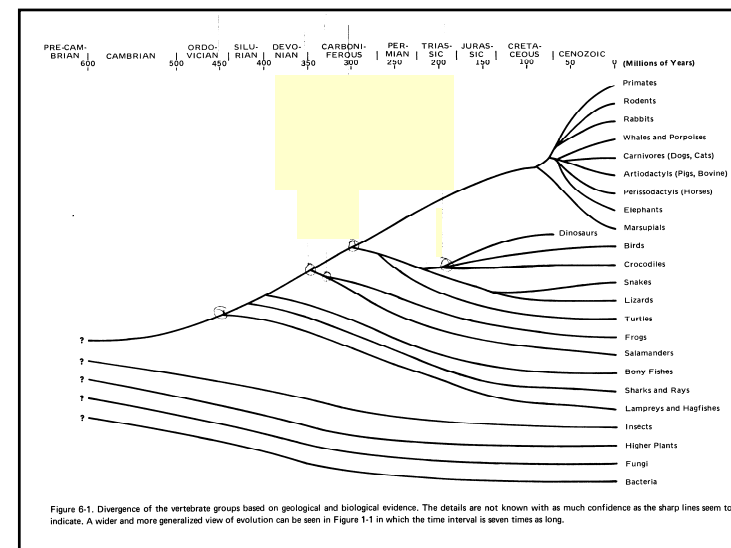
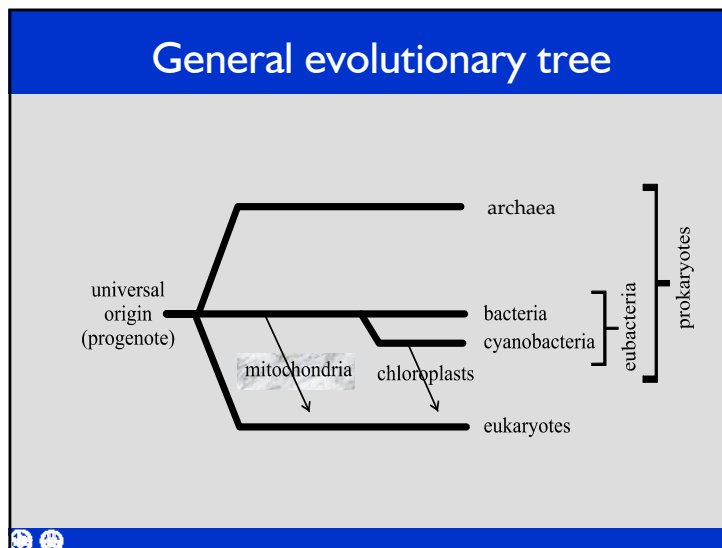
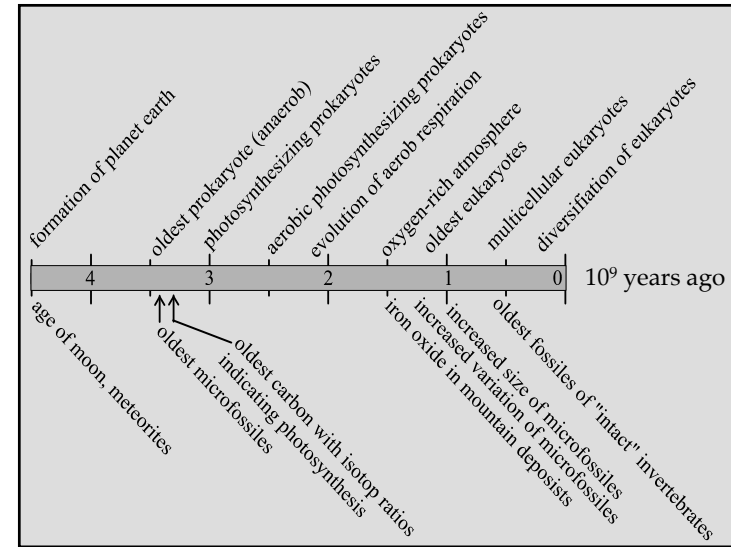


**Evolution**

Bengt Persson



### Natural selection

★ "Survival of the fittest", Darwin

### Nucleotide substitutions

- ★ Frequency of mutations
- ★ Fixation in population
- ★ Variation dependent on:
  - gene / not gene
  - site in gene
  - which gene

### Rates of evolution in different proteins

Constant speed of change for a certain gene in all species (Molecular clock)

Protein	Rate (changes in 10 <sup>9</sup> years)
Fibrinopeptides	8.3
Lysozyme	2.0
α-Globin	1.2
Insulin	0.44
Cytochrome <i>c</i>	0.3
Histone H4	0.01

*from Ridley: Evolution*

### Constant evolutionary speed

The rate of evolution of α-globin.

*from Ridley: Evolution*

## Increase of genome size

1. Genome duplications  
tetraploid organisms
2. Regional duplications
  - a) internally in genes  
Examples: collagen, immunoglobulin
  - b) gene duplications  
Examples: rRNA genes, lactalbumin & lysozyme, gene families (globins, immunoglobulins), pseudogenes

## Genome sizes (from Dayhoff, 1972)

	Average Total Genome (10 <sup>9</sup> gm genome)	Highly Repetitive of Nonrepetitive 10 <sup>6</sup> genes	Number of Genes
Mammals	6.5	3.2	102
Man	3.4	1.2	52
Plant	6.0	2.4	9
Algae	6.0	2.4	2
Liver	3.0	1.0	4
Bacteria	3.0	1.0	0
Archaea	12.0	6.0	24
Plant	6.0	2.4	9
Others*	40-100	21-80	6
Birds	1.8	0.9	20
Fish	3.2	2.0	11
Clay and Goldfish	2.2	1.0	11
Beetles	2.0	1.1	1
Insects	6.0	2.1	4
Collembola	0.8	0.7	1
Long/Val. 2 species	100-200	48.5-111.7	2
Others	6.0	2.9	2
Algae	2.0-5.1	1.2-3.5	2
Protozoans	1.1	0.6	1
Amoebae	0.3	0.2	2
Eukaryotes	1.1	0.6	10
Archaea	3.0	1.6	2
Insects - Crustal	11.1	6.7	2
Others - Fungi etc	0.3	0.1	1
Molluscs	0.0	0.4	1
Reptiles	1.2	0.6	2
Others	2.0	1.4	1
Cnidarians	0.7	0.3	1
Slugs	0.12	0.06	1
Protists	14-1610	0	1
Others*	0.00	0	1
Amoebae	0.00	0	1
Paramecia	2.4	1.0	10
Algae	0.3-300	0	10
Fungi (Ascomycetes)	0.004	0.001	14
Bacteria	0.012	0.013	54

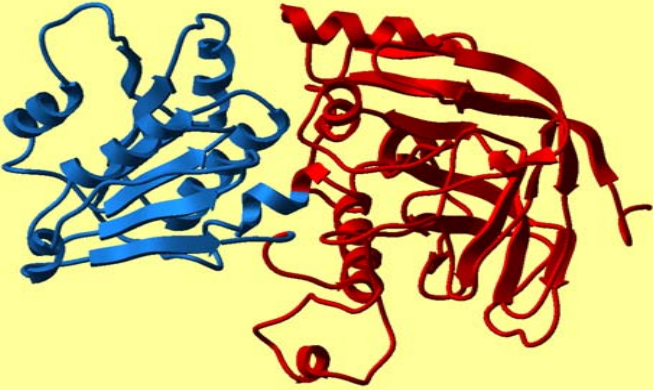
	Average Total Genome (10 <sup>9</sup> gm genome)	Highly Repetitive of Nonrepetitive 10 <sup>6</sup> genes	Number of Genes
Microplasma*	0.5-20x10 <sup>9</sup>	0.5-10x10 <sup>6</sup>	4
Plant Virus	0.2-3x10 <sup>6</sup>	0.2-3x10 <sup>6</sup>	6
Animal Virus*	0.2-3x10 <sup>6</sup>	0.2-3x10 <sup>6</sup>	20
Phage*	0.2-3x10 <sup>6</sup>	0.2-3x10 <sup>6</sup>	12
Chloroplast DNA*	100-200x10 <sup>6</sup>	0.1-1.0x10 <sup>6</sup>	1
Algae	0.1-10 <sup>9</sup>	0.1-10 <sup>6</sup>	1
Slime mold	0.1-10 <sup>9</sup>	0.1-10 <sup>6</sup>	1
Flowering plants	10-10 <sup>9</sup>	10 <sup>6</sup>	2

## Effects of gene duplications

- ★ Same function in a new context  
– e.g. lysozyme
- ★ Extended physiological repertoire  
– e.g. globin





## Evolutionary tree, hemoglobin

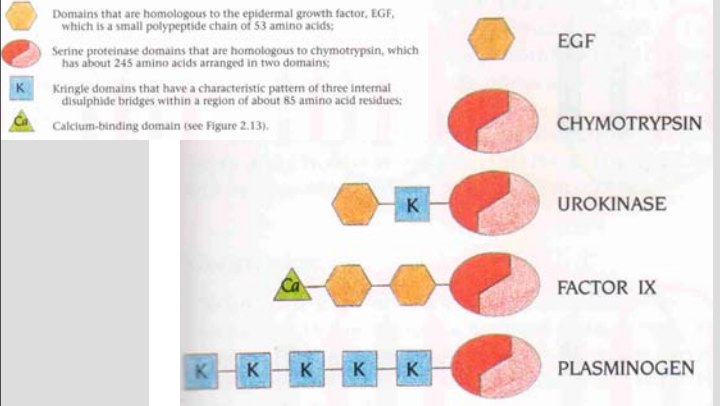
### Domains



Linköping University & Karolinska Institutet 13

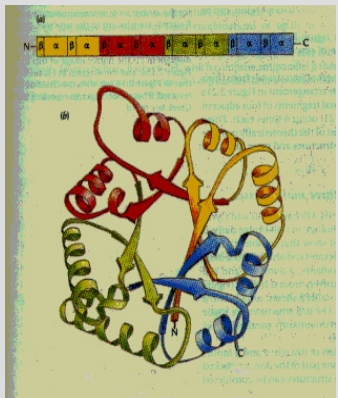
### Modular architecture I

-  Domains that are homologous to the epidermal growth factor, EGF, which is a small polypeptide chain of 53 amino acids:
-  Serine protease domains that are homologous to chymotrypsin, which has about 245 amino acids arranged in two domains:
-  Kringle domains that have a characteristic pattern of three internal disulphide bridges within a region of about 85 amino acid residues:
-  Calcium-binding domain (see Figure 2.13).



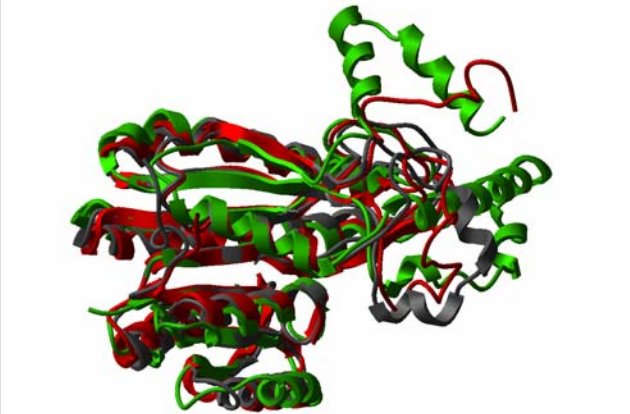
Linköping University & Karolinska Institutet 14

### Modular architecture 2

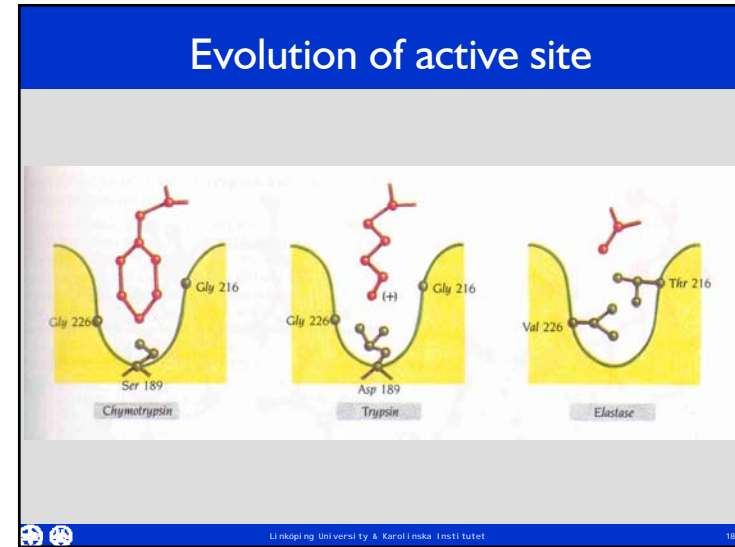
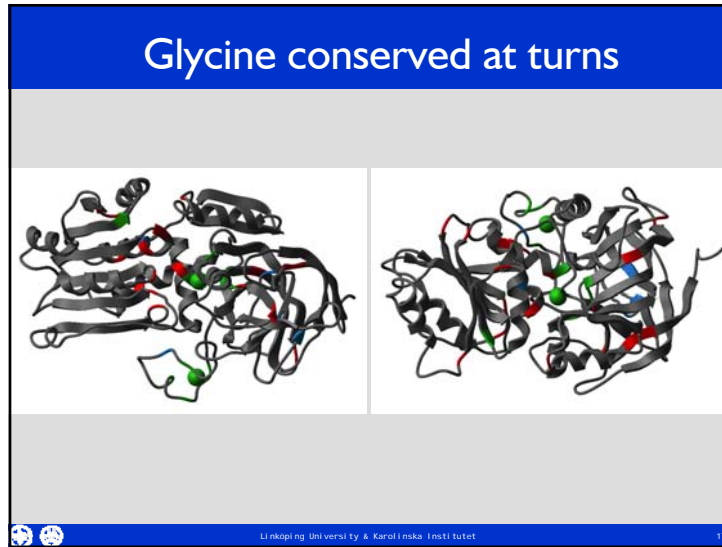


Linköping University & Karolinska Institutet 15

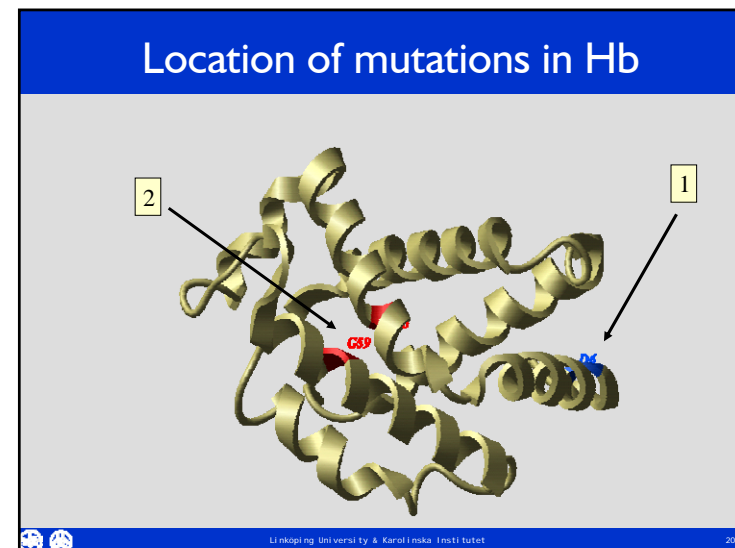
### Conserved folds

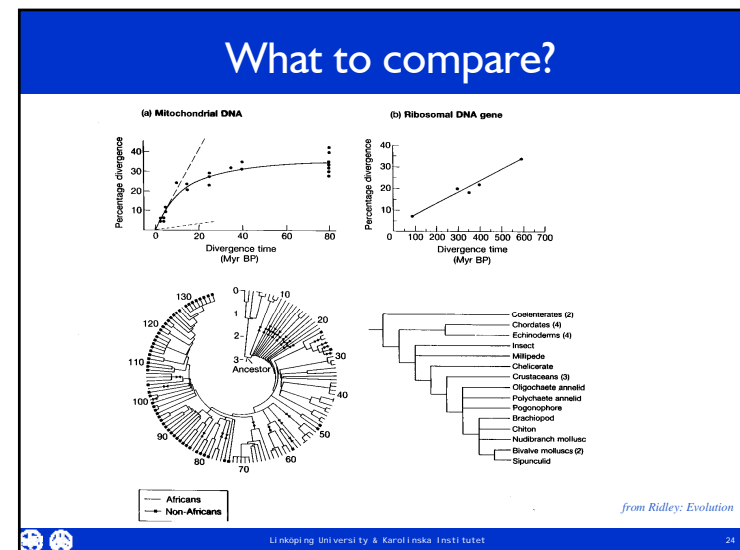
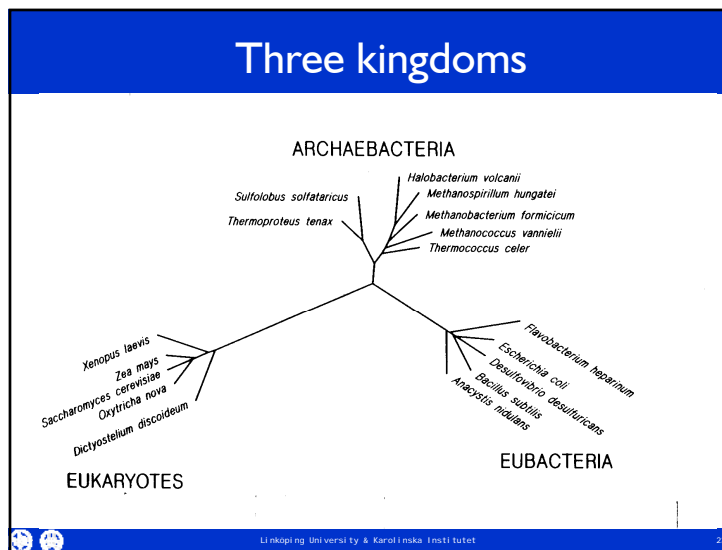
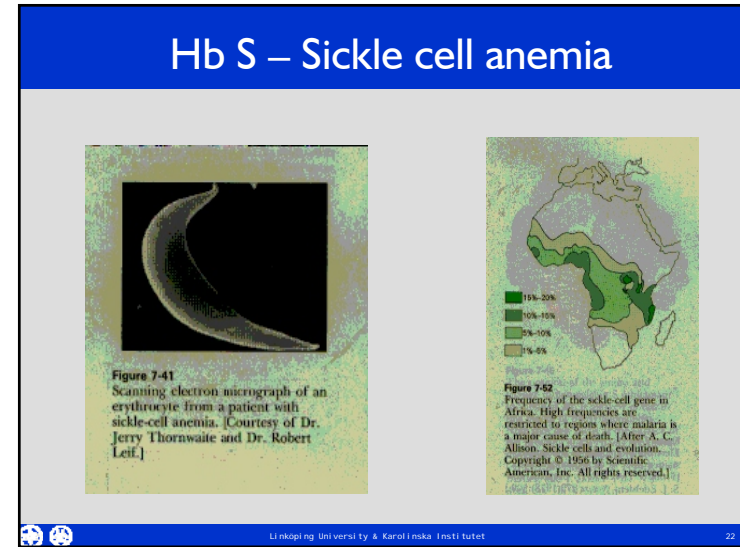
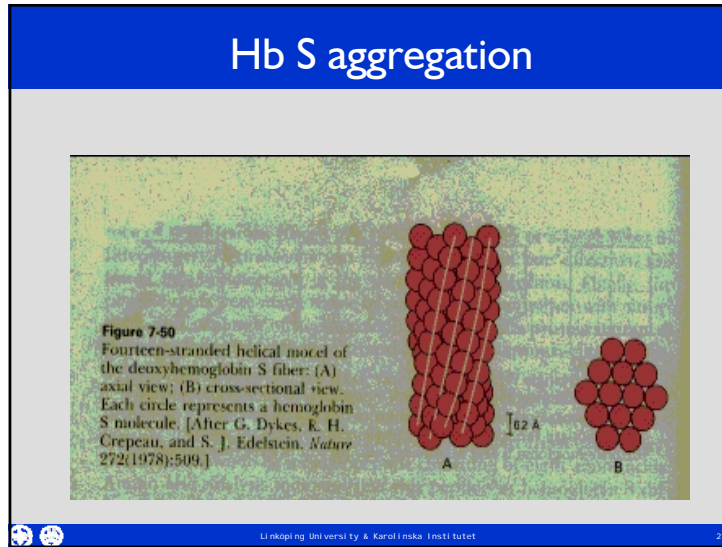


Linköping University & Karolinska Institutet 16



- ### Mutations and disease
- ★ Altered exterior
    - normally harmless but might cause disease 1
    - e.g. Hemoglobin S
  - ★ Altered active site
    - often critical, e.g. change of substrate specificity
  - ★ Altered tertiary structure
    - e.g. unstable forms (Hb Riverdale-Bronx with G->R mutation) 2
  - ★ Altered quaternary structure
    - change of allosteric properties
- Linköping University & Karolinska Institutet 19





## Parsimony

1 Ala  
2 Ser  
3 Ala  
4 Ser

2

1

2

Linköping University & Karolinska Institutet 25

## Number of unrooted trees

TABLE III  
SOME INTERESTING NUMBERS ASSOCIATED WITH  $n$  SEQUENCES

$n$	Possible unrooted trees <sup>a</sup>	Sets of four taxa <sup>b</sup>	Nearest neighbors <sup>c</sup>	Most possible <sup>d</sup>	Pairwise comparisons <sup>e</sup>	Segments in tree <sup>f</sup>
4	3	1	2	1	6	5
5	15	5	10	3	10	7
6	105	15	30	6	15	9
7	945	35	70	10	21	11
8	10,395	70	140	15	28	13
9	135,135	126	252	21	36	15
10	$2.0 \times 10^6$	210	420	28	45	17
11	$3.4 \times 10^7$	330	660	36	55	19
12	$6.6 \times 10^8$	495	990	45	66	21
13	$1.4 \times 10^9$	715	1430	55	78	23
14	$3.2 \times 10^{11}$	1001	2002	66	91	25
15	$7.9 \times 10^{12}$	1365	2730	78	105	27

<sup>a</sup> The number of unrooted trees is equal to  $[1 \times 3 \times 5 \dots (2n - 5)]$ .  
<sup>b</sup> The number of four-taxon arrangements is equal to  $n!/(n-4)! \times 4!$ .  
<sup>c</sup> Two sets of nearest neighbors are provided by each four-taxon tree.  
<sup>d</sup> The maximum number of times two taxa may be nearest neighbors is  $(n-3)(n-2)/2$ .  
<sup>e</sup> The number of pairwise comparisons is equal to  $n(n-1)/2$ .  
<sup>f</sup> The total number of segments in any unrooted tree is equal to  $2n-3$ .

from Meth. Enzymol. 183

Linköping University & Karolinska Institutet 26

## The terms *homology* and *analogy*

### Homology

Property of several species also present in *common* ancestor  
Divergent evolution

### Analogy

Property of several species *not* present in *common* ancestor  
Convergent evolution

Linköping University & Karolinska Institutet 27

## Programs for tree calculations

- ★ ClustalW
  - distance method *Des Higgins*
- ★ Phylip/Protpars
  - parsimony *Joe Felsenstein*
- ★ Tree, Papa etc.
  - progressive alignments *Russ Doolittle*
  - followed by parsimony

Linköping University & Karolinska Institutet 28

## ClustalW – main menu

```

*****
***** CLUSTAL W(1.5) Multiple Sequence Alignments *****
*****

1. Sequence Input From Disc
2. Multiple Alignments
3. Profile Alignments
4. Phylogenetic trees

S. Execute a system command
H. HELP
X. EXIT (leave program)

Your choice: 1

Sequences should all be in 1 file.

6 formats accepted:
NBRF/PIR, EMBL/SwissProt, Pearson (Fasta), GDE, Clustal, GCG/MSF.
Enter the name of the sequence file:
    
```

## Phylogenetic tree menu

1. Input an alignment
2. Exclude positions with gaps? = OFF
3. Correct for multiple substitutions? = OFF
4. Draw tree now
5. Bootstrap tree
6. Output format options

S. Execute a system command  
H. HELP  
or press [RETURN] to go back to main menu

Your choice:

## Phylogenetic tree file

```

(
(
(
(
(
(
ADH1_ALLMI:0.09794.      ADHA_HUMAN:0.02645.
(                       ADH_MACMU:0.03491)
(                       672:0.00793.
(                       ADHB_HUMAN:0.02422)
(                       ADH1_STRCA:0.03000)
(                       862:0.01244.
(                       ADHG_HUMAN:0.02419)
(                       971:0.02669.
(                       ADH1_CHICK:0.07911)
(                       995:0.03574)
(                       899:0.01605.
(                       999:0.03896.
(                       ADHE_HORSE:0.06256)
(                       675:0.00839)
(                       700:0.01223)TRICHOTOMY;
)
)
)
)
)
)
    
```

