Protein evolution

Chapter 3

Proteins

TE Creighton

Evolution

- In nature the macroscopic diversity we see stems from a molecular unity.
  - All organisms use the same 20 aminoacids
  - All organisms use the same nucleotides for DNA, RNA
  - The aminoacids are all L amino acids
  - Ribose units of DNA and RNA arre all D isomers.
  - The basic properties of the proteins in E.coli and Humans are the same.
  - The function of DNA and RNA pertains, but the differences between procaryotes and eucaryotes becomes significant

Evolution

- The biochemical similarities extend to even higher levels of organisation
  - Biochemical pathways
  - Metabolism

- All this can be explained so far, with the presence of a common ancestor for all living organisms.
- This cell had already acquired all basic biochemical features.

Evolution

- The diversity at higher level of organisation has arisen
- From Darwinian divergence
- But the most basic functions of an organism have been conserved.
- Reason for this: any alteration in these functions
- LETHAL !!!!!!!!

Evolution

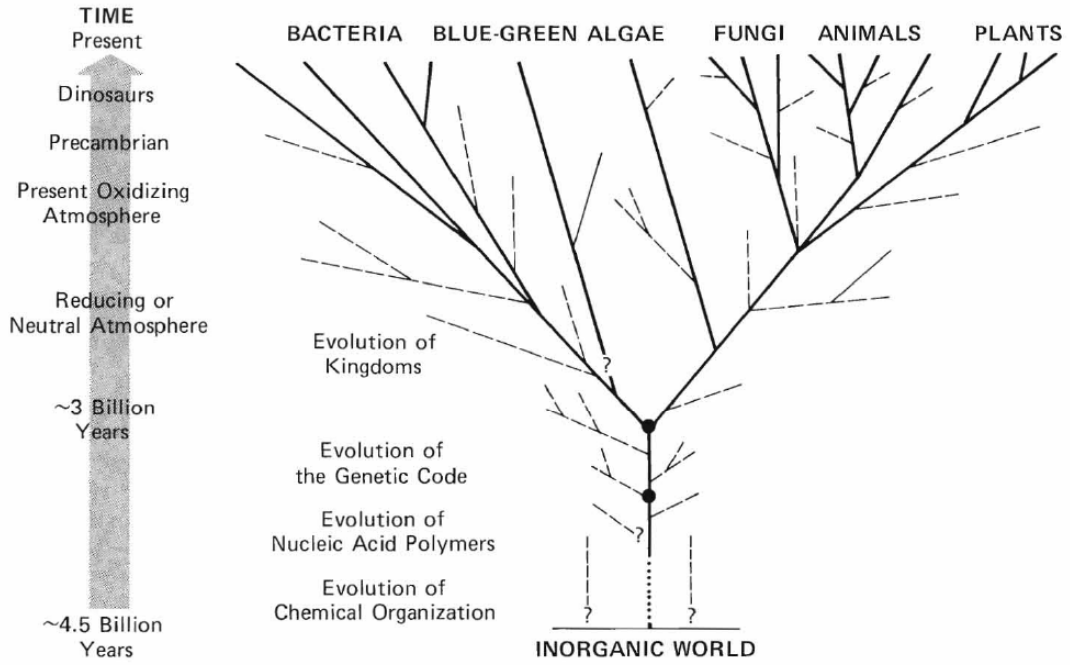


FIGURE 3.1
General phylogenetic tree of how life on earth might have evolved. The dashed lines represent hypothetical lineages that have not survived to the present day. (Adapted from M. O. Dayhoff, *Atlas of Protein Sequence and Structure,* National Biomedical Research Foundation, Washington, D.C., 1972.)

Evolution

- Comparisons at the macroscale:
- comparisons of the organsims,
- the fossil record
- But even more compelling is the comparison on the protein level.
- These studies provide deeper insights into how evolution works

Evolution

- E.g. A DNA chain of of only 1000 nucleotides
- Can exist with any one of 4^{1000} different nucleotide sequences.
- A small protein of 100 aminoacids can have 20^{100} different sequences.
- Compared to the mere numbers of possible proteins the existing proteins are determined by their ancesters
- An existing protein serves its function. The variation is within the limit of function.'

Primordial origins of life

- Life started out of a primordial soup
- Where all components have been made out of simple molecules
- Forming the primordial soup
- All the processes have been deomnstrated in the laboratory
- The chicken-egg problem of molecular biology:
- Who cames first DNA,RNA or protein

The accepted picture

- Since the finding of Ribozymes:
- First RNA and Ribozymes
- Thereafter proteins
- And DNA since DNA is chemically more stable as RNA
- Remainders of thsi RNA world are:
- mRNA
- T-RNA
- Ribosomes (hybrids RNA-Protein)
- Coenzymes: ATP, NAD^+, FAD composed of RNA moieties

Evolutionary divergence of proteins

- Similar sequences mean descent from a common ancestor.
- The number of possible RNA,DNA or protein sequences is so great that it is implausible that similar long sequences could have arisen by any mechanism other than divergence from the same ancestor.
- Proteins and nucleic acids that have evolved from a common ancestor are homologous.
- After the first replication, mutations have started in the two sequences

Homologous genes and proteins

- Proteins and genes are either Homologous or **NOT**.
- Either or not from a common ancestor.

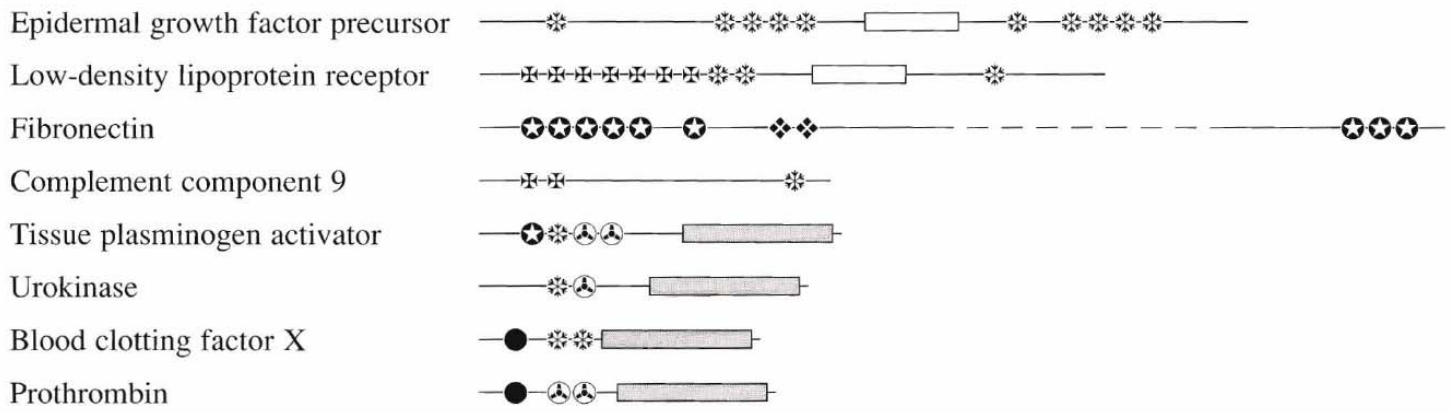
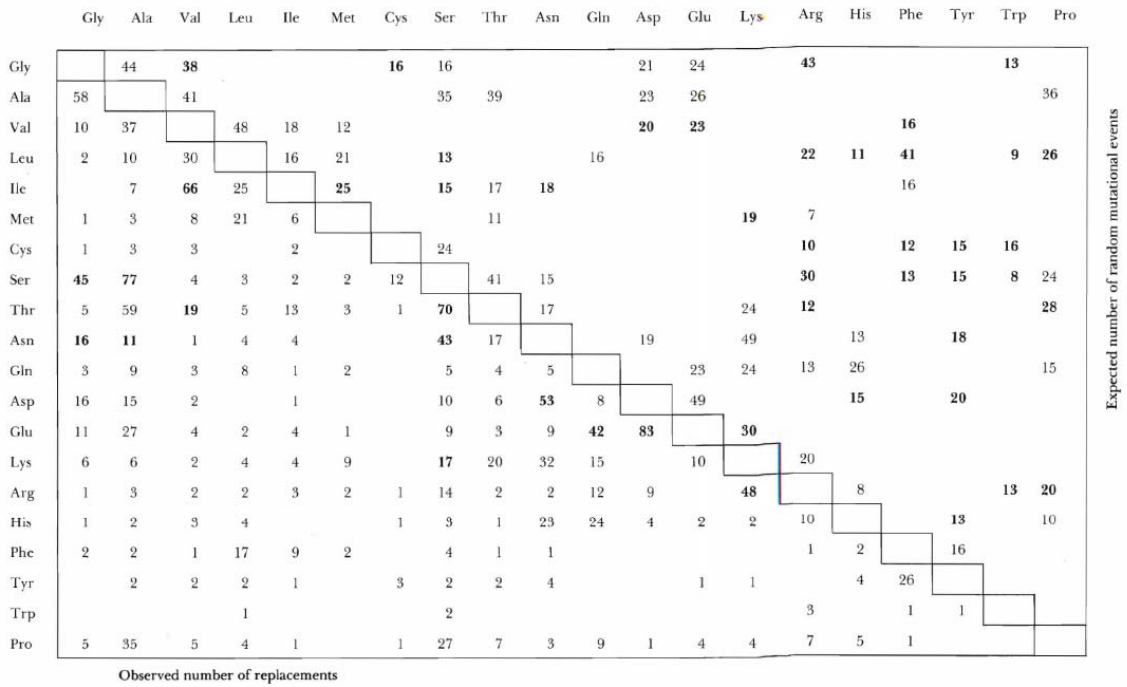


FIGURE 3.16
Mosaic structures of some vertebrate proteins. Each symbol represents members of a set of homologous segments of polypeptide chain, which in these examples are also independent structural units, or domains, of the protein. Those with distinct structural or functional properties are the serine protease domains (shaded rectangles), Ca^{2+}-binding domains containing γ-carboxy-Glu residues (●), so-called kringle domains (✇), and epidermal growth factor (EGF) domains (❄).

Homologous Genes and Proteins

- The other explanation Than divergence is
- Convergence
- Convergence means two different sequences have acquired the same function ality and share a certain degree of homology.
- It exists at the macroscopic level but there are no instances where it could have been shown conclusively for nucleic acids and proteins.
- But it can not be dismissed

Sequence alignments



	Gly	Ala	Val	Leu	Ile	Met	Cys	Ser	Thr	Asn	Gln	Asp	Glu	Lys	Arg	His	Phe	Tyr	Trp	Pro
Gly		44	**38**				**16**	16				21	24		**43**				**13**	
Ala	58		41					35	39			23	26							36
Val	10	37		48	18	12						**20**	**23**				**16**			
Leu	2	10	30		16	21		**13**			16				**22**	**11**	**41**		**9**	**26**
Ile		7	**66**	25		**25**		**15**	17	**18**							16			
Met	1	3	8	21	6				11					**19**	7					
Cys	1	3	3		2			24							**10**		**12**	**15**	**16**	
Ser	**45**	**77**	4	3	2	2	12		41	15					**30**		**13**	**15**	8	24
Thr	5	59	**19**	5	13	3	1	**70**		17				24	**12**					**28**
Asn	**16**	**11**	1	4	4			**43**	17			19		49		13		**18**		
Gln	3	9	3	8	1	2		5	4	5			23	24	13	26				15
Asp	16	15	2		1			10	6	**53**	8		49			**15**		**20**		
Glu	11	27	4	2	4	1		9	3	9	**42**	**83**		**30**						
Lys	6	6	2	4	4	9		**17**	20	32	15		10		20					
Arg	1	3	2	2	3	2	1	14	2	2	12	9		**48**		8			**13**	**20**
His	1	2	3	4			1	3	1	23	24	4	2	2	10			**13**		10
Phe	2	2	1	17	9	2		4	1	1					1	2		16		
Tyr		2	2	2	1		3	2	2	4			1	1		4	26			
Trp				1				2							3		1	1		
Pro	5	35	5	4	1		1	27	7	3	9	1	4	4	7	5	1			

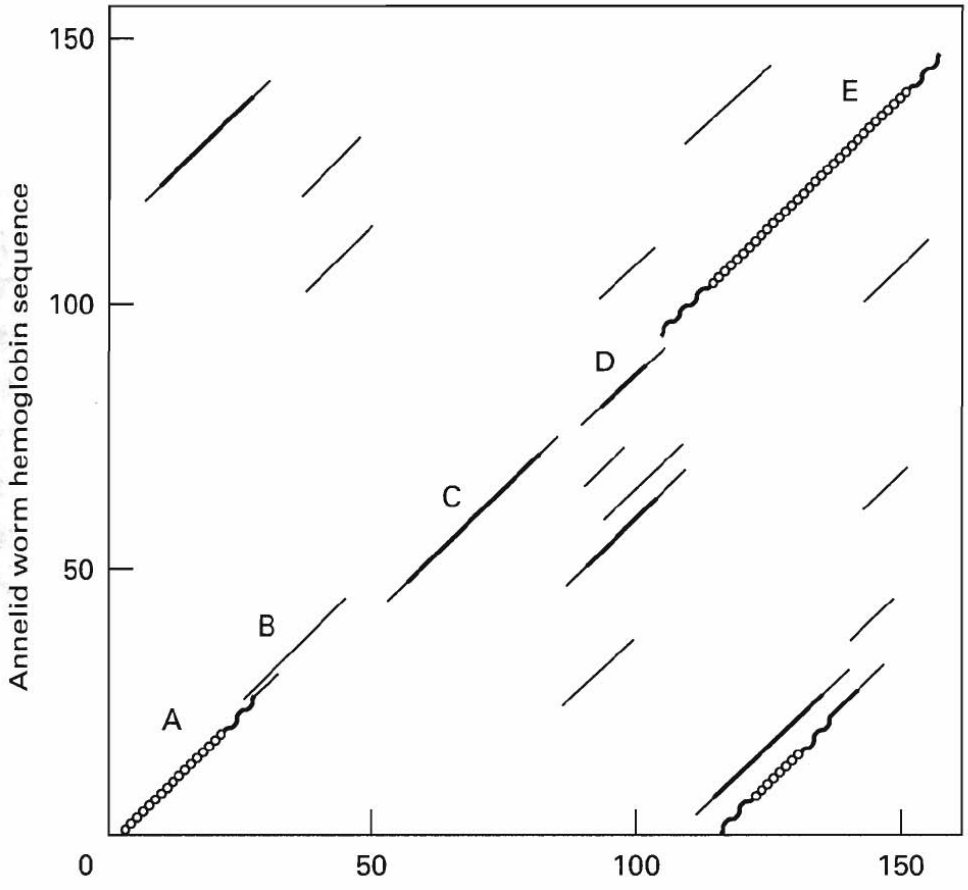
Observed number of replacements

Expected number of random mutational events

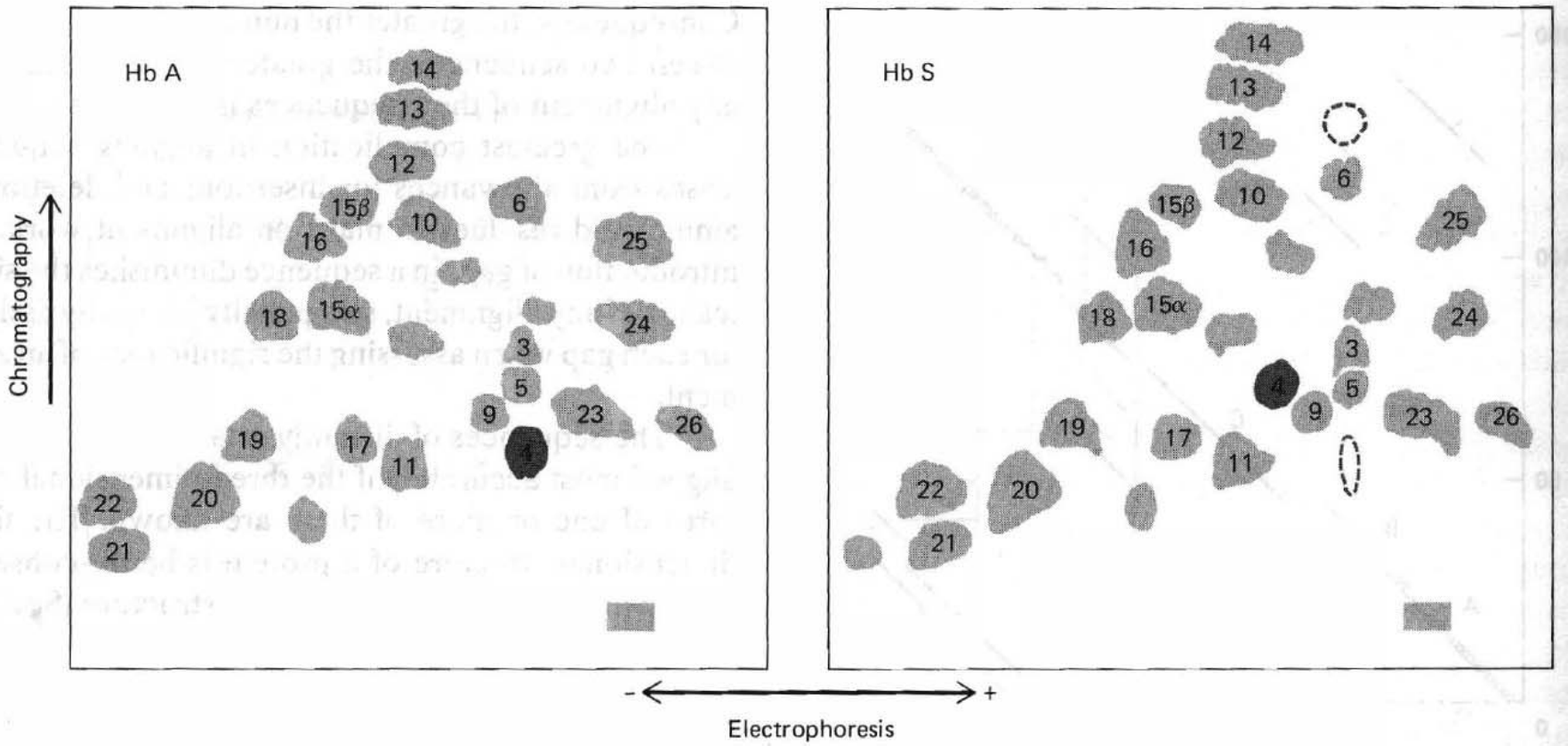
FIGURE 3.2

Relative frequences of amino acid replacements in a total of 1572 examples between closely related proteins that are observed *(bottom, left)* and those expected for random single-nucleotide mutations *(top, right)*. The greatest discrepancies between the observed and random replacements are shown in boldface type. Replacements involving chemically similar amino acids are generally observed to be much more frequent than expected for random mutations. (Observed replacements compiled by M. O. Dayhoff, *Atlas of Protein Sequence and Structure*, vol. 5, suppl. 3, National Biomedical Research Foundation, Washington, D.C., 1978.)

Diagonal dot blot



Mutation and Protein structure



Mutation and protein structure

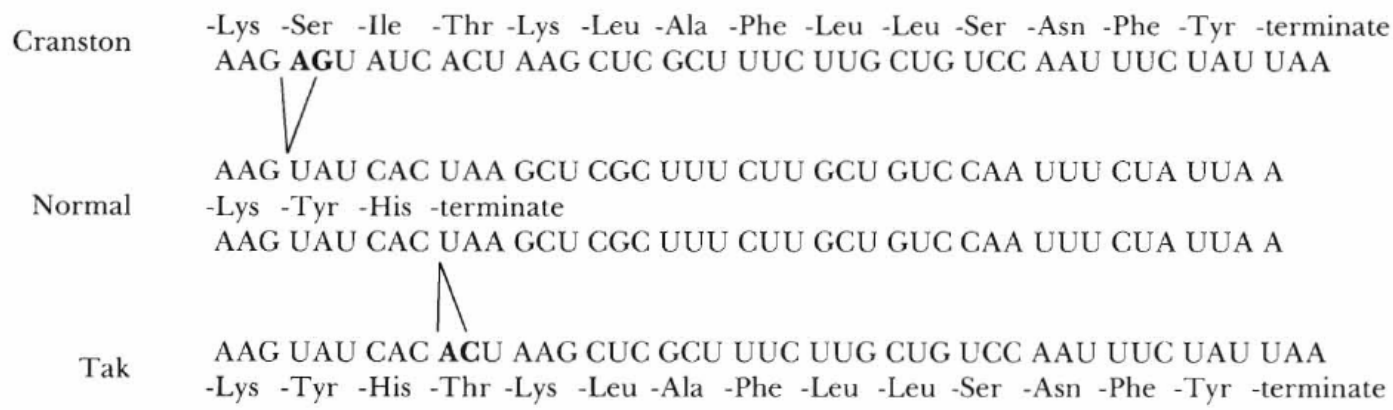


FIGURE 3.5

Insertion mutations that alter the reading frame of the gene for the β chain of hemoglobin. The normal sequences of mRNA and of the protein product are shown, as well as the sequences for two mutants, *Cranston* and *Tak*. The nucleotides inserted in the mutants are in boldface type; note that both insertions apparently occurred by duplication of the two preceding nucleotides. (Sequences from N. Proudfoot and G. Brownlee, *Brit. Med. Bull.* 32:251–256, 1976.)

Reconstructing evolution

	1									10										20										30										40										48
Human and chimpanzee	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	P	G	Y	S	Y
Pig, bovine, and sheep	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	P	G	*F*	S	Y
Gray kangaroo	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	*I*	F	G	R	K	T	G	Q	A	P	G	*F*	*T*	Y
Chicken and turkey	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	E	G	*F*	S	Y
Snapping turtle	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	E	G	*F*	S	Y
Puget Sound dogfish	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	Q	G	*F*	S	Y
Pacific lamprey	G	D	V	E	K	G	K	K	*V*	F	*V*	Q	K	C	S	Q	C	H	T	V	E	K	A	G	K	H	K	T	G	P	N	L	S	G	L	F	G	R	K	T	G	Q	A	P	G	*F*	S	Y
Garden snail	G	Z	A	Z	K	G	K	K	I	F	T	Q	K	C	L	Q	C	H	T	V	E	A	G	G	K	H	K	T	G	P	N	L	S	G	L	F	G	R	K	Q	G	Q	A	P	G	*F*	A	Y
Screw-worm fly	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	A	G	*F*	A	Y
Tobacco hornworm moth	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	P	G	*F*	S	Y
Candida krusei	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	E	G	Y	S	Y
Rust fungus	G	F	E	D	G	D	A	K	K	G	A	*R*	I	F	K	T	*R*	C	A	Q	C	H	T	*L*	G	A	G	E	P	N	K	V	G	P	N	L	H	G	L	F	G	R	R	*S*	G	T	V	E	G	*F*	S	Y
Rape and cauliflower	A	S	F	D	E	A	P	P	G	N	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	A	G	Y	S	Y

	49									60										70										80										90										100				104
Human and chimpanzee	T	A	A	N	K	N	K	G	I	I	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	V	G	I	K	K	K	E	E	R	A	D	L	I	A	Y	L	K	K	A	T	N	E
Pig, bovine, and sheep	T	D	A	N	K	N	K	G	I	T	W	G	E	*E*	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	G	E	R	E	D	L	I	A	Y	L	K	K	A	T	N	E
Gray kangaroo	T	D	A	N	K	N	K	G	I	I	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	G	E	R	A	D	L	I	A	Y	L	K	K	A	T	N	E
Chicken and turkey	T	D	A	N	K	N	K	G	I	T	W	G	E	D	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	S	E	R	V	D	L	I	A	Y	L	K	D	A	T	S	K
Snapping turtle	T	E	A	N	K	N	K	G	I	T	W	G	E	*E*	T	L	M	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	K	A	E	R	A	D	L	I	A	Y	L	K	D	A	T	S	K
Puget Sound dogfish	T	D	A	N	K	S	K	G	I	T	W	Q	Q	*E*	T	L	R	I	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	*L*	K	K	K	S	E	R	Q	D	L	I	A	Y	L	K	K	T	A	A	S
Pacific lamprey	T	D	A	N	K	S	K	G	I	*V*	W	N	Q	*E*	T	L	*F*	V	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	I	K	K	E	G	E	R	K	D	L	I	A	Y	L	K	K	S	T	S	E
Garden snail	T	D	A	N	K	G	K	G	I	T	W	K	N	Q	T	L	*F*	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	*V*	F	A	G	*L*	K	B	Z	T	E	R	V	D	L	I	A	Y	L	Z	Z	A	T	K	K
Screw-worm fly	T	N	A	N	K	A	K	G	I	T	W	Q	*D*	D	T	L	*F*	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	I	F	A	G	*L*	K	K	P	N	E	R	G	D	L	I	A	Y	L	K	S	A	T	K	
Tobacco hornworm moth	*S*	N	A	N	K	A	K	G	I	T	W	Q	*D*	D	T	L	*F*	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	*V*	F	A	G	*L*	K	K	A	N	E	R	A	D	L	I	A	Y	L	K	Q	A	T	K	
Candida krusei	T	D	A	N	K	R	A	G	*V*	E	W	A	E	P	T	*M*	S	*D*	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	A	F	G	G	*L*	K	K	A	K	*D*	R	N	D	L	*V*	T	Y	*M*	L	E	A	*S*	K
Rust fungus	T	D	A	N	K	K	A	G	Q	*V*	W	E	E	*E*	T	*F*	*L*	E	Y	L	E	N	P	K	K	Y	I	P	G	T	K	M	A	F	G	G	*L*	K	K	E	K	*D*	R	N	D	L	*V*	T	Y	L	*R*	E	E	T	K
Rape and cauliflower	*S*	A	A	N	K	N	K	A	*V*	E	W	E	E	K	T	L	Y	*D*	Y	L	L	N	P	K	K	Y	I	P	G	T	K	M	*V*	F	P	G	*L*	K	K	P	Q	*D*	R	A	D	L	I	A	Y	L	K	E	A	T	A	

FIGURE 3.6

Amino acid sequences of some cytochromes *c* from various eukaryotes. The one-letter code (see Table 1.1) is used; Z is either E or Q (Glu or Gln), B is either D or N (Asp or Asn). Residues that are identical to those in the human protein are boldface; chemically similar residues are in italics.

Reconstructing evolution

	Chimpanzee	Sheep	Rattlesnake	Carp	Snail	Moth	Yeast	Cauliflower	Parsnip
Human	0	10	14	18	29	31	44	44	43
Chimpanzee		10	14	18	29	31	44	44	43
Sheep			20	11	24	27	44	46	46
Rattlesnake				26	28	33	47	45	43
Carp					26	26	44	47	46
Garden snail						28	48	51	50
Tobacco hornworm moth							44	44	41
Baker's yeast (iso-1)								47	47
Cauliflower									13

FIGURE 3.7
Amino acid difference matrix for cytochromes *c* of a few representative species. These proteins all consist of at least 104 amino acid residues, so at least half the residues of each pair are identical.

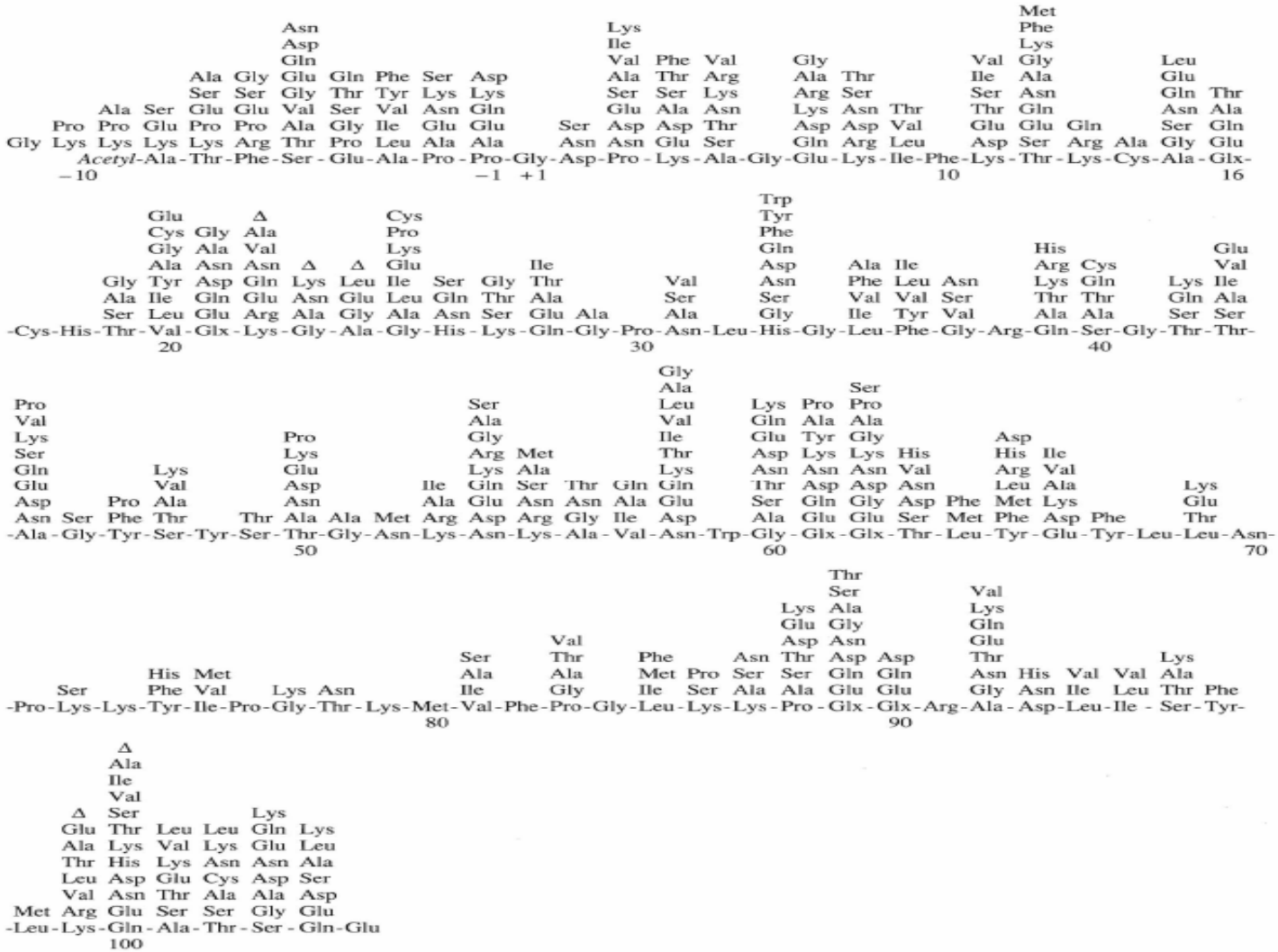


FIGURE 3.8

Composite amino acid sequence of cytochromes *c* from 92 eukaryotic species. The continuous sequence of 113 residues is the longest cytochrome *c*, that of *Gingko bilboa*. The numbering of the residues is that of the mammalian proteins, which start at position 1 and end at 104. The proteins included here start at positions −11 to +1 and end at positions 101 to 105. Deletions of residues are indicated by Δ. (The sequences were obtained from M. O. Dayhoff, *Atlas of Protein Sequences and Structure*, National Biomedical Research Foundation, Washington, D.C., 1972; and from D. M. Hampsey et al., *J. Biol. Chem.* 261:3259–3271, 1986.)

Reconstructing evolution

Table 3.1 *Relative Variabilities of Amino Acid Residues during Divergence of Homologous Proteins*[a]

Residue	Variability	Residue	Variability	Residue	Variability
Asn	100	Met	70	Gly	37
Ser	90	Gln	69	Tyr	31
Asp	79	Val	55	Phe	31
Glu	76	His	49	Leu	30
Ala	75	Arg	49	Cys	15
Thr	72	Lys	42	Trp	13
Ile	72	Pro	42		

[a] The number of times that a given amino acid residue has changed during the evolution of various proteins has been divided by the number of times the residue occurred. These values have been normalized by setting the highest to 100.

Adapted from M. O. Dayhoff, *Atlas of Protein Sequence and Structure,* vol. 5, suppl. 3, National Biomedical Research Foundation, Washington, D.C., 1978.

Phylogenetic trees

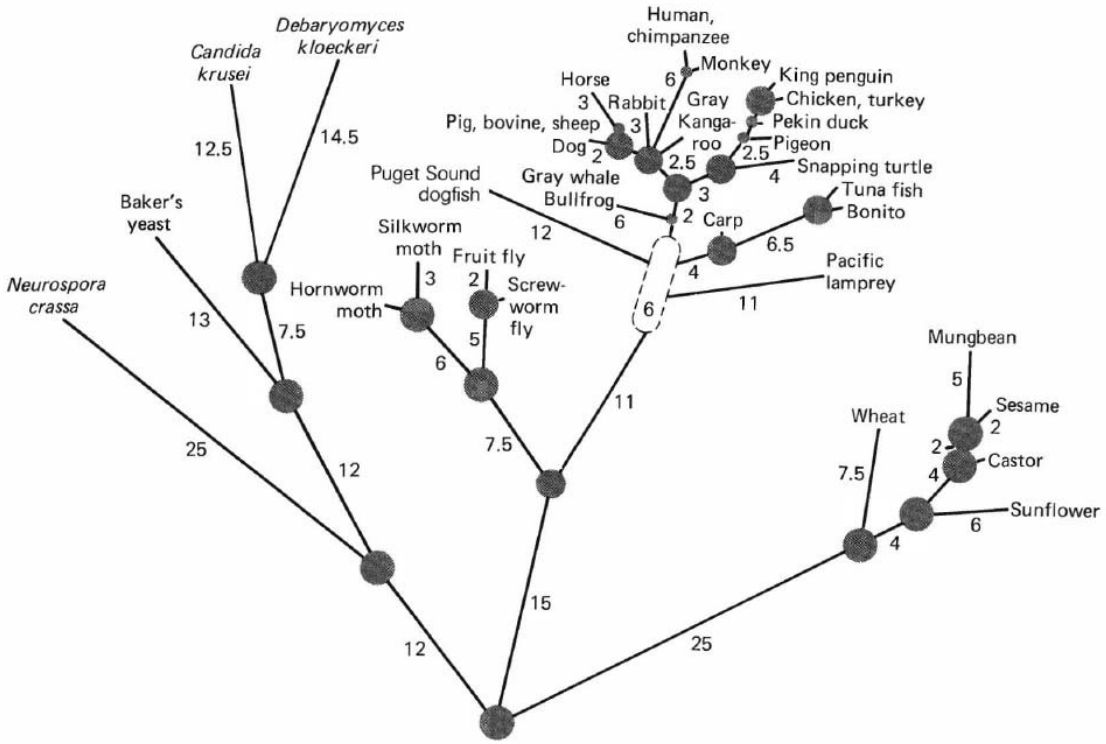
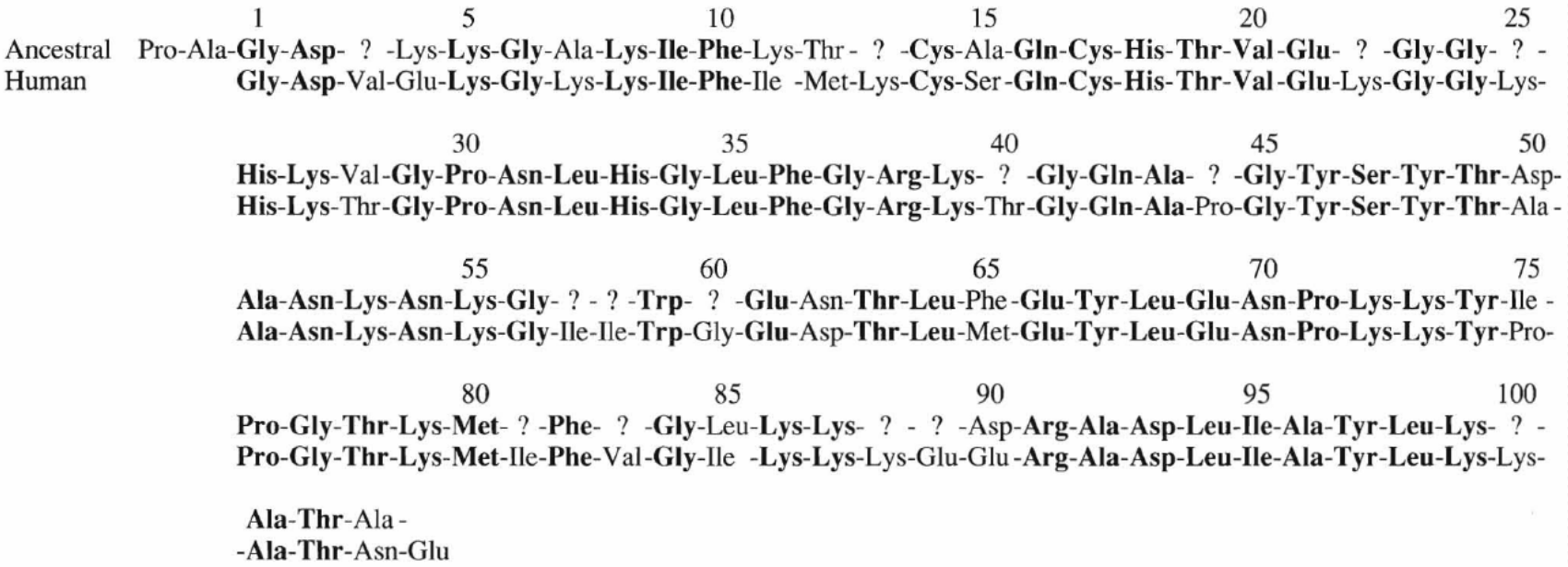


FIGURE 3.9

Phylogenetic tree constructed from the sequences of cytochromes *c* by minimizing the total number of amino acid replacements. The sequences of the ancestors at each branch point were inferred from the present-day sequences. The length of each branch is proportional to the indicated number of amino acid changes that are believed to have occurred. The branch points within the dashed oval were not adequately defined by the sequences. (Adapted from M. O. Dayhoff, *Atlas of Protein Sequence and Structure,* National Biomedical Research Foundation, Washington, D.C., 1972.)

Phylogenetic trees



Phylogenetic trees

Table 3.2 *Relative Extents of Divergence for Several Proteins from Different Species*

Protein	Residues Differing from Human Protein (%)								
	Rhesus monkey	Cow	Pig	Rabbit	Chicken	Frog	Fish	Fruit fly	Yeast
Cytochrome *c*	1	10	10	9	13	17	20	27	41
Hemoglobin α chain	3	12	13	18	25		50		
Hemoglobin β chain	5	17	16	10	26	46			
Fibrinopeptides A and B	30	70	67	70					

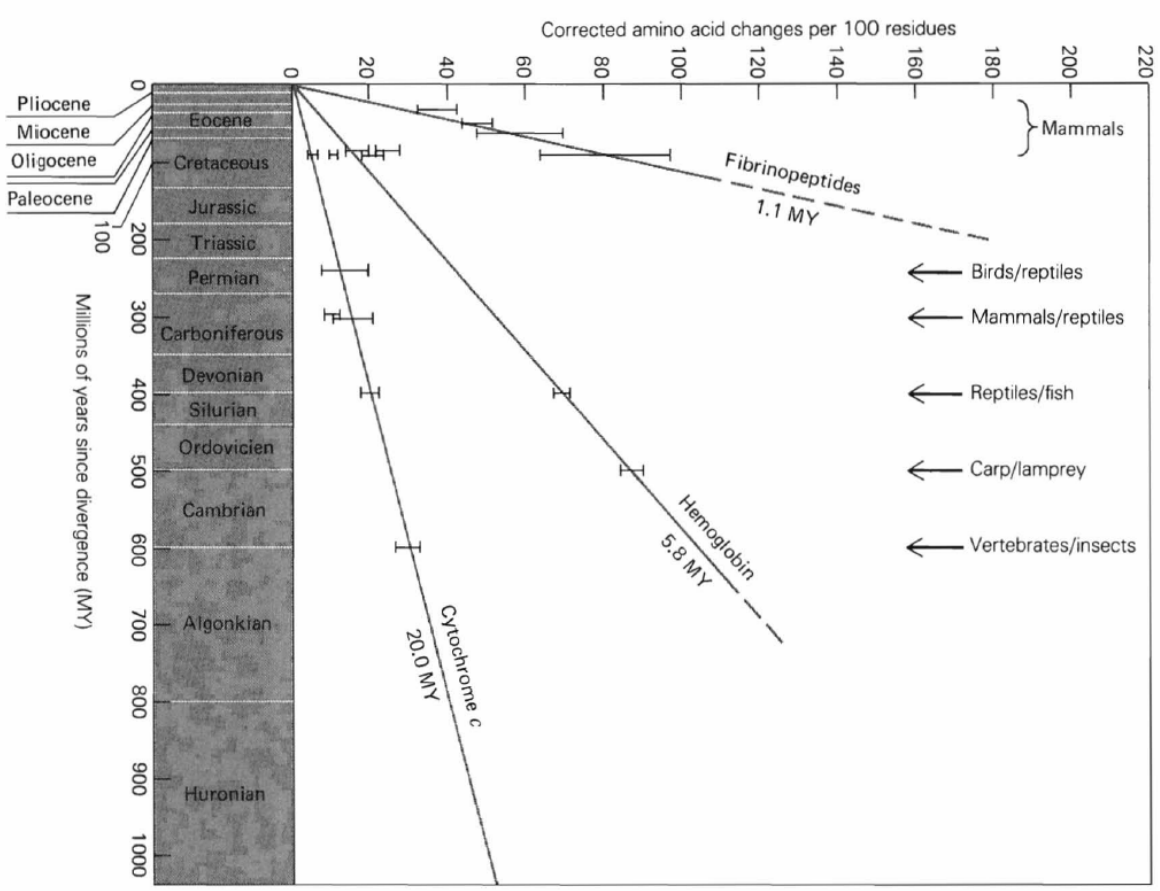
Rates of Divergence

Table 3.3 *Rates of Evolution of Some Proteins*

Protein	Accepted point mutations [(number)/(100 residues) · (10^8 years)]	Protein	Accepted point mutations [(number)/(100 residues) · (10^8 years)]
Histones		Hormones *cont.*	
H4	0.25	Proparathyrin	14
H3	0.30	Prolactin	20
H2A	1.7	Growth hormone	25
H2B	1.7	Lutropin β chain	33
H1	12	Insulin C peptide	53
Fibrous proteins		Oxygen-binding proteins	
Collagen (α-1)	2.8	Myoglobin	17
Crystallin (αA)	4.5	Hemoglobin α chain	27
Intracellular enzymes		Hemoglobin β chain	30
Glutamate dehydrogenase	1.8	Secreted enzymes	
Triosephosphate dehydrogenase	5.0	Trypsinogen	17
Triosephosphate isomerase	5.3	Lysozyme	40
Lactate dehydrogenase H	5.3	Ribonuclease A	43
Lactate dehydrogenase M	7.7	Immunoglobulins	
Carbonic anhydrase B	25	κ chains (V region)	100
Carbonic anhydrase C	48	κ chains (C region)	111
Electron carriers		λ chains (V region)	125
Cytochrome *c*	6.7	γ chains (V region)	143
Cytochrome b_5	9.1	Snake venom toxins	
Plastocyanin	14	Long neurotoxins	111
Ferredoxin	17	Cytotoxins	111
Hormones		Short neurotoxins	125
Glucagon	2.3	Other proteins	
Corticotropin	4.2	Parvalbumin	20
Insulin	7.1	Albumin	33
Thyrotropin β chain	11	α-Lactalbumin	43
Lipotropin β chain	13	Fibrinopeptide A	59
Lutropin α chain	14	Casein κ chain	71
		Fibrinopeptide B	91

From A. C. Wilson et al., *Ann. Rev. Biochem.* 46:573–639 (1977).

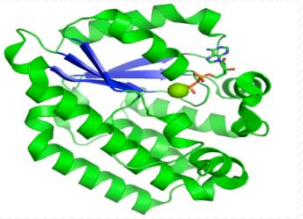
Rates of divergence



The nine most ancient protein folds

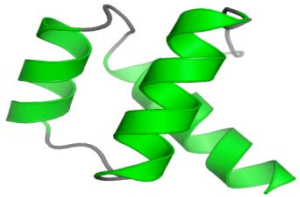
SCOP label	Fold
c.37	P-loop containing nucleoside triphosphate hydrolases *3 layers with α/β/α arrangement, parallel or mixed β-sheets of variable sizes*
a.4	DNA/RNA-binding 3-helical bundle *Core: 3-helices; closed or partly opened bundle, right-handed twist; up-and-down*
c.1	TIM β/α-barrel *Closed barrel with parallel β-sheet and strand order 12345678; n=8, S=8*
c.2	NAD(P)-binding Rossmann-fold domains *Core: 3 layers in α/β/α arrangement; parallel β-sheet of 6 strands, order 321456*
d.58	Ferredoxin-like *Core: 3 helices; closed or partly opened bundle, right-handed twist; up-and-down*
c.23	Flavodoxin-like *3 layers with α/β/α arrangement; parallel β-sheet of 5 strands, order 21345*
c.55	Ribonuclease H-like motif *3 layers with α/β/α arrangement; mixed β-sheet of 5 strands, order 32145 with strand 2 antiparallel to the rest*
b.40	OB-fold *Closed or partly opened barrel, n=5, S=10 or S=8 with greek-key motif*
c.66	S-adenosyl-L-methionine-dependent methyltransferases *Core: 3 layers with α/β/α arrangement; mixed β-sheet of 7 strands, order 3214576 with strand 7 antiparallel to the rest*

Deoxycytidine kinase



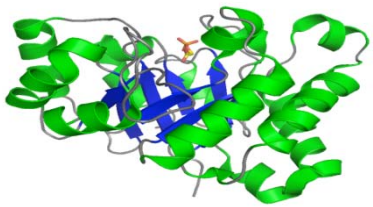
c.37: P-loop containing NTP hydrolase

Engrailed homeodomain



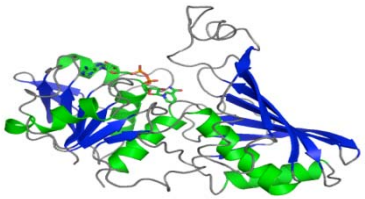
a.4: DNA/RNA binding 3 helix bundle

Triose phosphate isomerase



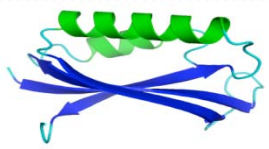
c.1: 8-stranded α/β barrel

Glyceraldehyde phosphate dehydrogenase



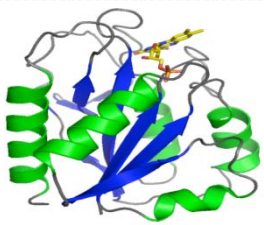
c.2: NAD(P) binding Rossman fold

Ribosomal protein S6



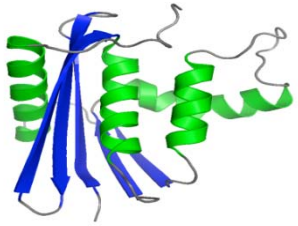
d.58: Ferredoxin-like

Flavodoxin



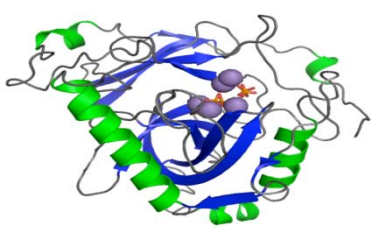
c.23: Flavodoxin-like

Ribonuclease H



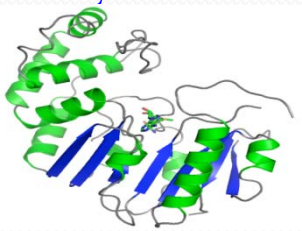
c.55: Ribonuclease H-like

Pyrophosphatase



b.40: OB fold

Adenine methyltransferase

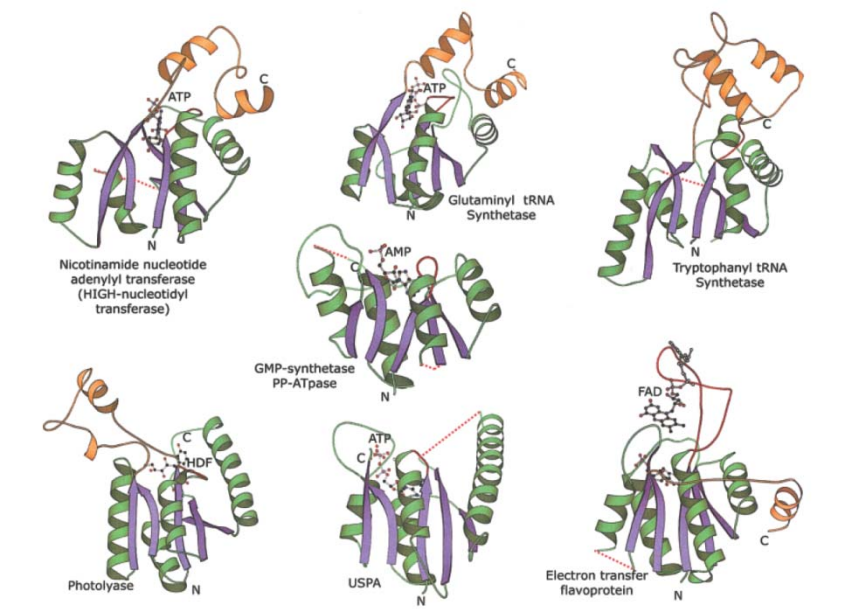


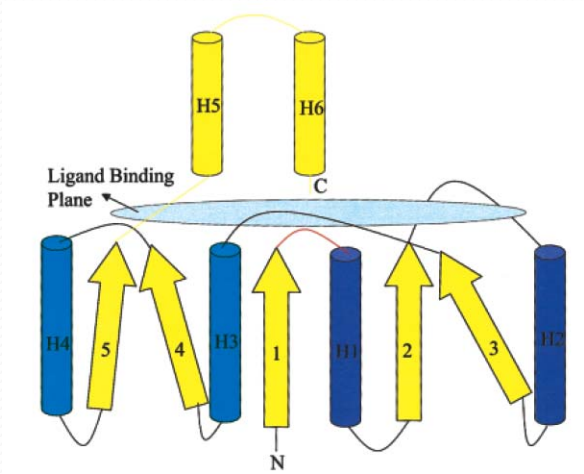
c.66: SAM-dependent methyltransferases

Not all widely diversified and broadly functional folds are ancient

- Adenine nucleotide binding protein

Scop C.28: also called 'HUP' domains





Gene rearrangements

Table 3.4 *Number of Amino Acid Identities between Various Members of the Human Globin Family*

	Hemoglobin				
	α	β	γ	δ	ϵ
Myoglobin	38	36	36	37	34
Hemoglobin α		64	59	63	55
Hemoglobin β			107	136	110
Hemoglobin γ				105	116
Hemoglobin δ					106

Gene rearrangements

FIGURE 3.12

Amino acid sequences of members of the human globin family, myoglobin plus the hemoglobin polypeptide chains. Positions where three or more of the amino acid residues are identical are in boldface; chemically similar residues are in italics.

Gene rearrangements

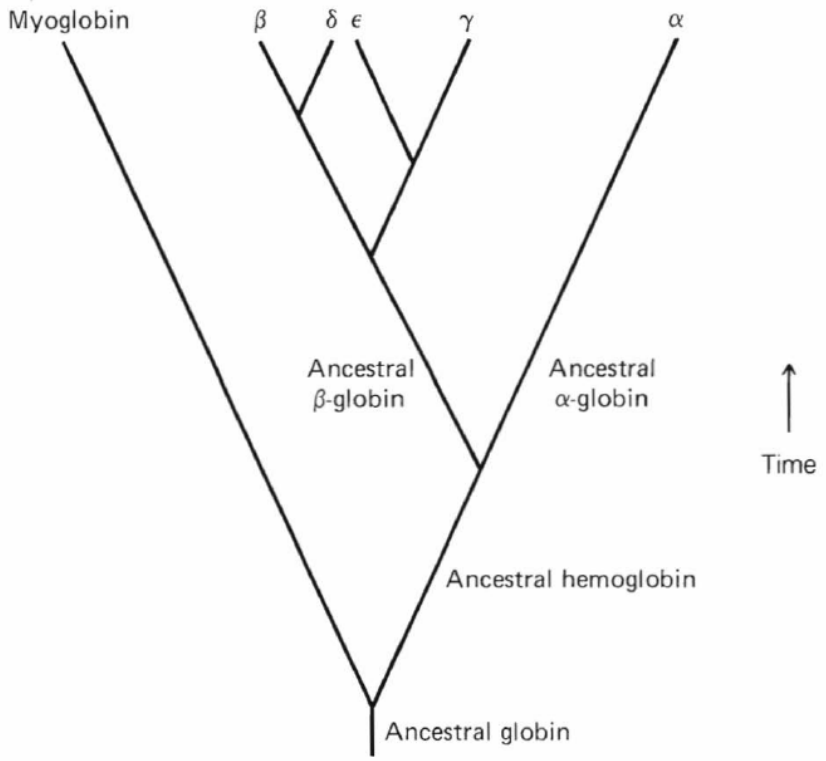
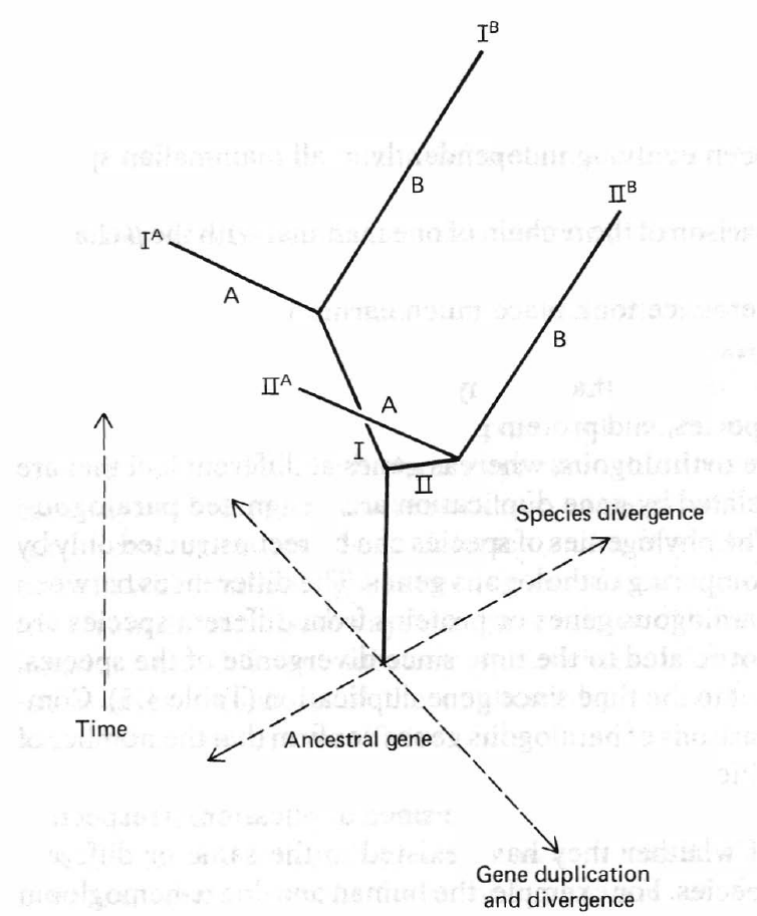


FIGURE 3.13
Evolutionary scheme for the divergence of the globin chains, inferred from the relative similarities of their gene and protein sequences. Time increases from the bottom to the top of the figure. The separation of the $^{A}\gamma$ and $^{G}\gamma$ genes is not depicted here. (Adapted from A. Efstratiadis et al., *Cell* 21:653–668, 1980.)

Gene rearrangements



Paralogous comparison

Table 3.5 *Paralogous Comparisons of Globin Chains*[a]

Comparison protein	Number of Differences in Amino Acid Sequence: Human α-globin	Number of Differences in Amino Acid Sequence: Dog α-globin
Human β-globin	84	84
Human γ-globin	89	83
Human myoglobin	115	119
Dog β-globin	88	87
Dog myoglobin	115	118

[a] The globin genes duplicated and diverged long before the human and dog species diverged, so comparisons of the different globins are relevant to the time since the genes duplicated. Consequently, very similar results are obtained irrespective of which species is used.

Elongation by intragene duplication

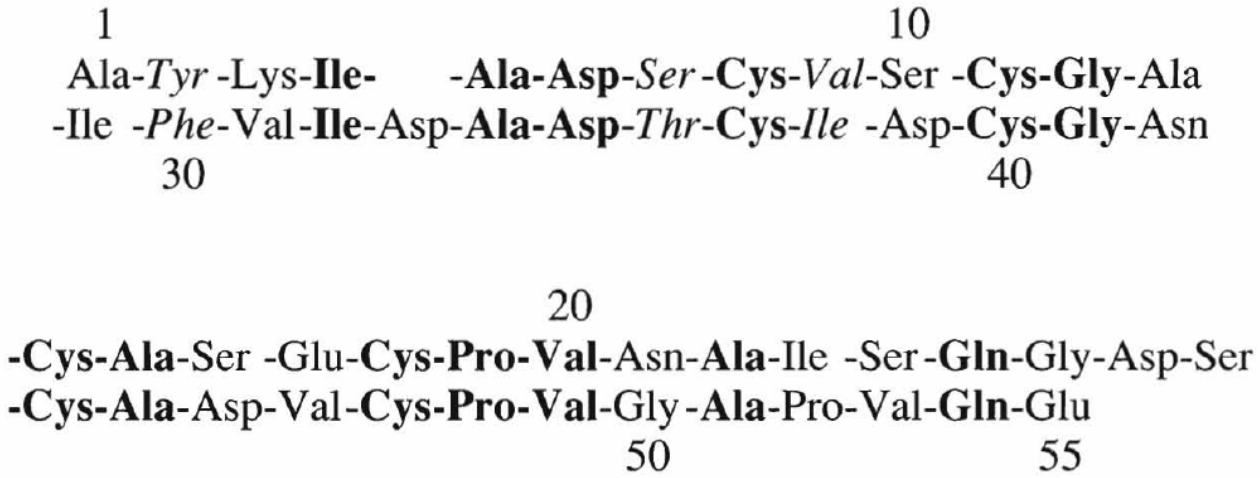


FIGURE 3.15
Homology between the two halves of the primary structure of *Clostridium pasteurianum* ferredoxin. Identical residues are in boldface, similar residues in italics.

Gene fusion and division

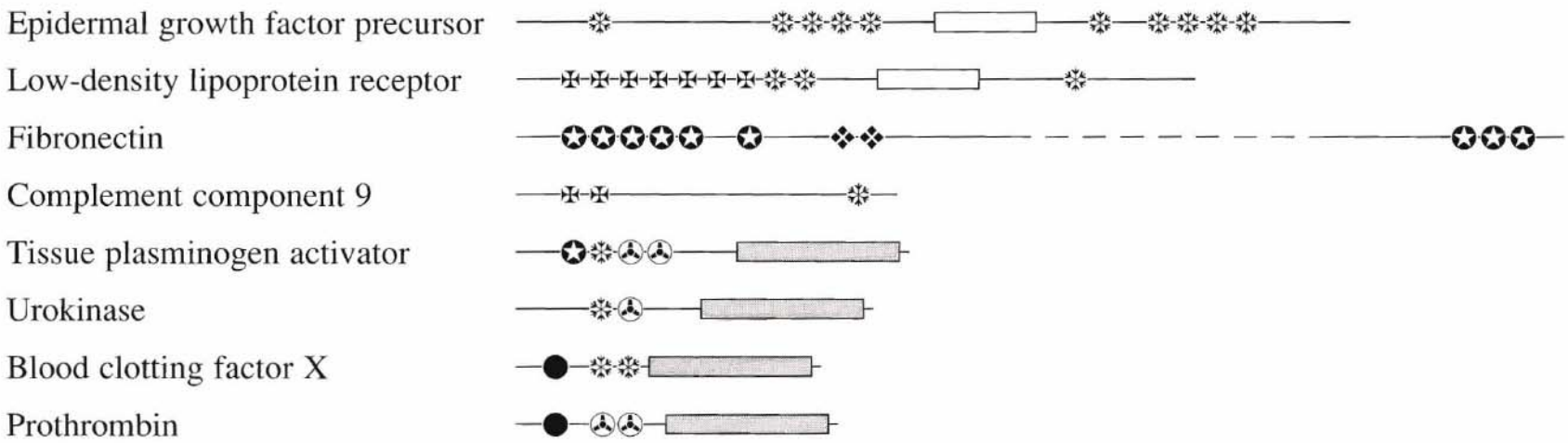


FIGURE 3.16
Mosaic structures of some vertebrate proteins. Each symbol represents members of a set of homologous segments of polypeptide chain, which in these examples are also independent structural units, or domains, of the protein. Those with distinct structural or functional properties are the serine protease domains (shaded rectangles), Ca^{2+}-binding domains containing γ-carboxy-Glu residues (●), so-called kringle domains (✇), and epidermal growth factor (EGF) domains (❄).

Hybrid protein by mismatch recombination

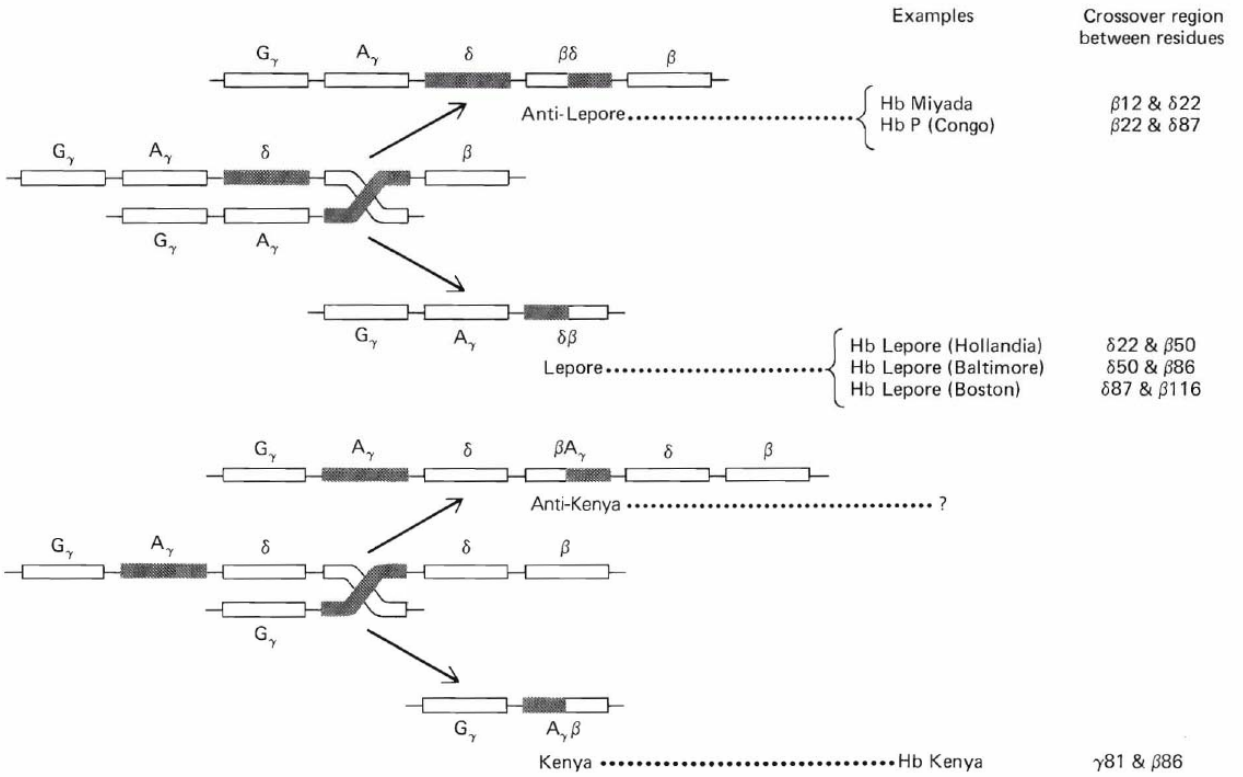


FIGURE 3.17
Genetic mechanism for production of hybrid proteins by mismatched recombination between duplicated genes. The chromosomal arrangements of the β-like globin genes are depicted, with recombination between the δ and β genes at the top and between the $^{A}\gamma$ and β genes at the bottom. These events are inferred from the amino acid sequences of the hybrid proteins produced. (From D. J. Weatherall and J. B. Clegg, *Cell* 16:467–479, 1979.)