In Biology

How can Similarity occur?

3 Possibilities

For questions, comments contact Joanne K. Kelleher, jkk@mit.edu

In Biology

How can Similarity occur?

3 Possibilities

1. **Chance**
2. **Evolutionary Homology or…**

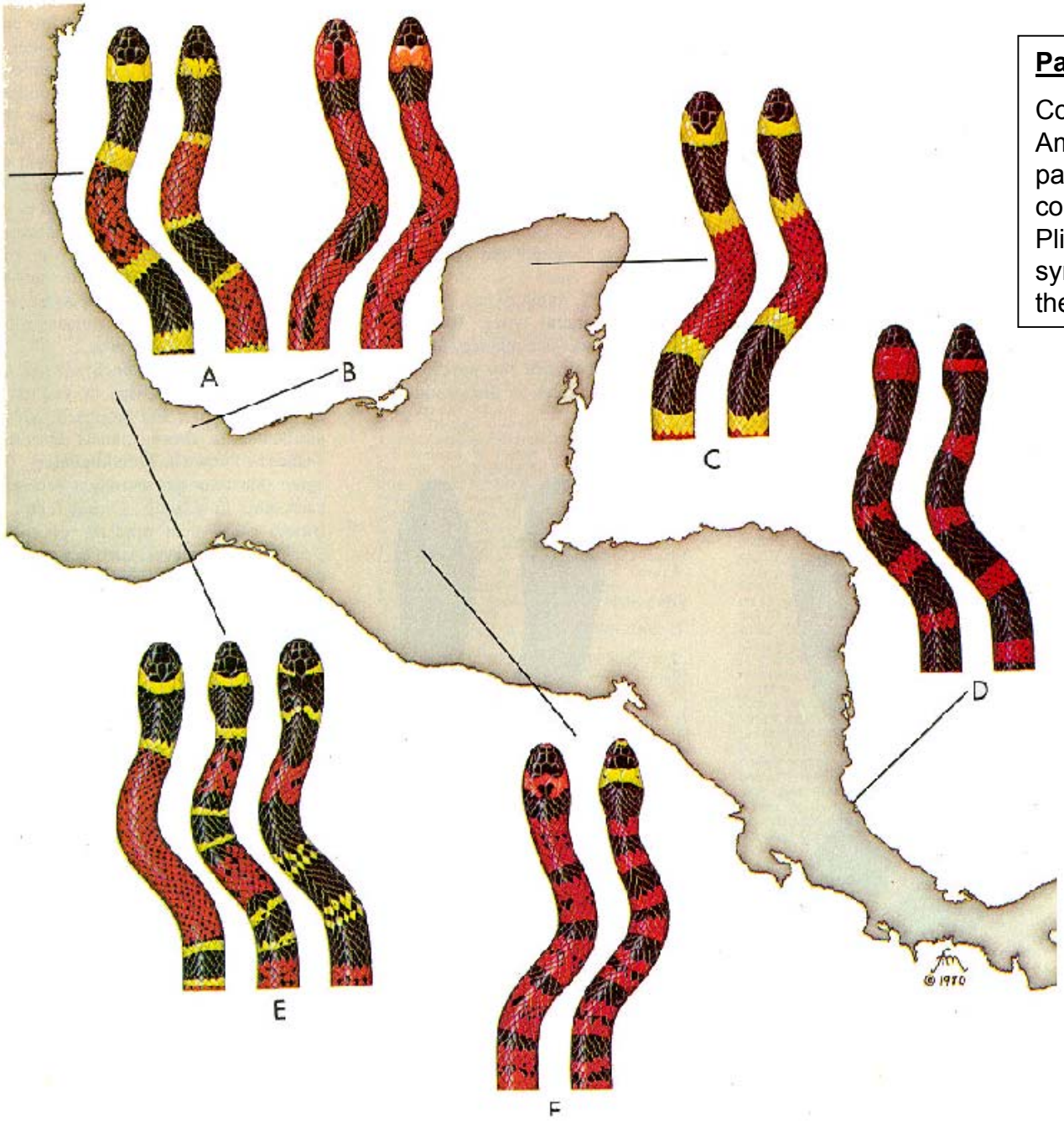
Similarity occurs by either …

1. **chance,**
2. **evolutionary homology or**
3. **convergence.**

Convergence depends on adaptive replacements that are positively selected.

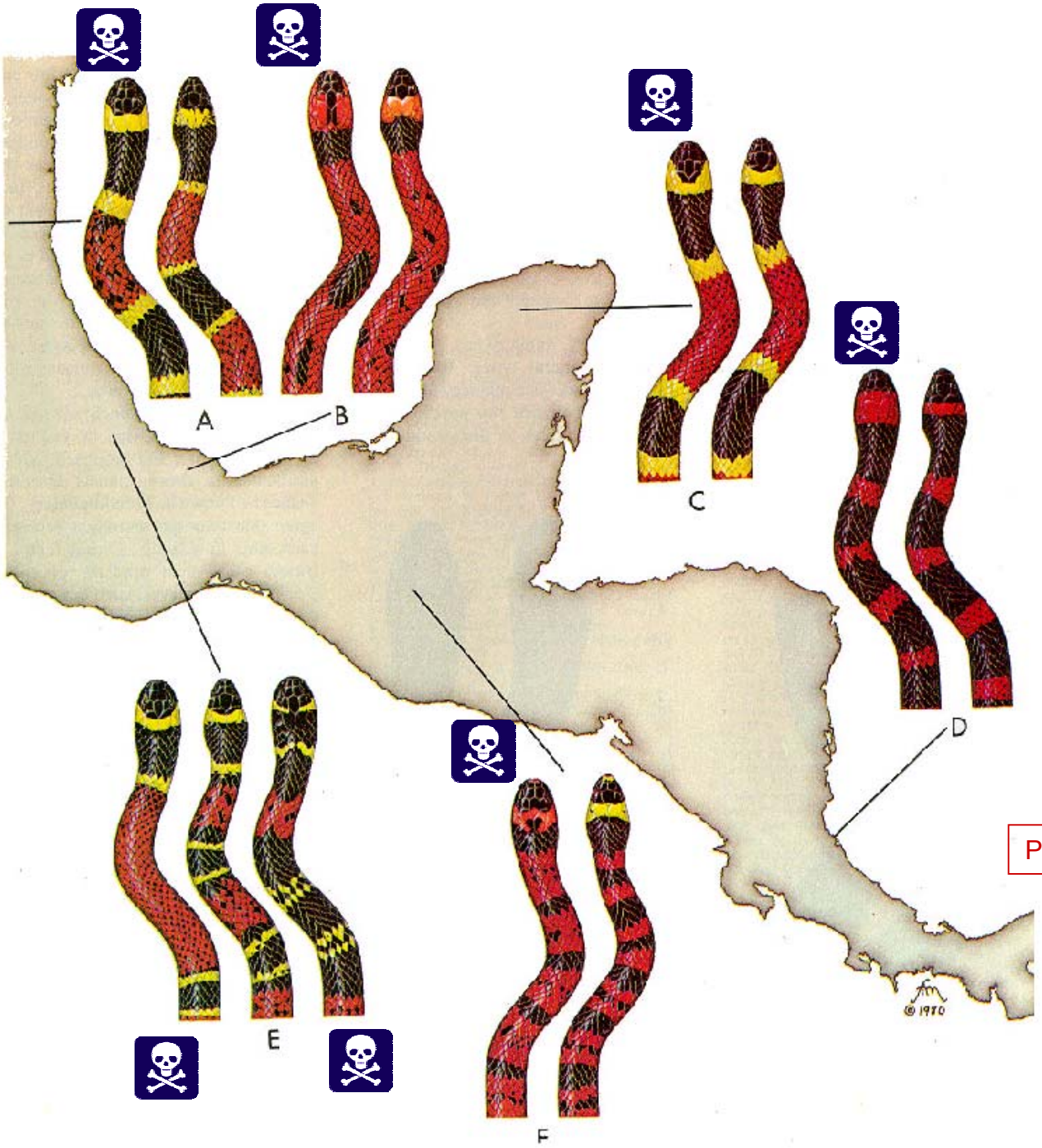
The starting point…..

R. F. Doolittle. 1994. Convergent evolution: the need to be explicit. Trends Biochem.Sci. 19(1):15-18.



Pattern Discovery by Hawks.

Coral Snakes of Central America. Geographic color pattern variation in rear-fanged colubrid snakes of the genus Pliocercus in relation to sympatric front-fanged snakes of the genus Micrurus.

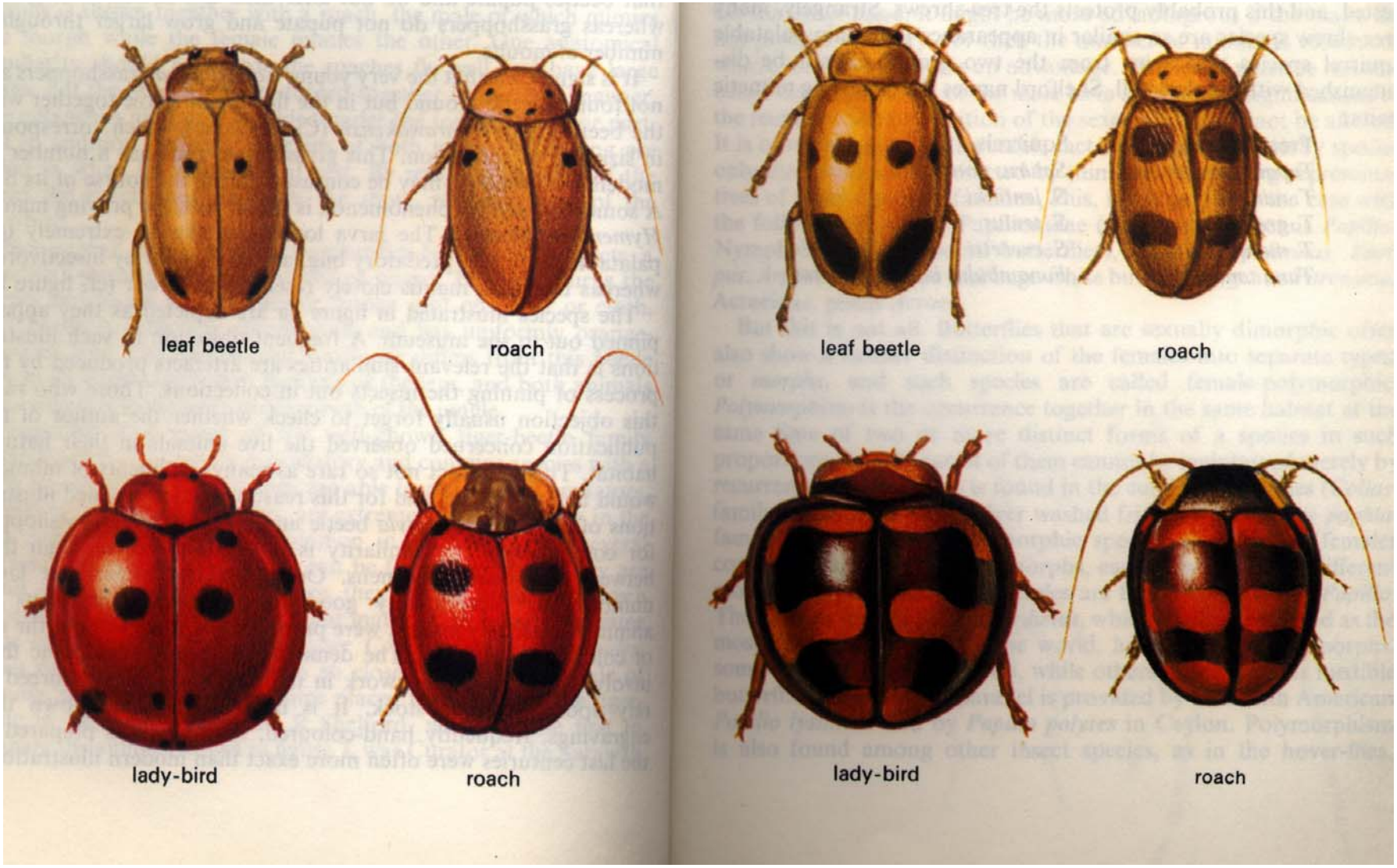


The right snake in each pair (and central snake in diagram E) are rear-fanged snakes of the genus Pliocerus. All are nonpoisonous. The left snake in each pair (and the right and left snakes in diagram E) are front-fanged poisonous snakes of the genus Micrurus.

These snakes are examples of Batesian mimicry

Pattern Discovery by Hawks ?

Leaf Beetles and roaches from the Philippines. Beetles (unpalatable) on right for each pair serve as models for the roach mimics.



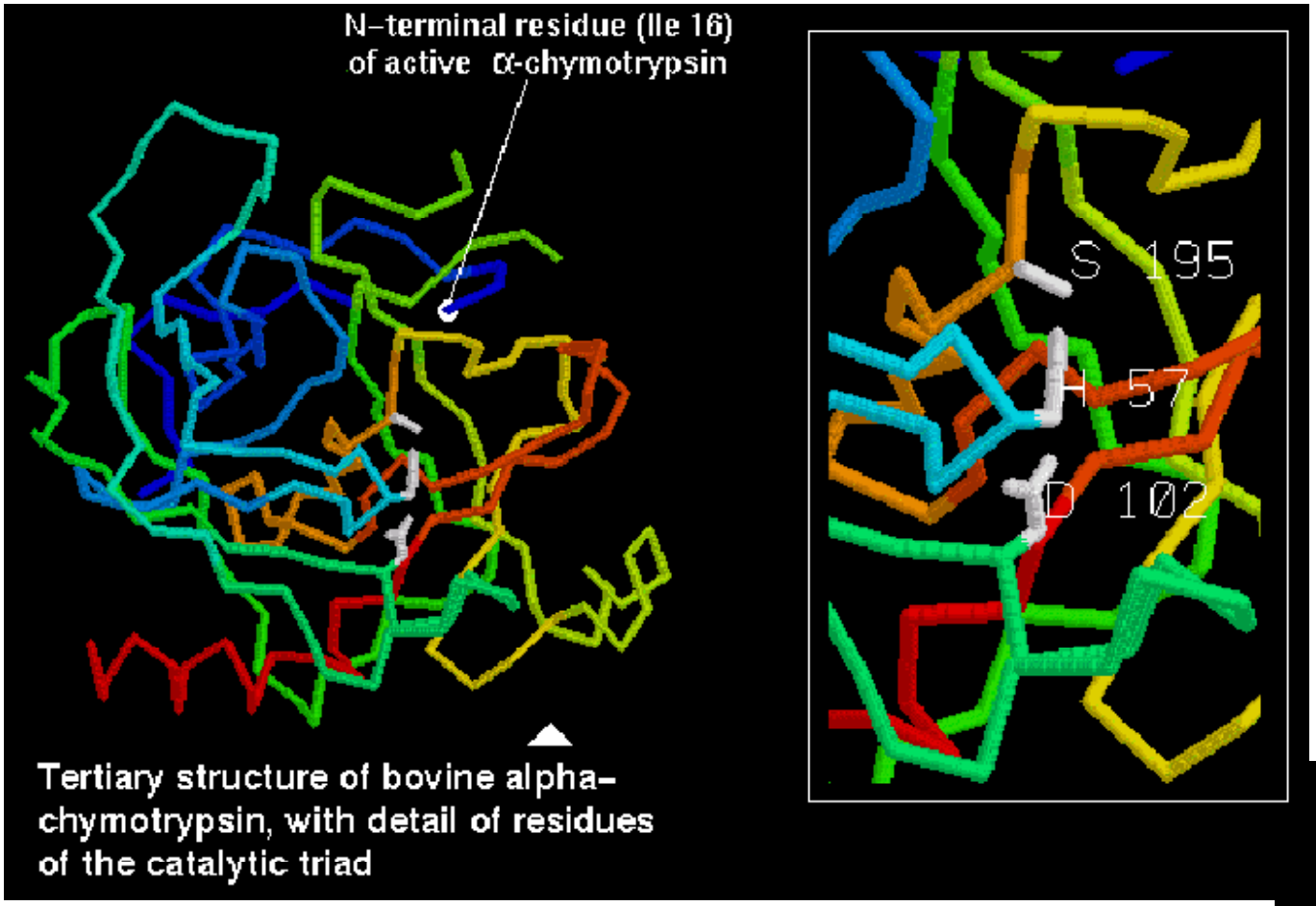
Similarity occurs by either …

1. **chance,**
2. **evolutionary homology or**
3. **convergence.**

The starting point…..

R. F. Doolittle. 1994. Convergent evolution: the need to be explicit. Trends Biochem.Sci. 19(1):15-18.

Careful study of many proteins has indicated that true convergent evolution of long sequences of amino acids is very rare (may not have been documented to date). Very similar 3D protein structures often result from unrelated amino acid sequences.



Tertiary structure of bovine alpha–chymotrypsin, with detail of residues of the catalytic triad

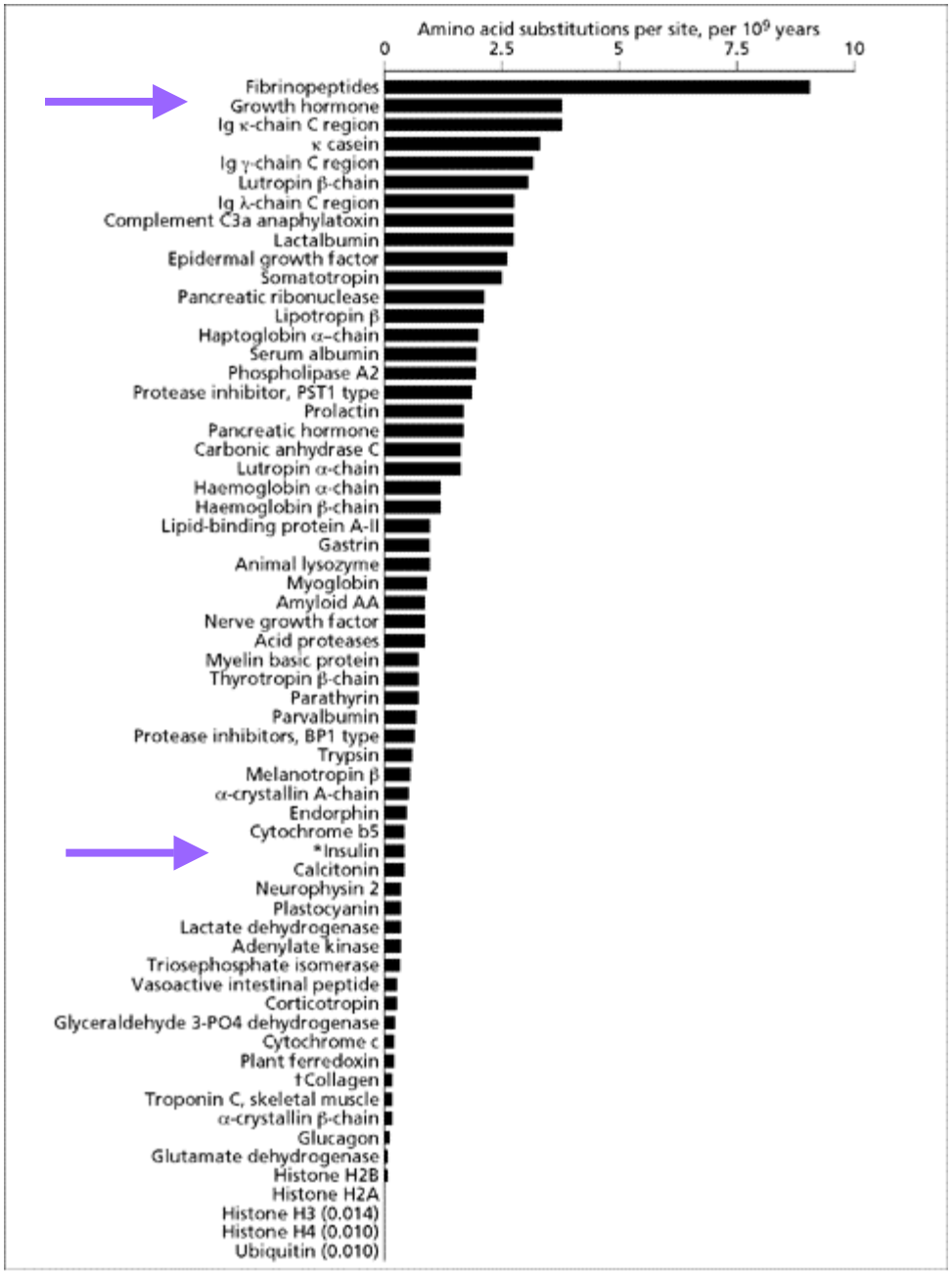
Subtilisin (*Bacillus amyloliquefaciens*), indicating catalytic triad

Some pointers from biology when considering

Protein sequences and gene sequences of proteins?

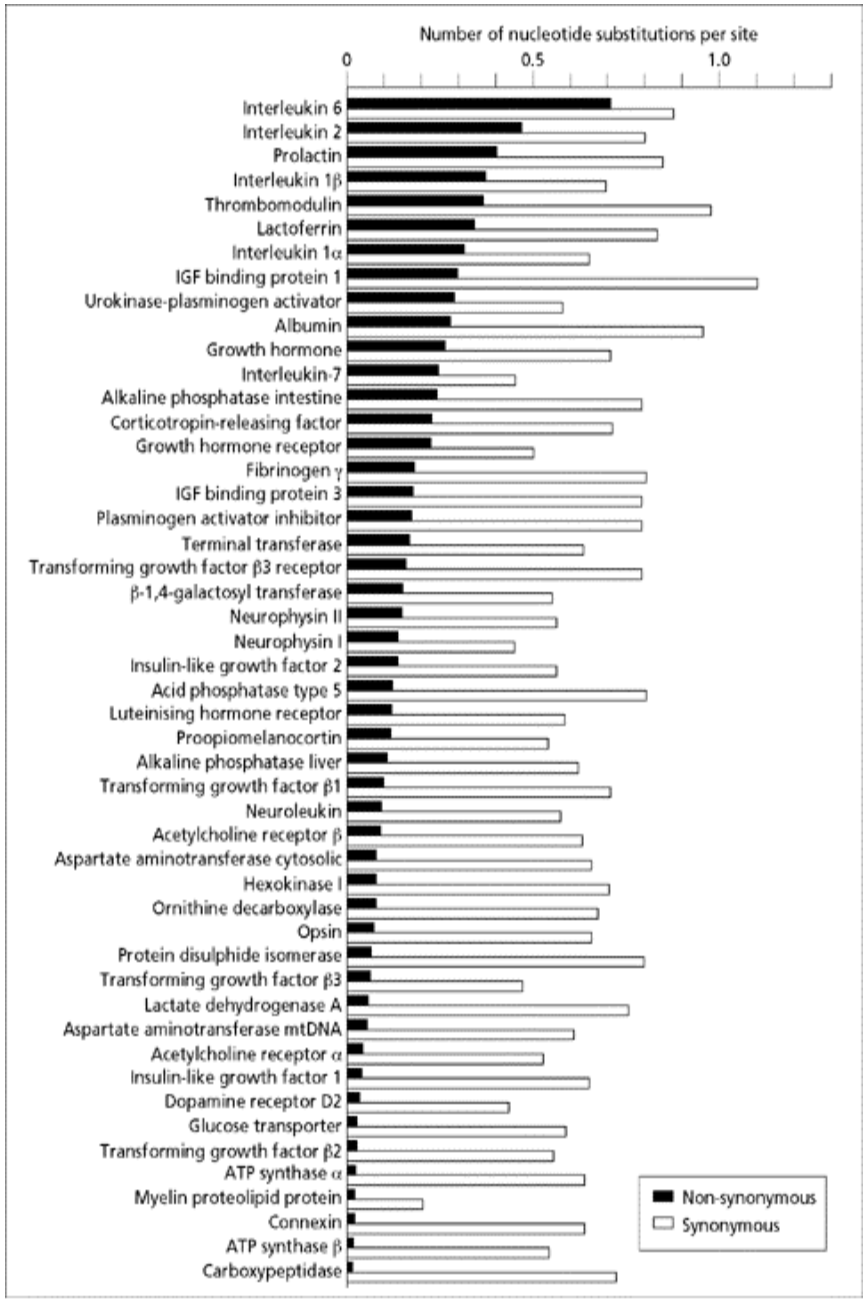
1. Rates of change in amino acid sequence over time

2. Changes in DNA sequence may be Synonymous or Non-synonymous.



Some pointers from Biology about protein sequences

1. Rates of change in protein sequences over evolutionary time are not similar for all proteins



Page, Holmes
Molecular Evolution

Blackwell Science

At the nucleic acid level one may consider both Synonymous and Non-synonymous mutations.

Synonymous mutation are base changes that do not result in aa sequence changes

AAA -> Lysine

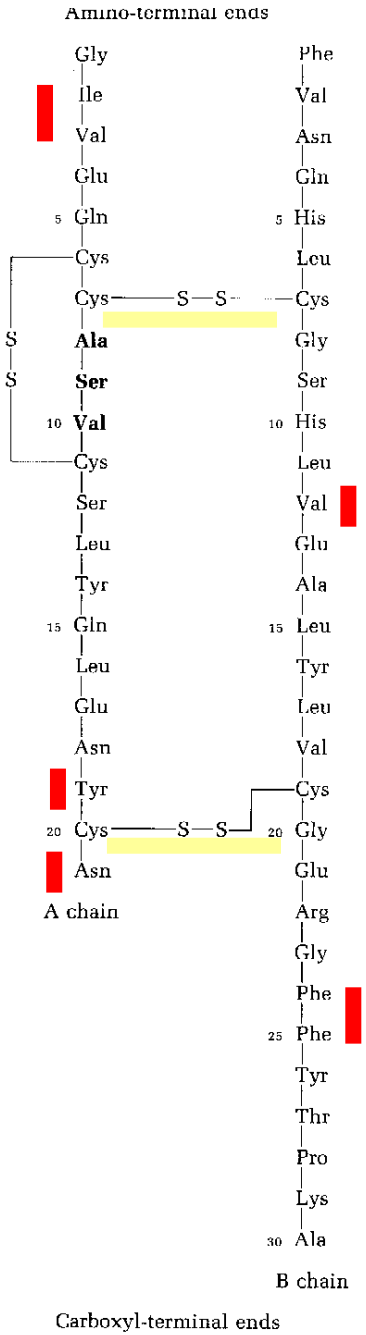
↓

AAG -> Lysine

Non-synonymous mutations produce aa sequence changes

If all nucleotide substitutions are equally likely this data suggests that many Non-synonymous changes are deleterious mutations are selected against….

Data for combined primates rodents and artiodactyls



What parts of the amino acid sequence are important for biological activity?

Insulin - Insulin receptor Binding in Mammals

A current consensus is that A2 isoleucine, A3 valine, B12 valine, B24 and B25 phenylalanine, A19 tyrosine, A21 asparagine, and the partially buried residues A16 and B15 leucine are the major determinants of the receptor binding site, with A8 threonine, B9 serine, B10 histidine, B13 glutamate, and B16 tyrosine making minor contributions.

A chain

```
GIVEQCCTSICSLYQLENYCN     Human
ARIVQQCTSGICSLYQENYCN     Sponge
GIVEQCCHKRCSIYDLENYCN     Hagfish
GLVEECCYNVCDYSQLESYCN     Amphioxus
```

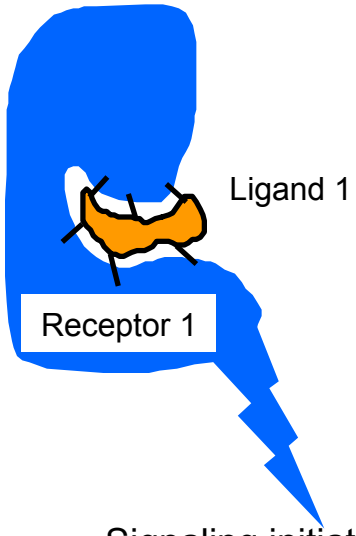
B chain

```
FVNQHLCGSHLVEALYLVCGERGFFYTPKT     Human
FVNQHLCGSHLVEALYILVCGERGFFYTPMS    Sponge
RTTGHLCGKDLVNALYIACGVRGFFYDPTKM    Hagfish
TQAEYLCGSTLADVLSFVCGNRGYNSQP       Amphioxus
```

Protein - protein interactions differ from protein small common ligand interactions

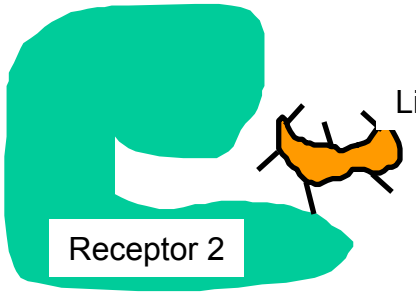
Here each of two proteins may evolve.

Case 1 Ligand (small protein) binds to a receptor (larger protein) very specifically leading to signaling, downstream actions

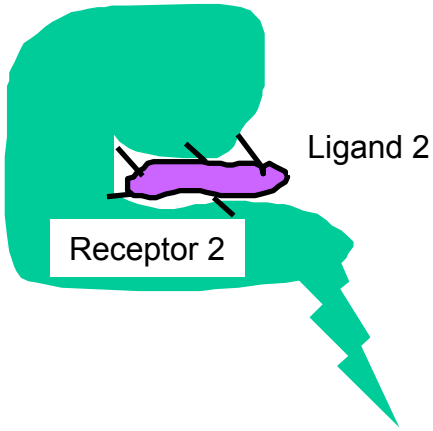


Signaling initiated.

Case 2 Receptor Has mutated and now original ligand cannot bind the receptor. No signaling, downstream actions



Case 3 Mutated receptor binds a different ligand leading to signaling, downstream actions



Signaling initiated.

Just one of many possible scenarios.

Biological systems have a remarkable potential to respond to natural selection.

If long sequences of amino acids have not converged perhaps there is no selective pressure to do so.

What factors influence the similarity between shorter sequences?

Are short amino acid sequences candidates for convergent evolution ?

Perhaps binding to simple universal ligands is an example.

ATP, NADH, Fe, O_2, etc.,

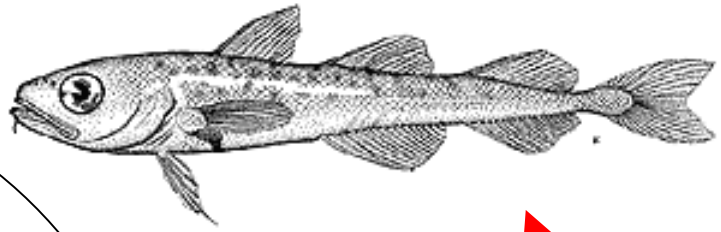
All organisms need to design proteins to bind these compounds. Is this an opportunity for convergent evolution of small peptide sequences ?

J. M. Logsdon, Jr. and W. F. Doolittle. 1997. Origin of antifreeze protein genes: a cool tale in molecular evolution. *Proc.Natl.Acad.Sci.U.S.A* 94(8):3485-3487

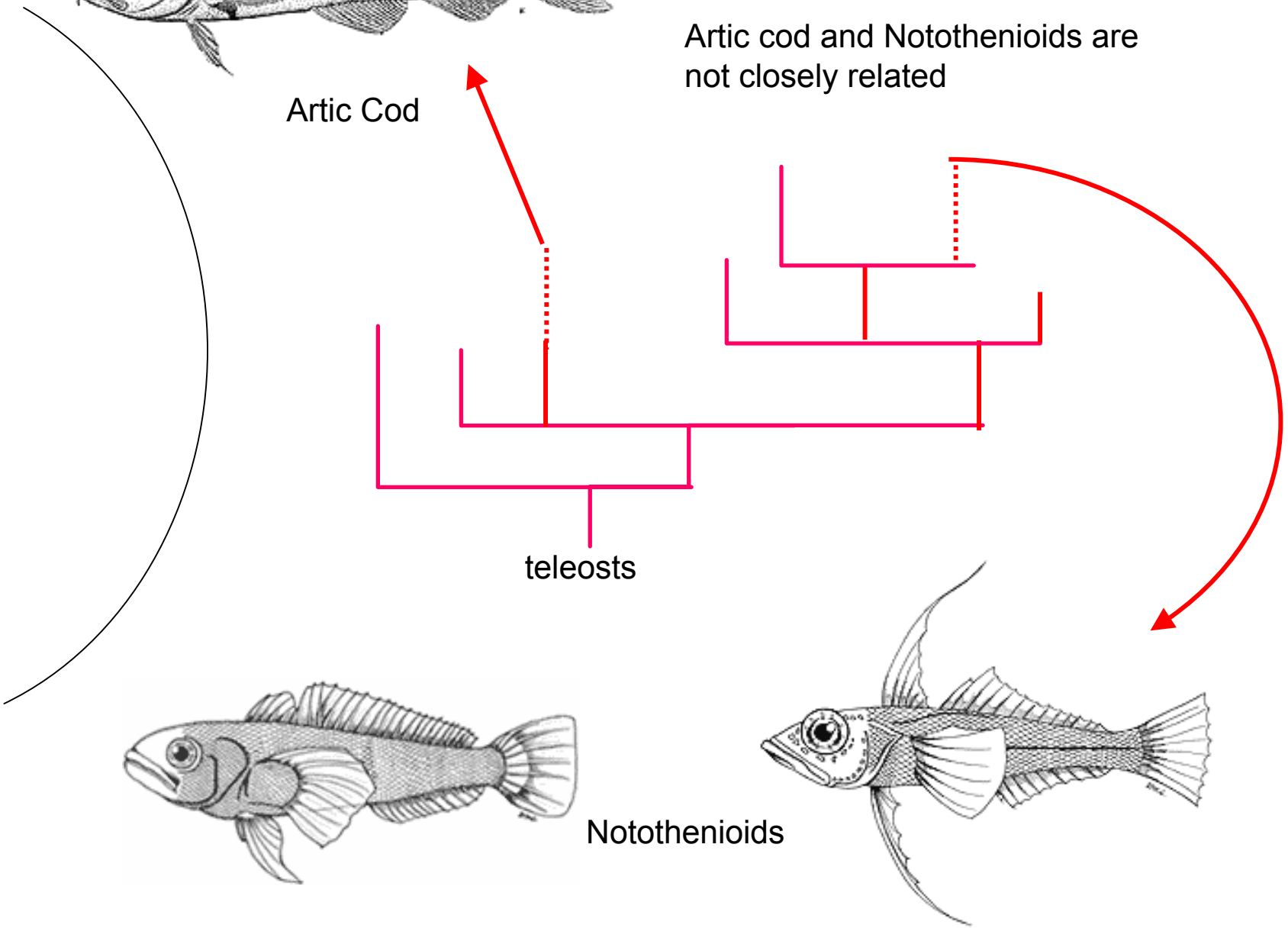
L. Chen, A. L. DeVries, and *C. H. Cheng*. 1997. Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. *Proc.Natl.Acad.Sci.U.S.A* 94(8):3817-3822.

How to survive in subfreezing water
Antifreeze: proteins that bind ice as ligand

Artic cod and Notothenioids are not closely related

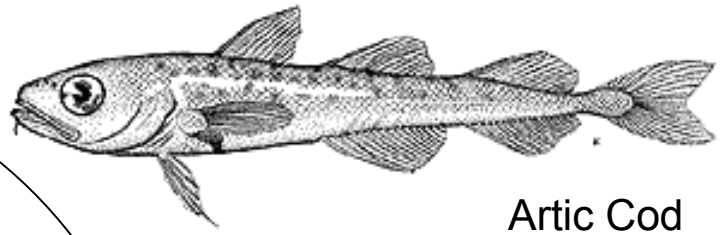


Artic Cod



Notothenioids

Artic



Artic Cod

TAATAPTAATAATAATAARAAT

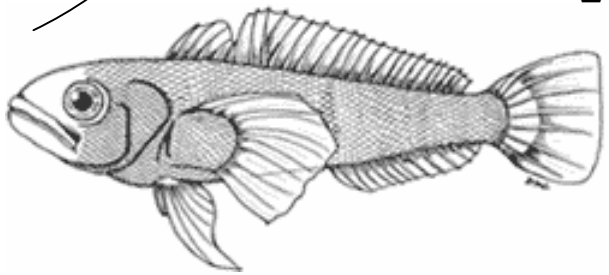
The cod and the notothenioids antifreeze molecules illustrate convergent evolution of short amino acid sequences targeted to a common ligand, ice.

The antifreeze is the same three amino acids repeated over and over with occasional proline. (Threonine has disaccharide ⬟–⬟ attached posttranslationally) R replaces T occasionally in cod antifreeze.

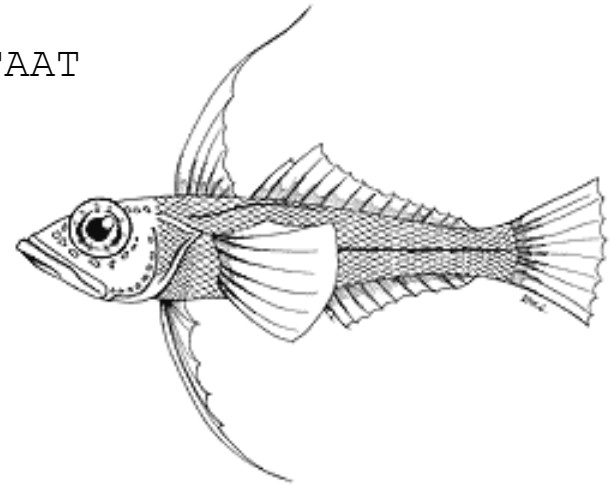
The proteins evolved independently. The notothenioid protein arose about 7 million to 15 million years ago, when Antarctic oceans were chilling to freezing, while the cod version probably evolved about 3 million years ago, during the glaciation of the Arctic seas.

Antarctic

TAATAATAATAATAAPAATAAT



Notothenioids



Does mimicry occur in protein structure ?

Two topics to considered here

1. Autoimmune response

2. Viral immune evasions strategies

Consider the following diseases:

Multiple Sclerosis

Diabetes Mellitus, Type I

Rheumatoid Arthritis

HIV AIDS

Graves Disease (hyperthyroidism)

Rheumatic fever???

What do they have in common ?

Molecular mimicry by infections agents has been suggested as a common mechanism in many autoimmune diseases.

How does this occur?

Infectious agents such as viruses produce proteins that are immunologically similar to host proteins. Antibodies raised by host react with both virus protein and host protein.

It is not required that these proteins are homologous by sequence.

…..but perhaps there is a similarity that may be revealed by pattern discovery.

Albert LJ and Inman RD. Molecular mimicry and autoimmunity. *N Engl J Med* 341: 2068-2074, 1999.

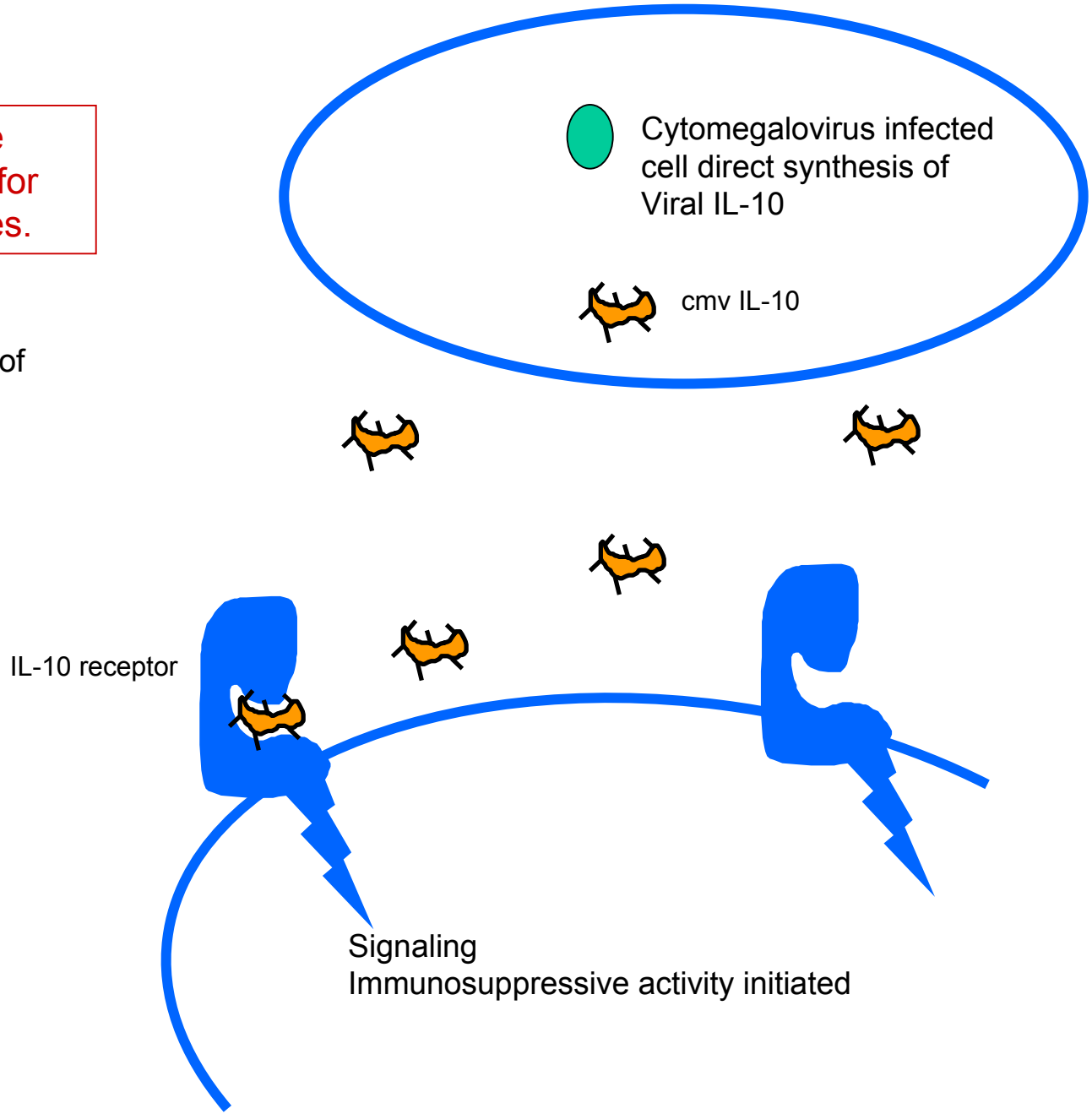
2. Viral immune evasions strategies

Alcami A. Viral mimicry of cytokines, chemokines and their receptors. *Nat Rev Immunol* 3: 36-50, 2003

Viruses have evolved elegant mechanisms to evade detection and destruction by the host immune system. One of the evasion strategies that have been adopted by large DNA viruses is to encode homologues of cytokines, chemokines and their receptors--molecules that have a crucial role in control of the immune response. **Viruses have captured host genes or evolved genes to target specific immune pathways**

Human genes are stolen by viruses for their own purposes.

Cmv IL-10 is a homolog of mammalian IL-10



Kotenko, S. V., S. Saccani, L. S. Izotova, O. V. Mirochnitchenko, and S. Pestka. 2000. **Human cytomegalovirus harbors its own unique IL-10 homolog (cmvIL-10).** Proc. Natl. Acad. Sci. U. S. A 97: 1695-1700.

We identified a viral IL-10 homolog encoded by an ORF (UL111a) within the human cytomegalovirus (CMV) genome, which we designated cmvIL-10. cmvIL-10 can bind to the human IL-10 receptor and can compete with human IL-10 for binding sites, despite the fact that these two proteins are only 27% identical. cmvIL-10 requires both subunits of the IL-10 receptor complex to induce signal transduction events and biological activities. The structure of the cmvIL-10 gene is unique by itself. The gene retained two of four introns of the IL-10 gene, but the length of the introns was reduced. We demonstrated that cmvIL-10 is expressed in CMV-infected cells. Thus, expression of cmvIL-10 extends the range of counter measures developed by CMV to circumvent detection and destruction by the host immune system

There are many viral IL-10 proteins.

There are additional immuno-evasive strategies where viruses produce proteins similar to human proteins

What is their similarity in amino acid patterns ?