

Homologous Proteins

Traditionally, phylogenies were based largely on anatomical or behavioral traits and biologists attempted to determine the relationships between organisms based on overall degree of similarity or by tracing the appearance of key characteristics. With the advent of molecular techniques, homologies can now be studied at the molecular level as well and these can be compared to the phylogenies established using other methods. Protein sequencing provides an excellent tool for establishing **homologies** (similarities resulting from shared ancestry). Each

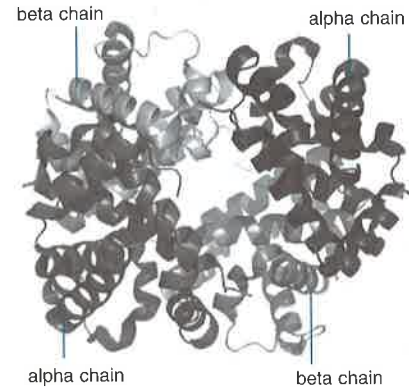
protein has a specific number of amino acids arranged in a specific order. Any differences in the sequence reflect changes in the DNA sequence. Commonly studied proteins include blood proteins, such as **hemoglobin** (below), and the respiratory protein **cytochrome c** (next page). Many of these proteins are **highly conserved**, meaning they change very little over time, presumably because mutations would be detrimental to basic function. Conservation of protein sequences is indicated by the identical amino acid residues at corresponding parts of proteins.

Amino Acid Differences in Hemoglobin

Human beta chain	0
Chimpanzee	0
Gorilla	1
Gibbon	2
Rhesus monkey	8
Squirrel monkey	9
Dog	15
Horse, cow	25
Mouse	27
Gray kangaroo	38
Chicken	45
Frog	67

When the sequence of the **beta hemoglobin chain** (right), which is 146 amino acids long, is compared between humans, five other primates, and six other vertebrates, the results support the phylogenies established using other methods. The numbers in the table (left) represent the number of amino acid differences between the beta chain of humans and those of other species. In general, the number of amino acid differences between the hemoglobins of different vertebrates is inversely proportional to genetic relatedness.

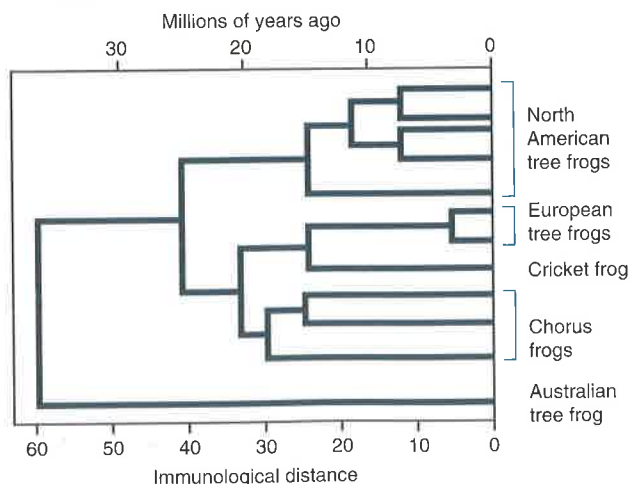
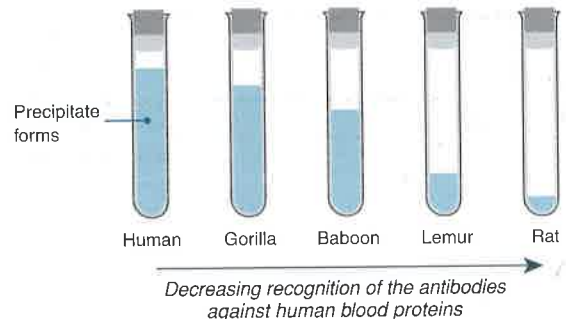
Shading indicates (from top) primates, non-primate placental mammals, marsupials, and non-mammals.



In most vertebrates, the oxygen-transporting blood protein hemoglobin is composed of four polypeptide chains, two alpha chains and two beta chains. Hemoglobin is derived from myoglobin, and ancestral species had just myoglobin for oxygen transport. When the amino acid sequences of myoglobin, the hemoglobin alpha chain, and the hemoglobin beta chain are compared, there are several amino acids that remain conserved between all three. These amino acid sequences must be essential for function because they have remained unchanged throughout evolution.

Using Immunology to Determine Phylogeny

The immune system of one species will recognise the blood proteins of another species as foreign and form antibodies against them. This property can be used to determine the extent of relatedness between species. Blood proteins, such as albumins, are used to prepare **antiserum** in rabbits. The antiserum contains antibodies against the test blood proteins (e.g. human) and will react to those proteins in any blood sample they are mixed with. The extent of the reaction indicates how similar the proteins are; the greater the reaction, the more similar the proteins. This principle is illustrated (right) for antiserum produced to human blood and its reaction with the blood of other primates and a rat.



The relationships among tree frogs have been established by immunological studies based on blood proteins such as immunoglobulins and albumins. The **immunological distance** is a measure of the number of amino acid substitutions between two groups. This, in turn, has been calibrated to provide a time scale showing when the various related groups diverged.

Cytochrome c and the Molecular Clock Theory

Evolutionary change at the molecular level occurs primarily through fixation of neutral mutations by genetic drift. The rate at which one neutral mutation replaces another depends on the mutation rate, which is fairly constant for any particular gene.

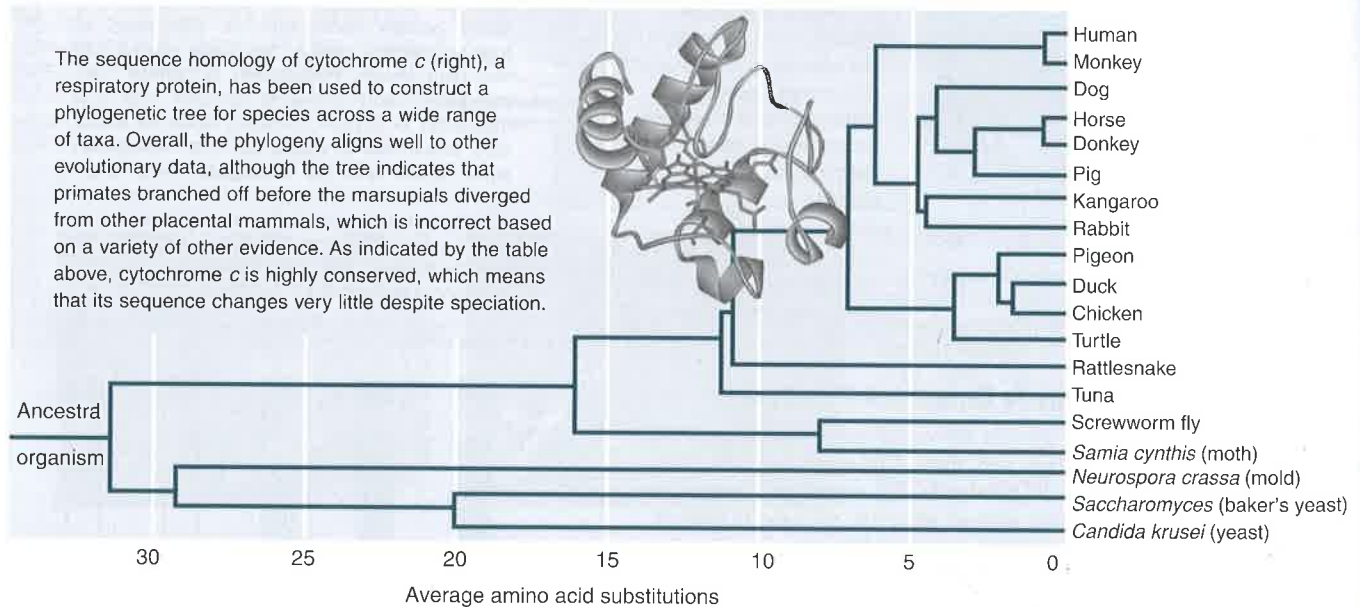
If the rate at which a protein evolves is roughly constant over time, the amount of molecular change that a protein shows can be used as a molecular clock to date evolutionary events, such as the divergence of species.

The molecular clock for each species, and each protein, may run at different rates, so scientists calibrate the molecular clock data with other evidence (morphological, molecular) to confirm phylogenetic relationships.

For example, 20 amino acid substitutions in a protein since two organisms diverged from a known common ancestor 400 mya indicates an average substitution rate of 5 substitutions per 100 my.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Human		Gly	Asp	Val	Glu	Lys	Gly	Lys	Lys	Ile	Phe	Ile	Met	Lys	Cys	Ser	Gln	Cys	His	Thr	Val	Glu	Lys
Pig												Val	Gln			Ala							
Chicken				Ile						Val		Val	Gln			Ala							
Dogfish									Val		Val	Gln				Ala							Asn
Drosophila	<<									Leu		Val	Gln	Arg		Ala							Ala
Wheat	<<		Asn	Pro	Asp	Ala		Ala				Lys	Thr	Arg		Ala						Asp	Ala
Yeast	<<		Ser	Ala	Lys			Ala	Thr	Leu		Lys	Thr	Arg		Glu	Leu						

This table shows the N-terminal 22 amino acid residues of human cytochrome c, with corresponding sequences from other organisms aligned beneath. Sequences are aligned to give the most position matches. A shaded square indicates no change. In every case, the cytochrome's heme group is attached to the Cys-14 and Cys-17. In *Drosophila*, wheat, and yeast, arrows indicate that several amino acids precede the sequence shown.



The sequence homology of cytochrome c (right), a respiratory protein, has been used to construct a phylogenetic tree for species across a wide range of taxa. Overall, the phylogeny aligns well to other evolutionary data, although the tree indicates that primates branched off before the marsupials diverged from other placental mammals, which is incorrect based on a variety of other evidence. As indicated by the table above, cytochrome c is highly conserved, which means that its sequence changes very little despite speciation.

1. Explain why chimpanzees and gorillas are considered most closely related to humans, while monkeys are less so:

2. (a) Why would a respiratory protein like cytochrome c be highly conserved?

(b) Why are highly conserved proteins good candidates for use in establishing protein homologies?

3. Discuss some of the limitations of using protein homology, specifically molecular clocks, to establish phylogeny:
