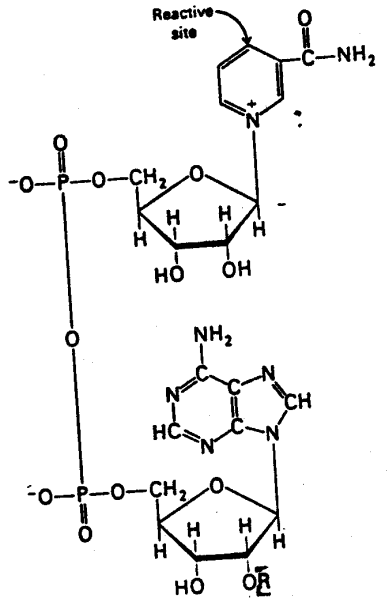
$$ATP + H_2O \xrightarrow[\text{ATP-ase}]{} ADP + P_i^- + H^+$$

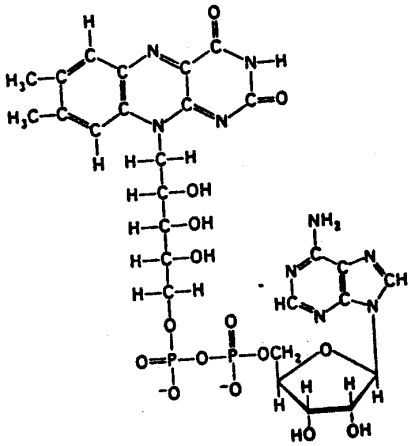
$\Delta G \sim -12$ kcal/mole



2) Reducing Power : Nicotinamide Adenine Dinucleotide (NAD^+)
($NADPH$)

Flavin Adenine Dinucleotide ($FADH_2$)





3) Activation : Coenzyme-A (CoA)