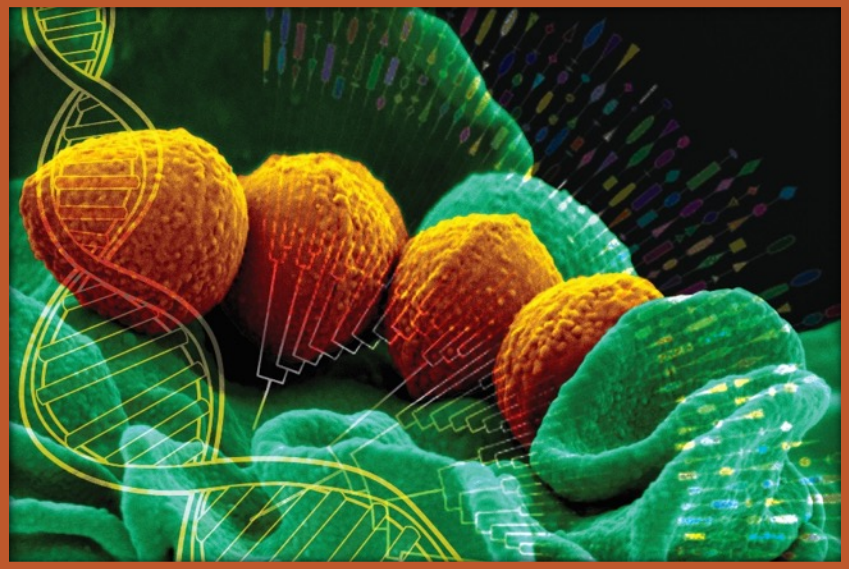
Computational Biology



Course topics & objectives

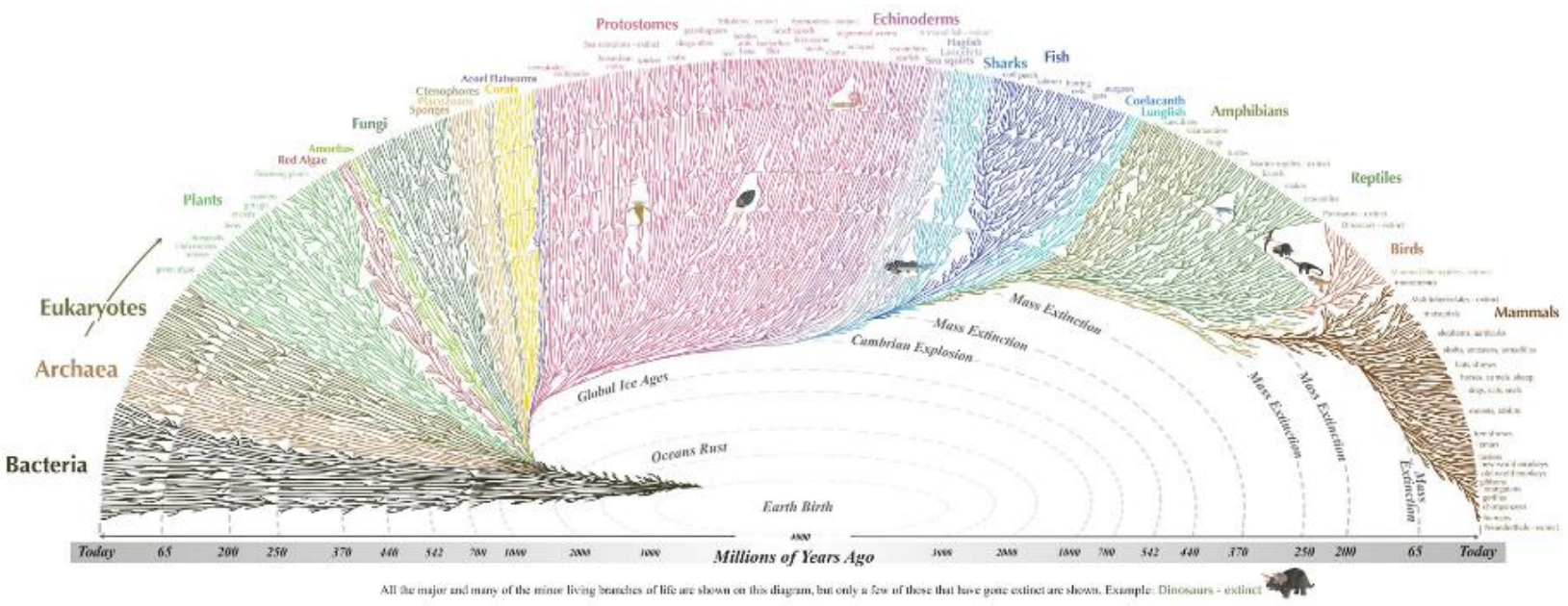
Computational biology

How to use computers to analyze biological data.

Activities

- CSI-Zoo Crime Unit --- Have a BLAST
- Protein Modeling

What is computational biology?



All the major and many of the minor living branches of life are shown on this diagram, but only a few of those that have gone extinct are shown. Example: Dinosaurs - extinct

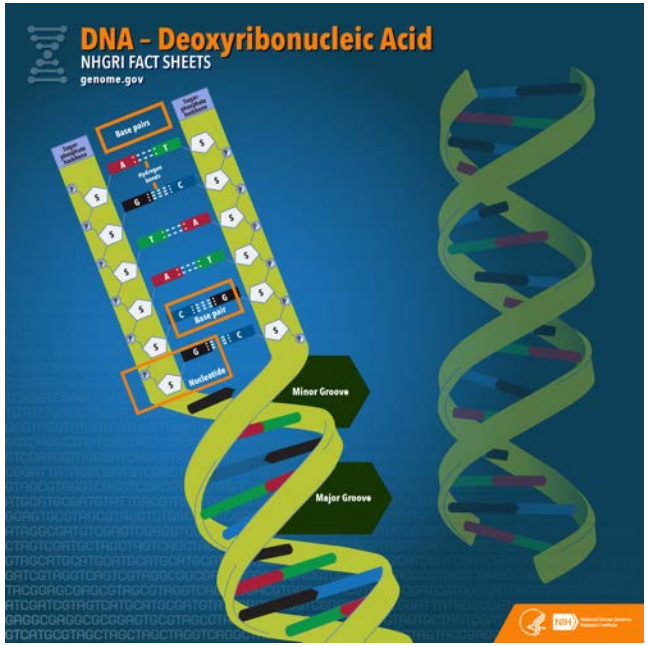
What about DNA sequencing?

Human Genome

- More than 3 billion base pairs
- 25,000 to 30,000 genes

Bacterial Genomes

- Between 1 million and 5 million base pairs
- 1,000 to 5,000 genes



What do we do with the data?

All this data is stored in very large databases, such as NCBI

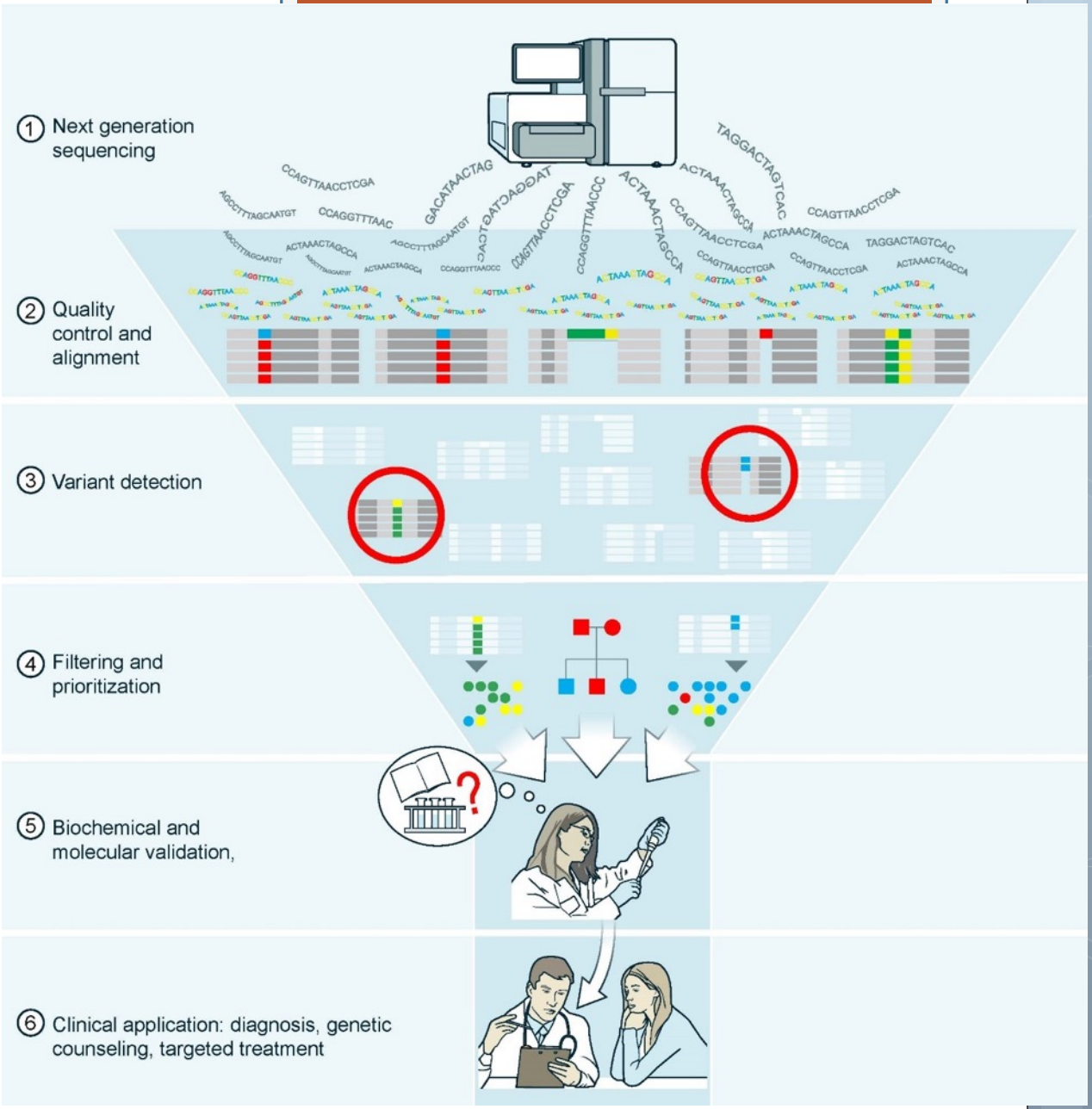
- 1.2 quadrillion basepairs of sequence data
- Representing many thousands of animal, plant, bacteria, and viral specimens

NCBI

National Center for

Biotechnology Information

Computational techniques can help us understand and organize large volumes of genetic data.



Comparing sequence information

Seq_1 ATCGGGGGT
Seq_2 ATGGGAGT

Seq_1 ATCGGGGGT
Seq_2 AT---GGGAGT

What do these sequences tell you?

```
seq_1 ATCGGGTGCTGCTGCTGGGCGCTGCGGCTGCGTGCGC
seq_2 ATGGGCGCGTCGTCTGGGGGTTGTCTGGGTCGTGCTT
seq_3 ATCGGGCTGGCGGCTGCGTGTGTGGGCTGTTGCGTCG
seq_4 ATGGGTGCGTCGCTGGGCTTGCTGGGGGCGTCTGGTG
seq_5 ATGGGCGCGTCGTCTGGGGGTTGTCTGGGTCGTGCTT
sea_6 ATCGGGTGCTGCTGCTGGGCGCTGCGGCTGCGTGCGC
seq_7 ATGGGCGCGTCGTCTGGGGGTTGTCTGGGTCGTGCTT
seq_8 ATCGGGTGCTGCTGCTGGGCGCTGCGGCTGCGTGCGC
```

Crime Scene Investigators: Zoo Crime Unit

- Crime scene at the Zoo
- A Panda was found dead
- DNA samples were taken from a knife and other evidence found at the scene

Computational Biology Activity 1

Use computational tools to discover perpetrator

CSI- ZCU activity

What are the DNA sequences?

- BLAST the DNA sequences
- Determine which DNA sequence belongs to primary suspect

Who did it?

- Align the DNA evidence sequences with those of all suspects

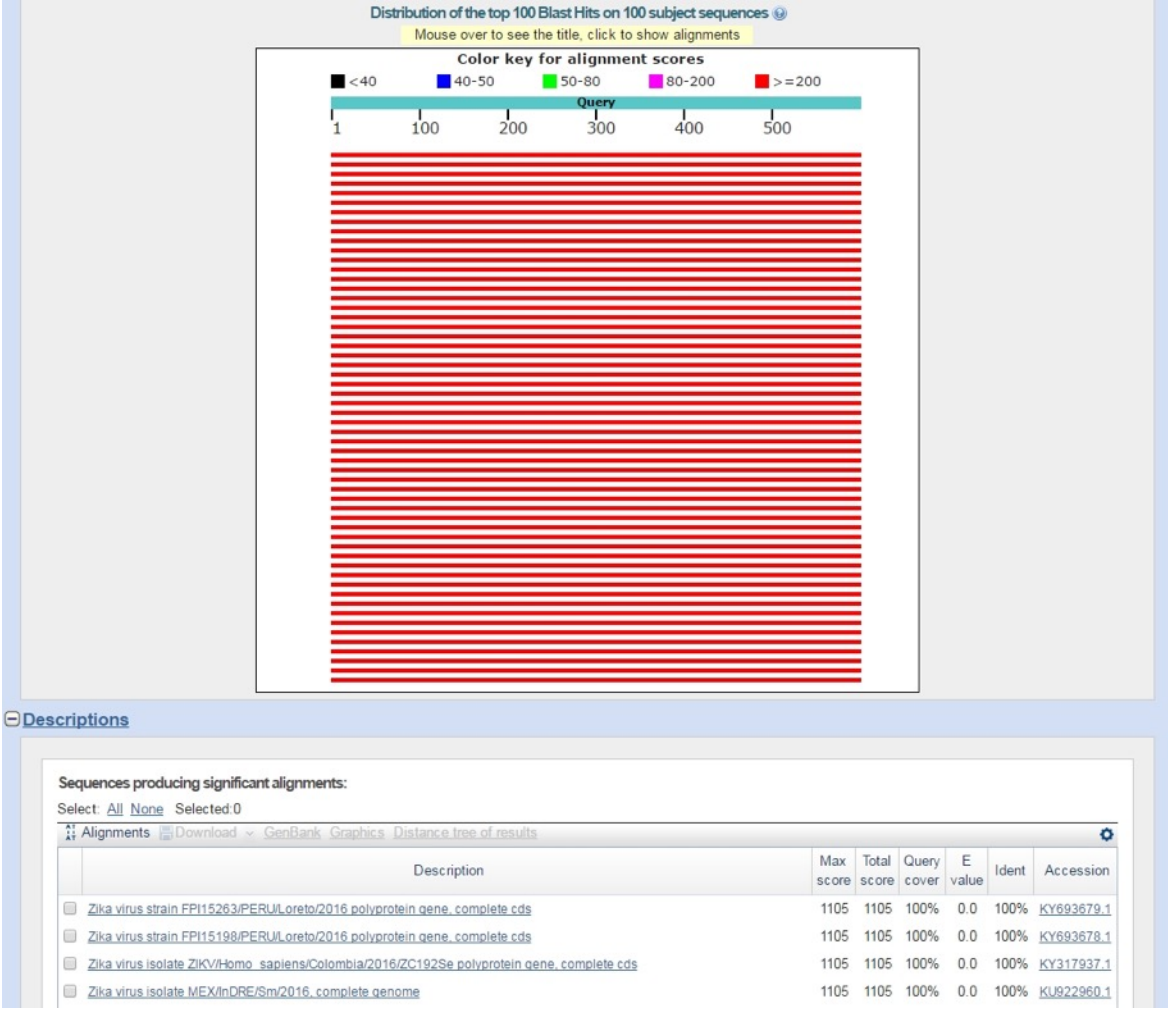
CSI- ZCU activity

BLAST – basic local alignment tool

- http://blast.ncbi.nlm.nih.gov/Blast.cgi
- Allows for comparison of sequences of interest with a huge database of known sequences

>sequence
AAGAAAGATCTGGCTGCCATGCTGAGAATAATCAATGCTAGGAAGGAGAAGAAGAGACGAGGCGCAGAAA
CTAGTGTCGGAATTGTTGGCCTCCTGCTGACCACAGCTATGGCAGCGGAGGTCACTAGACGTGGGAGTGC
ATACTATATGTACTTGGACAGAAACGATGCTGGGGAGGCCATATCTTTTCCAACCACATTGGGGATGAAT
AAGTGTTATATACAGATCATGGATCTTGGACACATGTGTGATGCCACCATGAGCTATGAATGCCCTATGC
TGGATGAGGGGGTGGAACCAGATGACGTCGATTGTTGGTGCAACACGACGTCAACTTGGGTTGTGTACGG
AACCTGCCATCACAAAAAAGGTGAAGCACGGAGATCTAGAAGAGCCGTGACGCTCCCCTCCCATTCCACT
AGGAAGCTGCAAACGCGGTCGCAAACCTGGTTGGAATCAAGAGAATACACAAAGCACTTGATTAGAGTCG
AAAATTGGATATTCAGGAACCCTGGTTTCGCTTTAGCAGCAGCTGCCATCGCGTGGCTTTTGGGAAGCTC
AACGAGCCAAAAAGTCATATACTTGGTCATGATACTGC

Example of BLAST results



CSI-ZCU activity

Align all sequences - DNA

- http://www.ebi.ac.uk/Tools/msa/muscle/
- Determine how two or more sequences are related

```
seq_1 GCAGGGCGGCACCAAGCTGGTGAAGGACTTAC
seq_2 GCAGGGCGGCCTGGTGAAGCACTTAC
```

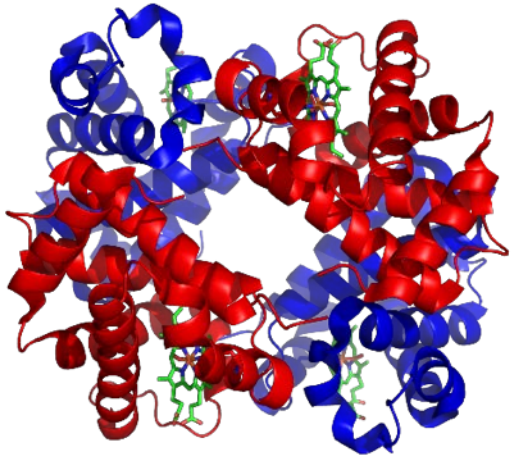
```
seq_1 GCAGGGCGGCACCAAGCTGGTGAAGGACTTAC
seq_2 GCAGGGCGGC------CTGGTGAAGCACTTAC
```

Proteins – the cellular workhorse

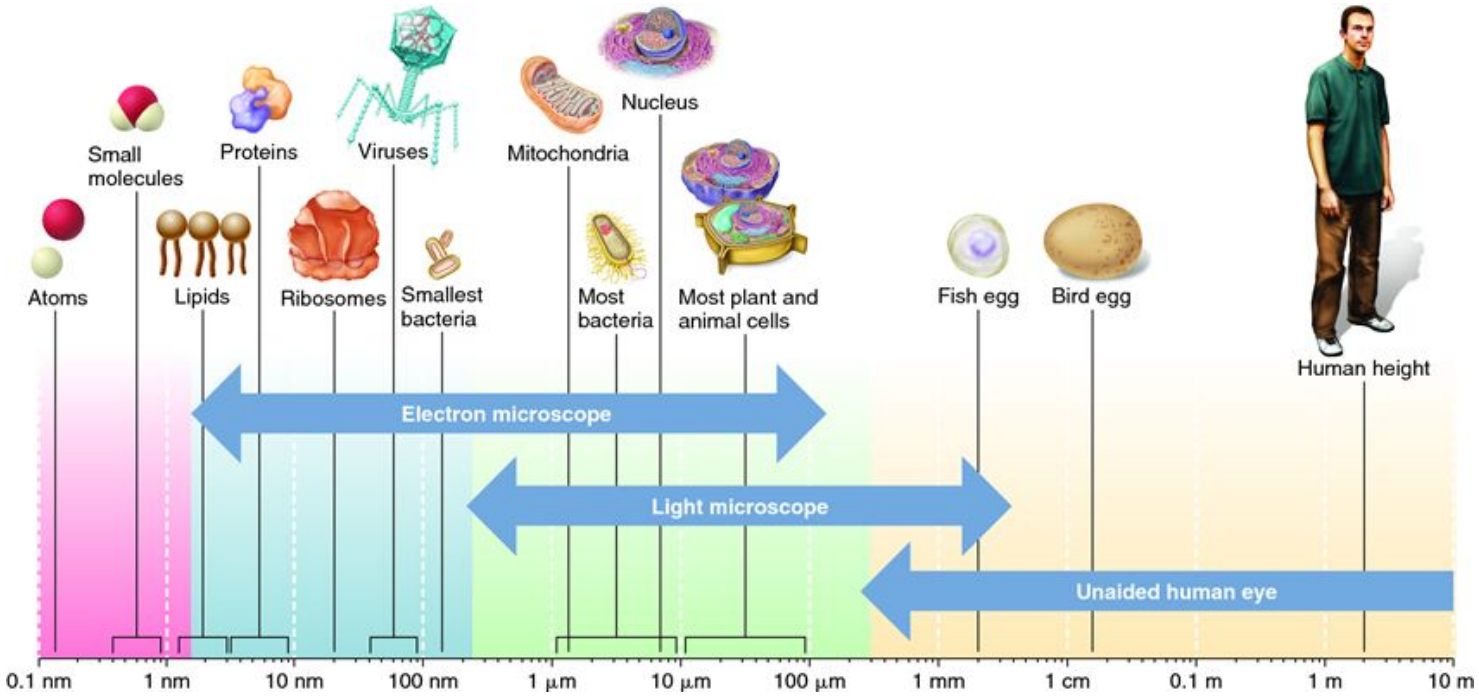
Proteins – the cellular workhorse

Proteins mediate virtually all cellular functions

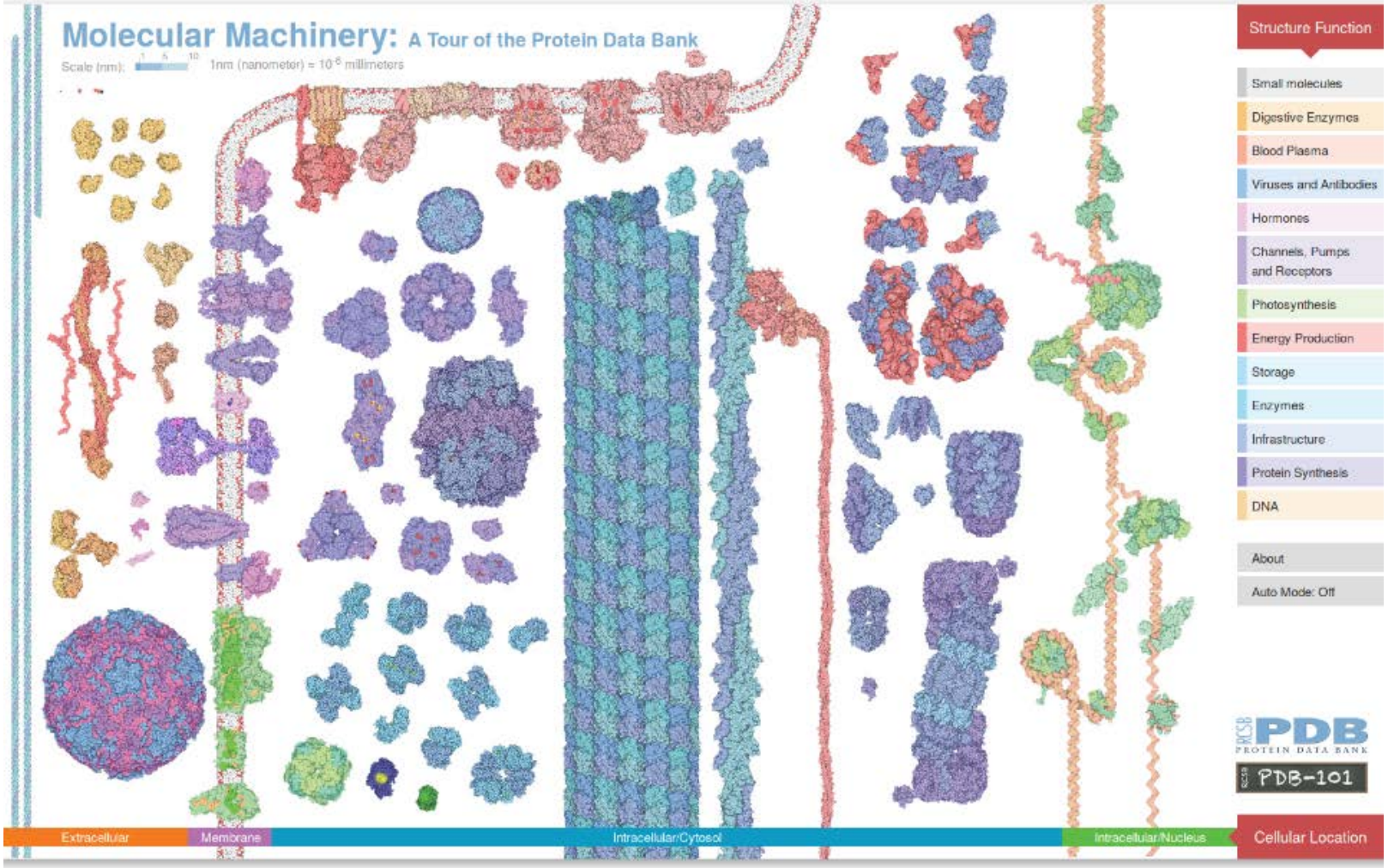
- Cellular Structure
- Transport across membranes
- Response to cellular environment
- Receptor ligands
- Energy metabolism
- Movement



How big is a protein molecule?

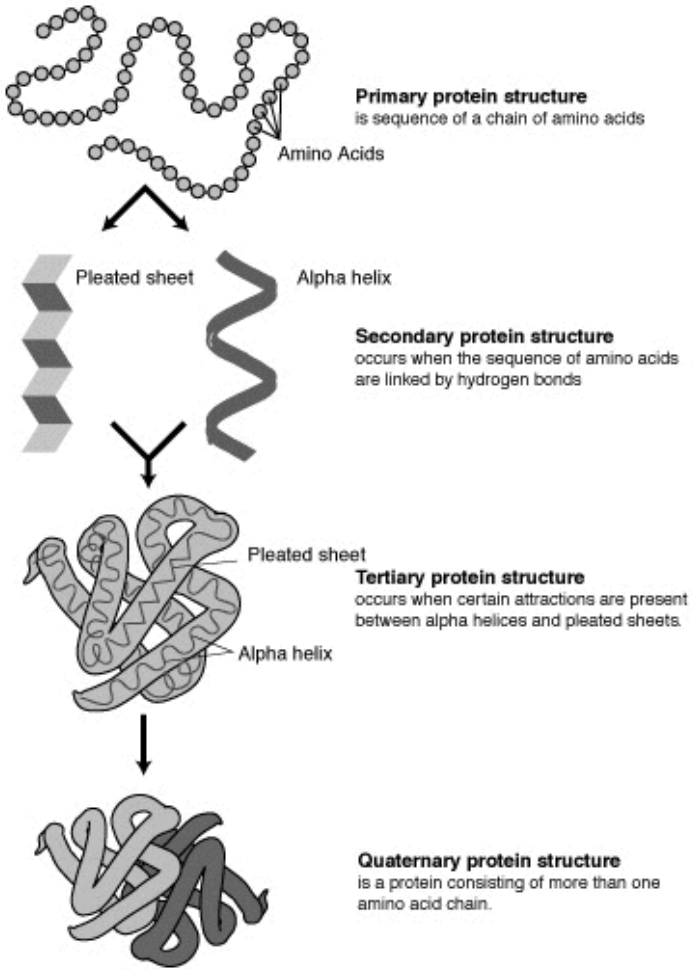


Proteins – molecular machinery



Structure of proteins

- Proteins are composed of 20 different amino acid
- There are four levels of protein structure



Structure of proteins

Hands on activity

- Go to http://www.ebi.ac.uk/pdbe/
- Type 2HYY in the search box.
- Click on 2hyy.
- Click on 3d-visualization.
- In “Polymer Visual -> Type” pulldown menu – choose surface.
- Rotate with left mouse click. Zoom in and out with right mouse click.
- You will see four BCR-ABL protein molecule. Rotate and try to locate the drug Gleevec. You can change the “HET Groups Visual -> Type” to VDW Balls to see the drug better.
- Play with other display option from “Polymer Visual -> Type” pulldown menu.
- Change “Water Visual -> Type” to VDW Balls to see the Oxygen atoms of the Water molecules around the protein molecule.