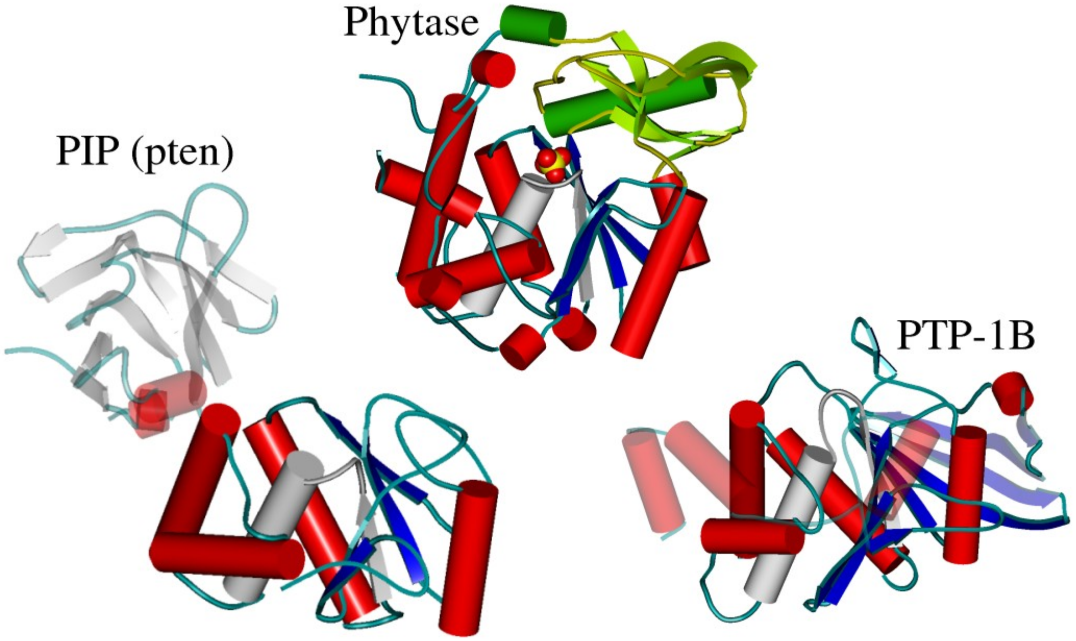
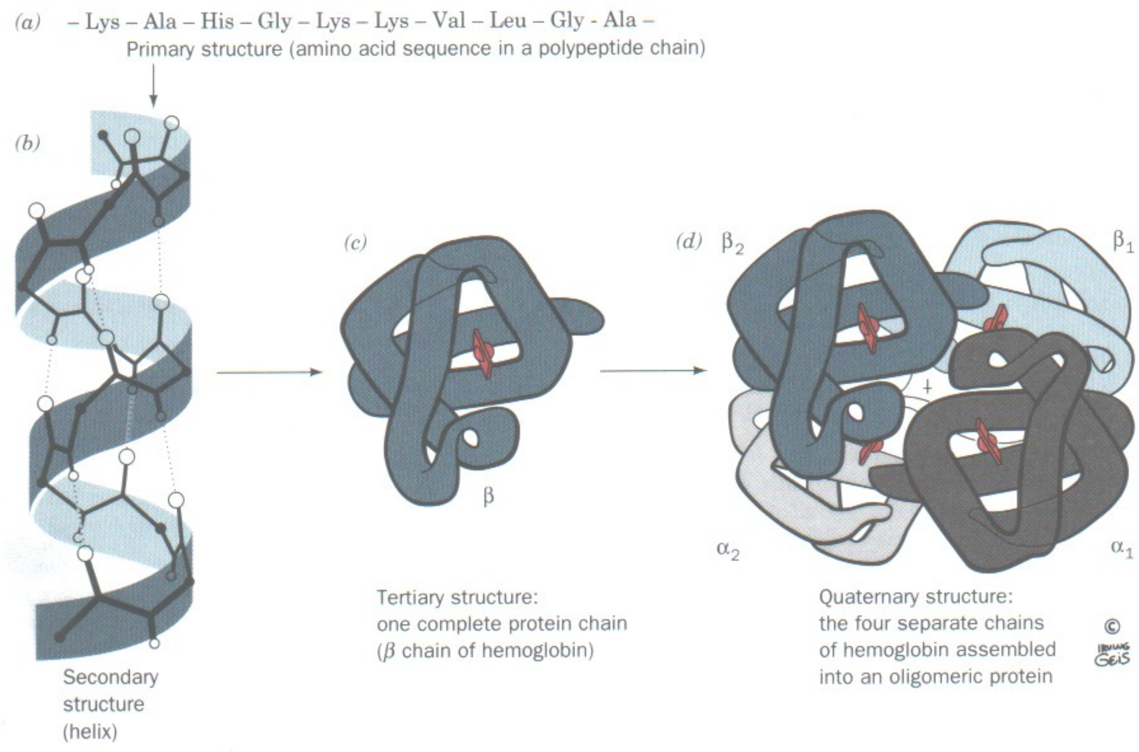
Chapter 7:
Covalent Structure of Proteins

Voet & Voet:
Pages 163-164, 185-194



Structure & Function



Primary structure (amino acid sequence in a polypeptide chain)

Secondary structure (helix)

Tertiary structure: one complete protein chain (β chain of hemoglobin)

Quaternary structure: the four separate chains of hemoglobin assembled into an oligomeric protein

Function is best understood in terms of structure

Four levels of structure that apply to proteins

- **Primary** – sequence of residues
- **Secondary** – local spatial conformation of backbone
- **Tertiary** – overall fold or 3D arrangement of polypeptide
- **Quaternary** – spatial arrangement of subunits in a multisubunit protein

Primary Structure Significance

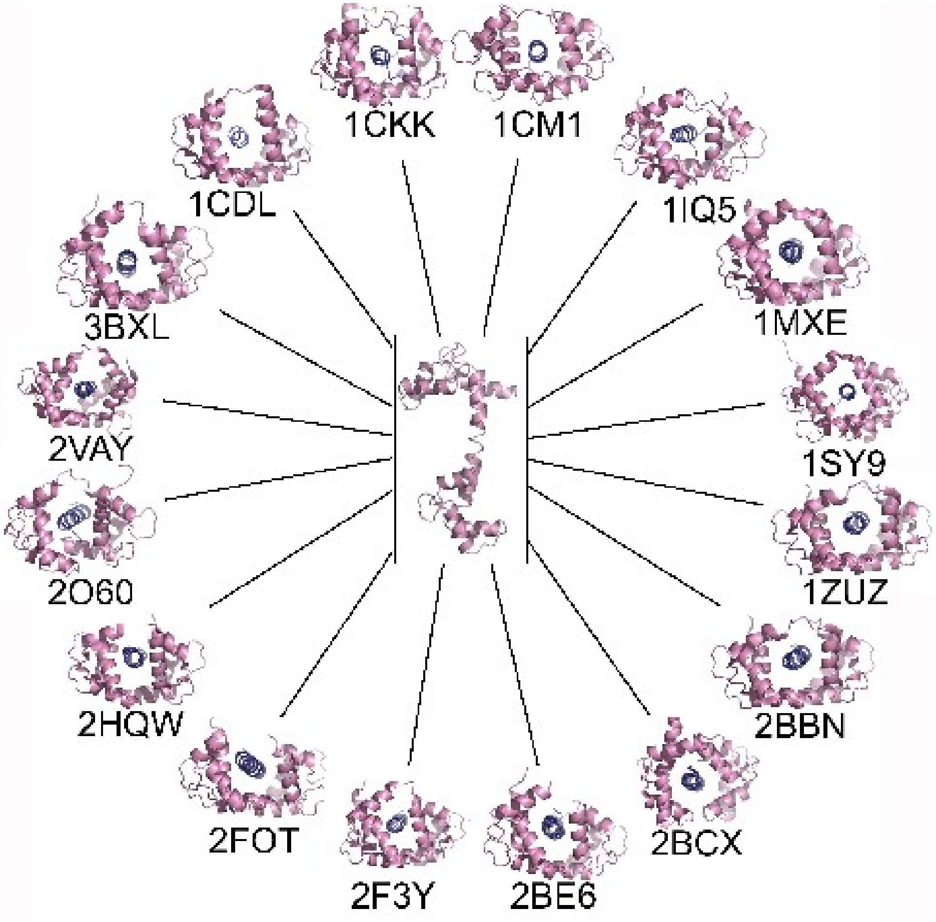
Primary Structure is central to the formulation of modern concepts of Biochemistry

(1) *Structure Determination* – primary sequence information is a prerequisite for structure determination and understanding molecular function

(2) *Evolutionary Relationships* – sequence comparisons of related proteins in different organisms shed light upon protein function and relationships between organisms

(3) *Clinical Applications* – many inherited diseases are caused by mutations that lead to amino acid changes in proteins; recognition of this fact has led to the development of therapies in many cases

Chemical Evolution



Voet & Voet calls the evolution of amino acid <u>primary sequences</u>

Chemical Evolution

Individuals/organisms are characterized by their genetic compositions

- ***Specifies proteins that can be expressed, their quantities and schedule of appearance***

Evolutionary changes (random mutations) often alter protein structure

- ***Mutations that are propagated somehow increase (or not decrease) the probability their owner will survive to reproduce***
- ***Mutations that are deleterious or lethal in their effect rapidly die out***

Sickle Cell Anemia:

An example of Chemical Evolution

Normal individuals

Red blood cells normally adopt a <u>flexible</u>, biconcave disk shape

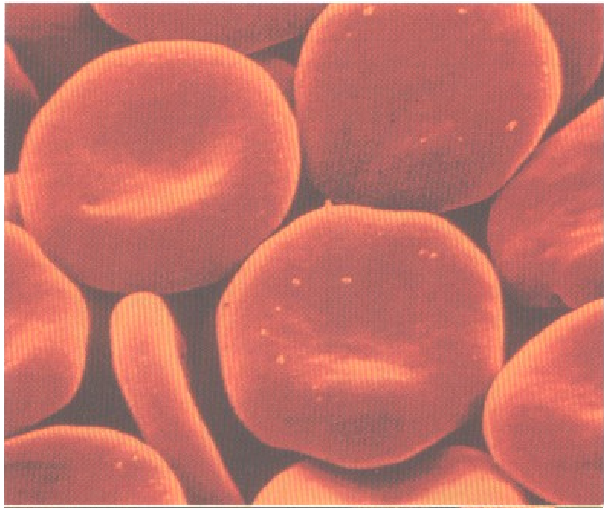
- **Allows cells to pass through very small capillaries**

Sickle Cell Anemia individual

Red blood cells form irregular, crescent-like shape under the low oxygen conditions typical in capillaries

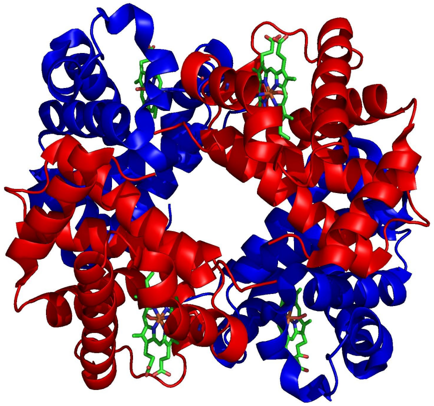
- **Increased red blood cell rigidity and impedes blood flow through capillaries**
- **Tissue damage, pain, hemolytic anemia, early death**

Normal Red Blood Cells



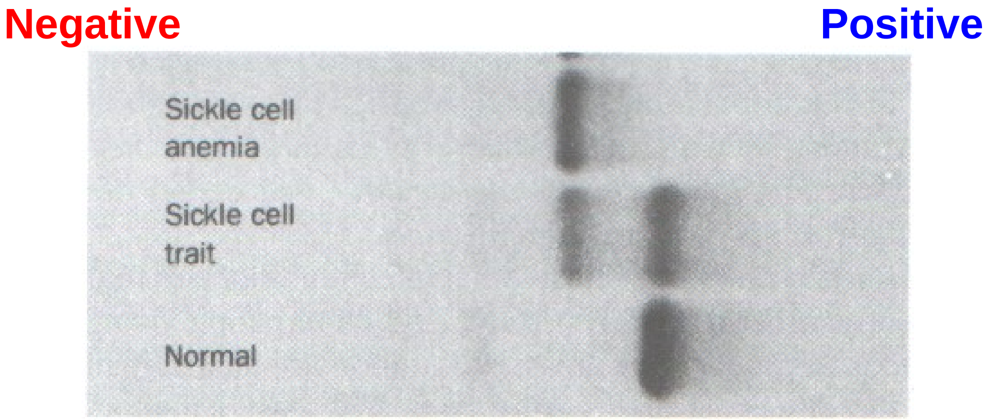
Sickle Red Blood Cells

Sickle Cell Anemia:
A Molecular Disease



Single (point) mutation in hemoglobin (Hb) β subunit

- **Pauling: HbS (sickle cell Hb) has ~2 fewer negative charges than Hb**
- **Trypsin (an enzyme) digests shows mutation is associated with β subunit**
  - **Trypsin specifically cuts polypeptides into smaller fragments**
  - **sequencing reveals nature of mutation; Glu 6 mutated to Val 6**
- **Deoxygenated HbS aggregates into filaments that deform red blood cells**



Non-denaturing
Gel Electrophoresis

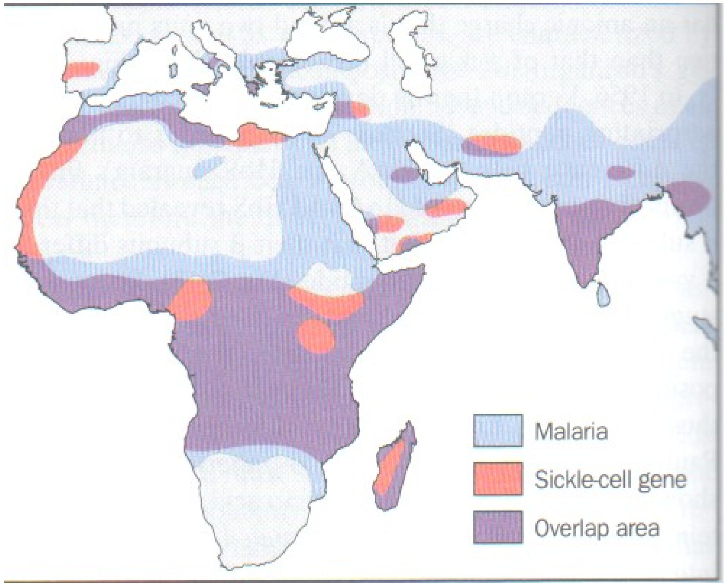
Sickle Cell Anemia:
Why has mutation propagated?

Sickle Cell Anemia is a deadly disease (both chromosomes encode HbS)

- **All red blood cells sickle**

Sickle Cell Trait confers resistance to malaria (one chromosome encodes HbS)

- **Some red blood cells sickle**
  - **Malaria kills ~ 1 million people per year (usually the very young)**
  - **Individuals with Sickle Cell Trait have normal lifespans (though red blood cells have shorter lifespan)**



Sickle cell trait occurs in the same areas of world as malaria

Sickle Cell Anemia:
How does it confer resistance?

Malaria is caused by protozoan, *Plasmodium falciparum*

In Normal Red Blood Cells

(1) Enters cell; reduces the pH

(2) Cells stick to 'walls' of blood vessels and avoid removal by the spleen

(3) Too many cells stick, blood flow is impeded and organs may fail

In Sickle cell trait Red Blood Cells

(1) Enters cell; reduces the pH

Increases tendency to sickle

(2) Sickle cells do not stick to walls of blood vessels and infected cells are removed by spleen

(3) Blood flow remains adequate

(4) Sickling also impairs *Plasmodium falciparum* replication

Neutral Drift: Species Variation

According to evolutionary theory, related species evolved from a common ancestor

- It follows that proteins in related species must also have evolved from a common ancestral protein
- **Homologous** – evolved or derived from a common ancestor

eg. Homologous primary sequences evolved from a common ancestor sequence.

Neutral Drift - random mutations in well adapted proteins that do not affect function

- All proteins continue to evolve over time

Neutral Drift: Sequence Variation

Comparison of primary sequence of homologous proteins from different organism reveals residues that are functionally important

Residue types in sequence comparisons

Invariant residues have essential function

- **cannot mutate without loss of function**

Conservative substitutions indicate less stringent functional role

- **can only mutate to residues with similar physiochemical properties (*eg.* polar uncharged to polar uncharged)**

Non-conservative substitutions and hypervariable regions indicate non-specific functional role

- **can mutate to residues with different physiochemical properties (*eg.* non-polar to polar charged)**

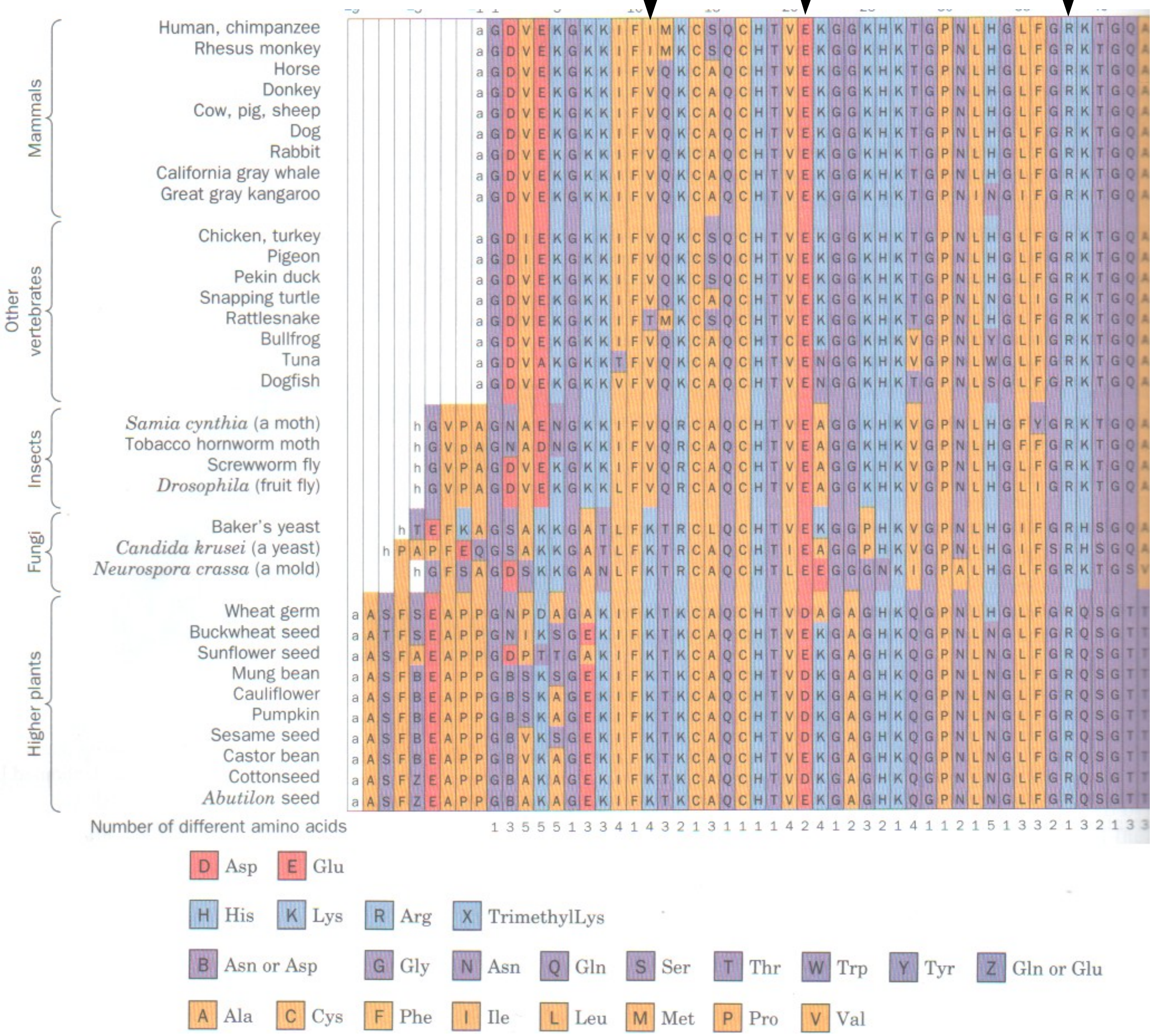
Neutral Drift: Species Variation

Cytochrome C primary sequences (38 organisms) aligned to each other

Non-conservative

Conservative

Invariant



		-9				-5				-1	1				5					10					15					20					25					30					35					40			
Mammals	Human, chimpanzee									a	G	D	V	E	K	G	K	K	I	F	I	M	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Rhesus monkey									a	G	D	V	E	K	G	K	K	I	F	I	M	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Horse									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Donkey									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Cow, pig, sheep									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Dog									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Rabbit									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	California gray whale									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Great gray kangaroo									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	I	N	G	I	F	G	R	K	T	G	Q	A
Other vertebrates	Chicken, turkey									a	G	D	I	E	K	G	K	K	I	F	V	Q	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Pigeon									a	G	D	I	E	K	G	K	K	I	F	V	Q	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Pekin duck									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Snapping turtle									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	N	G	L	I	G	R	K	T	G	Q	A
	Rattlesnake									a	G	D	V	E	K	G	K	K	I	F	T	M	K	C	S	Q	C	H	T	V	E	K	G	G	K	H	K	T	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Bullfrog									a	G	D	V	E	K	G	K	K	I	F	V	Q	K	C	A	Q	C	H	T	C	E	K	G	G	K	H	K	V	G	P	N	L	Y	G	L	I	G	R	K	T	G	Q	A
	Tuna									a	G	D	V	A	K	G	K	K	T	F	V	Q	K	C	A	Q	C	H	T	V	E	N	G	G	K	H	K	V	G	P	N	L	W	G	L	F	G	R	K	T	G	Q	A
	Dogfish									a	G	D	V	E	K	G	K	K	V	F	V	Q	K	C	A	Q	C	H	T	V	E	N	G	G	K	H	K	T	G	P	N	L	S	G	L	F	G	R	K	T	G	Q	A
Insects	*Samia cynthia* (a moth)					h	G	V	P	A	G	N	A	E	N	G	K	K	I	F	V	Q	R	C	A	Q	C	H	T	V	E	A	G	G	K	H	K	V	G	P	N	L	H	G	F	Y	G	R	K	T	G	Q	A
	Tobacco hornworm moth					h	G	V	p	A	G	N	A	D	N	G	K	K	I	F	V	Q	R	C	A	Q	C	H	T	V	E	A	G	G	K	H	K	V	G	P	N	L	H	G	F	F	G	R	K	T	G	Q	A
	Screwworm fly					h	G	V	P	A	G	D	V	E	K	G	K	K	I	F	V	Q	R	C	A	Q	C	H	T	V	E	A	G	G	K	H	K	V	G	P	N	L	H	G	L	F	G	R	K	T	G	Q	A
	Drosophila (fruit fly)					h	G	V	P	A	G	D	V	E	K	G	K	K	L	F	V	Q	R	C	A	Q	C	H	T	V	E	A	G	G	K	H	K	V	G	P	N	L	H	G	L	I	G	R	K	T	G	Q	A
Fungi	Baker's yeast				h	T	E	F	K	A	G	S	A	K	K	G	A	T	L	F	K	T	R	C	L	Q	C	H	T	V	E	K	G	G	P	H	K	V	G	P	N	L	H	G	I	F	G	R	H	S	G	Q	A
	Candida krusei (a yeast)			h	P	A	P	F	E	Q	G	S	A	K	K	G	A	T	L	F	K	T	R	C	A	Q	C	H	T	I	E	A	G	G	P	H	K	V	G	P	N	L	H	G	I	F	S	R	H	S	G	Q	A
	Neurospora crassa (a mold)					h	G	F	S	A	G	D	S	K	K	G	A	N	L	F	K	T	R	C	A	Q	C	H	T	L	E	E	G	G	G	N	K	I	G	P	A	L	H	G	L	F	G	R	K	T	G	S	V
Higher plants	Wheat germ	a	A	S	F	S	E	A	P	P	G	N	P	D	A	G	A	K	I	F	K	T	K	C	A	Q	C	H	T	V	D	A	G	A	G	H	K	Q	G	P	N	L	H	G	L	F	G	R	Q	S	G	T	T
	Buckwheat seed	a	A	T	F	S	E	A	P	P	G	N	I	K	S	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	E	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T	T
	Sunflower seed	a	A	S	F	A	E	A	P	P	G	D	P	T	T	G	A	K	I	F	K	T	K	C	A	Q	C	H	T	V	E	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T	T
	Mung bean	a	A	S	F	B	E	A	P	P	G	B	S	K	S	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	D	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T	T
	Cauliflower	a	A	S	F	B	E	A	P	P	G	B	S	K	A	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	D	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T	T
	Pumpkin	a	A	S	F	B	E	A	P	P	G	B	S	K	A	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	D	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T	T
	Sesame seed	a	A	S	F	B	E	A	P	P	G	B	V	K	S	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	D	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T	T
	Castor bean	a	A	S	F	B	E	A	P	P	G	B	V	K	A	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	E	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T	T
	Cottonseed	a	A	S	F	Z	E	A	P	P	G	B	A	K	A	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	D	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T	T
	Abutilon seed	a	A	S	F	Z	E	A	P	P	G	B	A	K	A	G	E	K	I	F	K	T	K	C	A	Q	C	H	T	V	E	K	G	A	G	H	K	Q	G	P	N	L	N	G	L	F	G	R	Q	S	G	T	T
	Number of different amino acids										1	3	5	5	5	1	3	3	4	1	4	3	2	1	3	1	1	1	1	4	2	4	1	2	3	2	1	4	1	1	2	1	5	1	3	3	2	1	3	2	1	3	3

D Asp
E Glu

H His
K Lys
R Arg
X TrimethylLys

B Asn or Asp
G Gly
N Asn
Q Gln
S Ser
T Thr
W Trp
Y Tyr
Z Gln or Glu

A Ala
C Cys
F Phe
I Ile
L Leu
M Met
P Pro
V Val

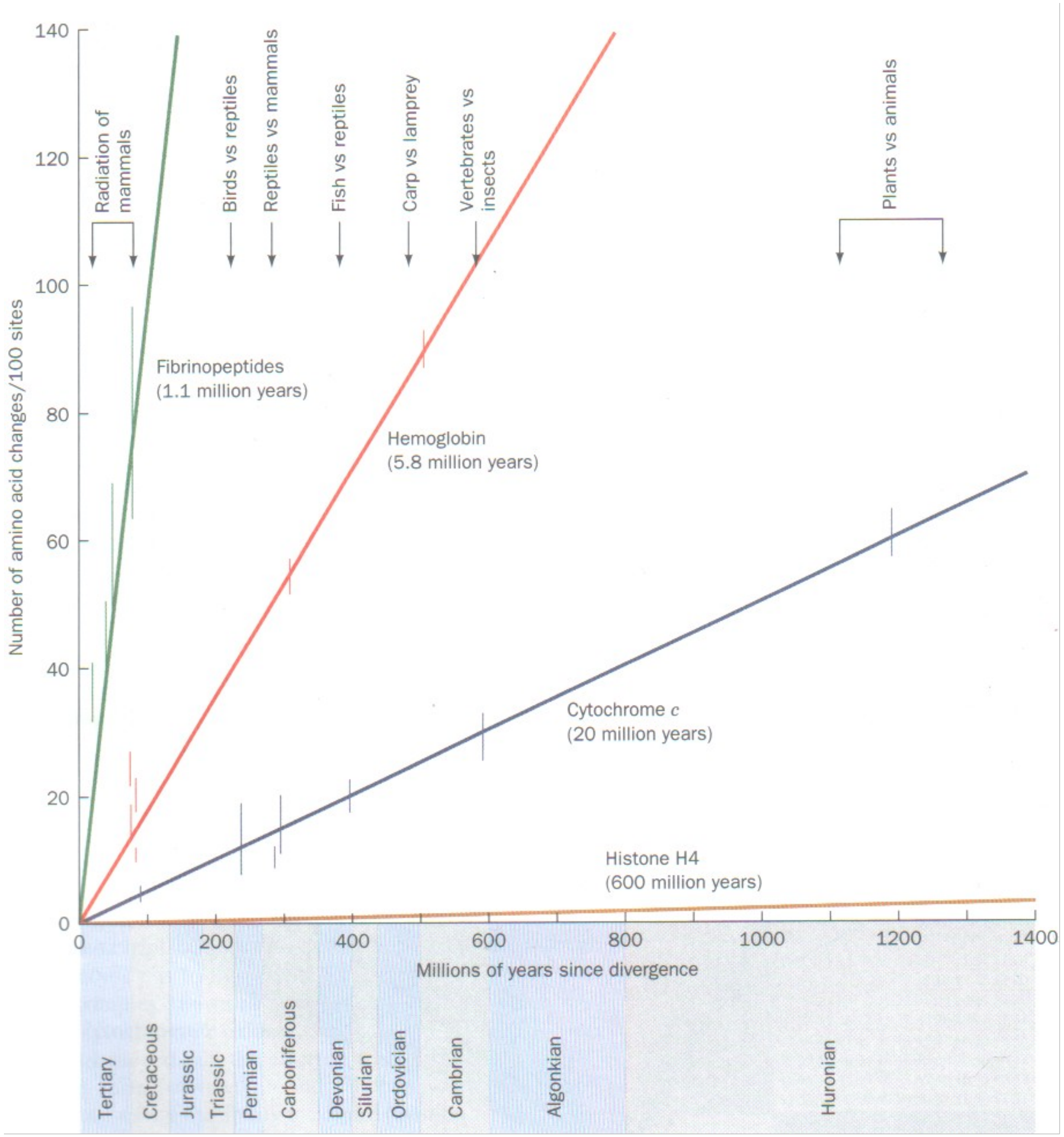
Evolutionary Rates

Evolution distance (accumulated changes) can be plotted vs time since divergence to yield an evolutionary rate (slope of line)

- Rate expressed as the time req'd to accumulate 1% change since divergence

Rate of accumulation < rate of random mutation

- Random mutations can be reversed
- Rate of accumulation limited by invariant and conservative residues



Mutational Rates

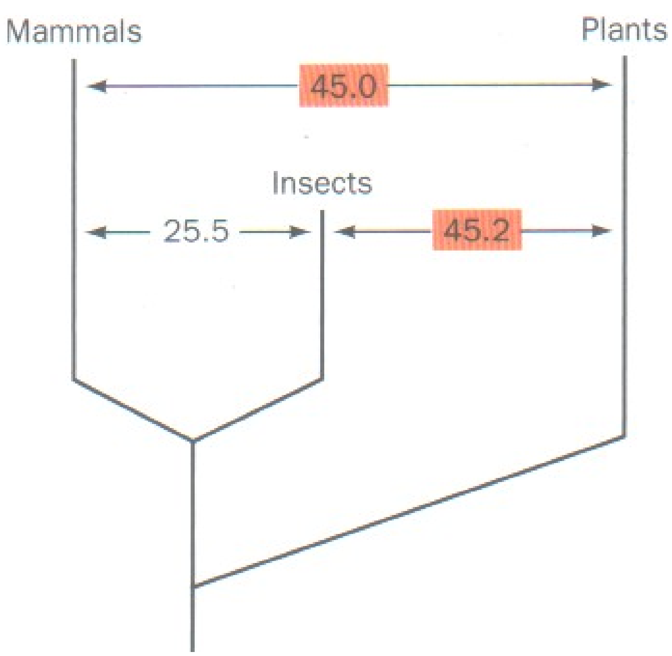
What mechanism accounts for observed mutations?

Model 1 - Mutations primarily arise from replication errors and the rate of mutation depends upon the number of organism generations

Model 2 – Mutations primarily arise from chemical modification and the rate of mutation depends upon the absolute time

Comparing the rate of mutation of insects (short generations) and mammals (long generations) for cytochrome C (figure)

- **Each is an equal evolutionary distance from plants suggesting mutations primarily arise from random CHEMICAL MODIFICATION of DNA**



Insects and mammals diverge from plants at same point yet are same evolutionary distance from plants.

Organismal Evolution

- **Protein evolution and taxonomical analysis generate very similar phylogenetic trees**
- **Protein evolution is not sufficient to explain all differences in organismal evolution**

eg. **Small evolutionary distances between Human and Chimpanzee (>99% sequence identity among proteins) suggest a closer relationship than anatomical and behavioral studies**

Mutations affecting regulation and scheduling of protein production affect the evolution of organisms more strongly than the average rate of mutation of proteins