Some Early Sequence Comparisons

pituitary hormones
insulins
hemoglobins
fibrinopeptides
cytochromes C
ferredoxins
ribonucleases
Fitch & Margoliash (1967)
Phylogenetic Trees

Align the protein sequences properly.

Count the differences between every pair.

Make a difference (= distance) matrix.

Find the shortest distance between pairs and designate the nearest two taxa as a subgroup.

Make a new matrix with one less entry, (the subgroup replaces the two nearest taxa).

Find the next most closely related pair, etc., continuing the process until the branching order is complete.
**MODEL I**

One Protein: Various Species

```
A → B → C → D
```

Common Ancestor

```
m = n + o = n + p + g = n + p + r  (mutation occurrence)
```

**MODEL II**

One Organism: Several Proteins

```
A′ → B′ → C′
```

Common Ancestor

```
m = n + o = n + p  (mutation occurrence)
```

```
a, b, c, d = f (selection)  (mutation survival)
```
gene duplication
(400 million years ago)

early hemoglobin gene

α chain gene

β chain gene

Frog

Chicken α

Mouse α

Frog β

Chicken β

Mouse β

orthologues

paralogues

homologues
Some Gene Duplications Uncovered Without Computers

- vasopressin/oxytocin (1955)
- hemoglobin chains (1957)
- vitamin K clotting factors* (1961)
- glucagon/secretin (1963)
- trypsin/chymotrypsin (1964)
- immunoglobulin chains (1966)

*based on unique adsorptive properties.
Some Changes in Function After Gene Duplication

Lysozyme to lactalbumin

Serine protease to haptoglobin

Transglutaminase to rbc structural protein 4.2

Ribonuclease to angiogenin

Ceruloplasmin to clotting factors

Serpin to hormone transporter

Serpin to hormone (angiotensinogen)

Lipocalin to prostaglandin D synthase
EVOLUTIONARY DISTANCE VS SEQUENCE DISSIMILARITY

TWILIGHT - ZONE
A Simple sequence comparison.
B Weighted scales, statistics.
C Multiple alignments, profile analysis.
D 3-D structure comparison.
### VERTEBRATE HEMOGLOBIN ALPHA CHAINS

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HUMAN</td>
<td>VLSPADKTNVKAAWGKVGAHAGEYGAEALE..</td>
<td></td>
</tr>
<tr>
<td>RABBIT</td>
<td>VLSPADKTNIKTAWEKIGSGHGGEGYGAEEAVE..</td>
<td></td>
</tr>
<tr>
<td>KANGAROO</td>
<td>VLSAADKGVKAIWGBKGVGHAGEYAAEGLE..</td>
<td></td>
</tr>
<tr>
<td>CHICKEN</td>
<td>VLSNAADKNNVKGIFTKIAGHAEETYGAETLE..</td>
<td></td>
</tr>
<tr>
<td>FROG</td>
<td>SLSASEKAAVLSTIVGBKIGSQQGSALGSEALT..</td>
<td></td>
</tr>
<tr>
<td>LUNGFISH</td>
<td>MRFSQDDDEVLIKEAWGLLHPQINAGGEALA..</td>
<td></td>
</tr>
</tbody>
</table>

### HUMAN HEMOGLOBIN CHAINS

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPHA</td>
<td>VLSPADKTNVKAAWGKVGAHAGEYGAEALE..</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BETA</td>
<td>VHLTPEEKSAVTALWGBKVNVDENVGGEALG..</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAMMA</td>
<td>GHFTEEDKATITSLWGBKVNVEDAGGETLG..</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DELTA</td>
<td>VHLTPEEKTAVNALWGBKVNVDAVGGEALG..</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ZETA</td>
<td>SLTKTERTIIVSMWAKISTQADTIGTETL..</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MYOGL</td>
<td>GLSDGEWQLVLNVWGKVVEDIPGHGQEVL..</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Percent Identity of human hemoglobin chains

<table>
<thead>
<tr>
<th></th>
<th>alpha</th>
<th>beta</th>
<th>gamma</th>
<th>delta</th>
<th>zeta</th>
</tr>
</thead>
<tbody>
<tr>
<td>myoglobin</td>
<td>27</td>
<td>25</td>
<td>24</td>
<td>26</td>
<td>28</td>
</tr>
<tr>
<td>alpha</td>
<td>45</td>
<td>42</td>
<td>45</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>beta</td>
<td></td>
<td>73</td>
<td>93</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>gamma</td>
<td></td>
<td></td>
<td>72</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>delta</td>
<td></td>
<td></td>
<td></td>
<td>40</td>
<td></td>
</tr>
</tbody>
</table>
The jawless vertebrates (lampreys and hagfish) have single-chain hemoglobins.
Using 3-D Structures to Determine Phylogenies
Some Homologies Found by 3-D Searching

Actin-Heat Shock 70

Globins-phycocyanins

Lysozyme-Chitinase

Ferredoxin-cold shock protein

PDGF-TGFβ-NGF

Urease-adenosine deaminase
Prothrombin

Protein C, factor VII. Factor IX, Factor X

Factor XII

Plasminogen

Prothrombin

Protein C, factor VII. Factor IX, Factor X

Factor XII

Plasminogen

Prothrombin

Protein C, factor VII. Factor IX, Factor X

Factor XII

Plasminogen

Prothrombin

Protein C, factor VII. Factor IX, Factor X

Factor XII

Plasminogen

Prothrombin

Protein C, factor VII. Factor IX, Factor X

Factor XII

Plasminogen

Prothrombin

Protein C, factor VII. Factor IX, Factor X

Factor XII

Plasminogen
INTERMISSION