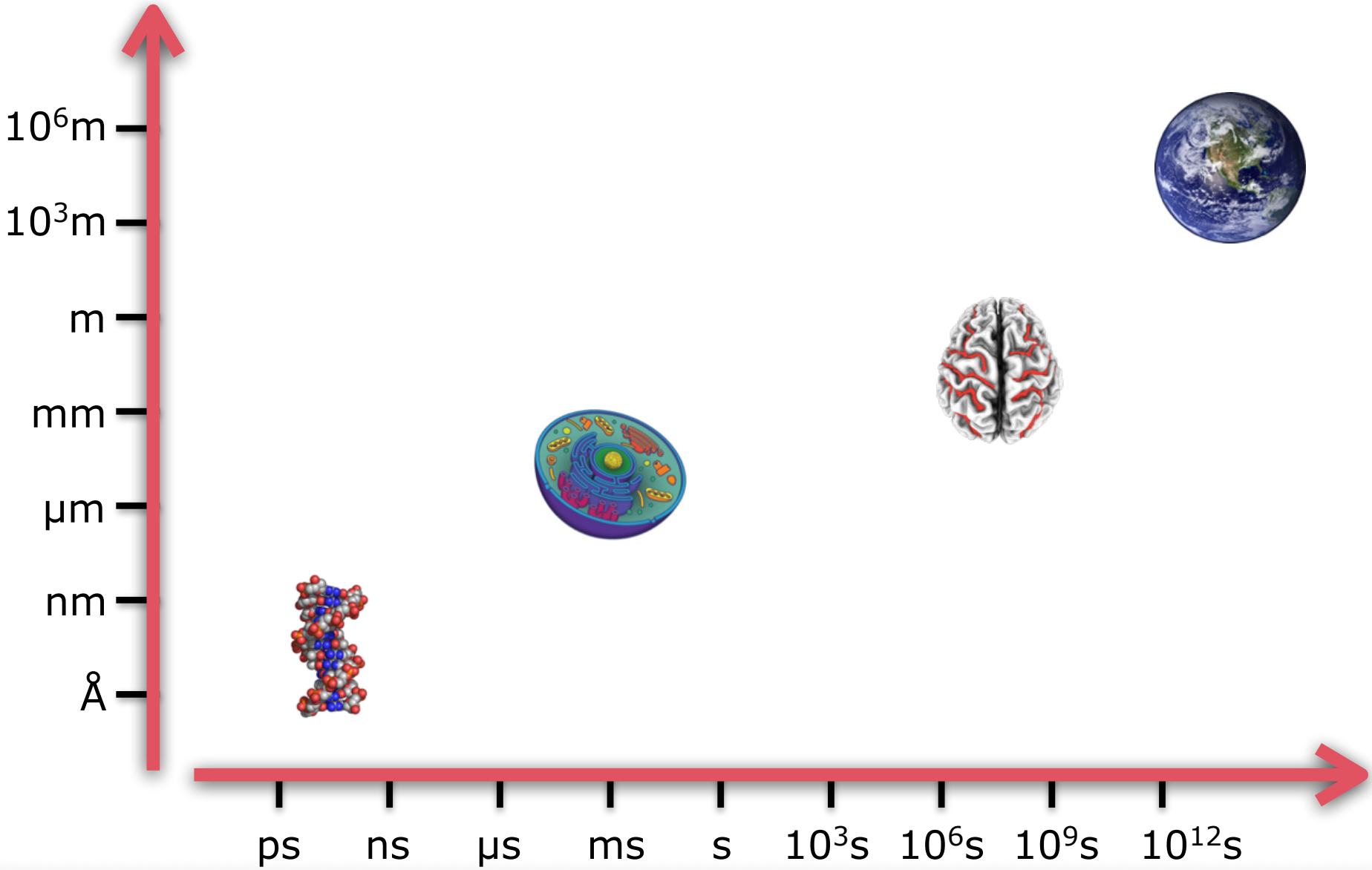
Tutorial 1
Geometry, Topology, and Biology

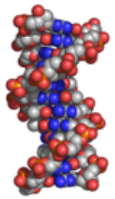
Patrice Koehl and Joel Hass

University of California, Davis, USA
http://www.cs.ucdavis.edu/~koehl/IMS2017/

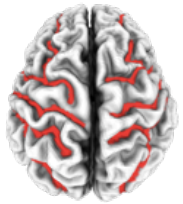
Biology = Multiscale ….



Biology = Quantitative Science ….



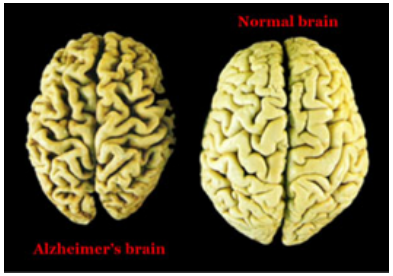


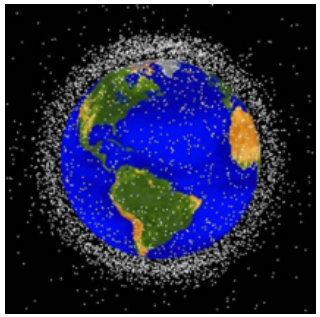


Acetylaminofluorene

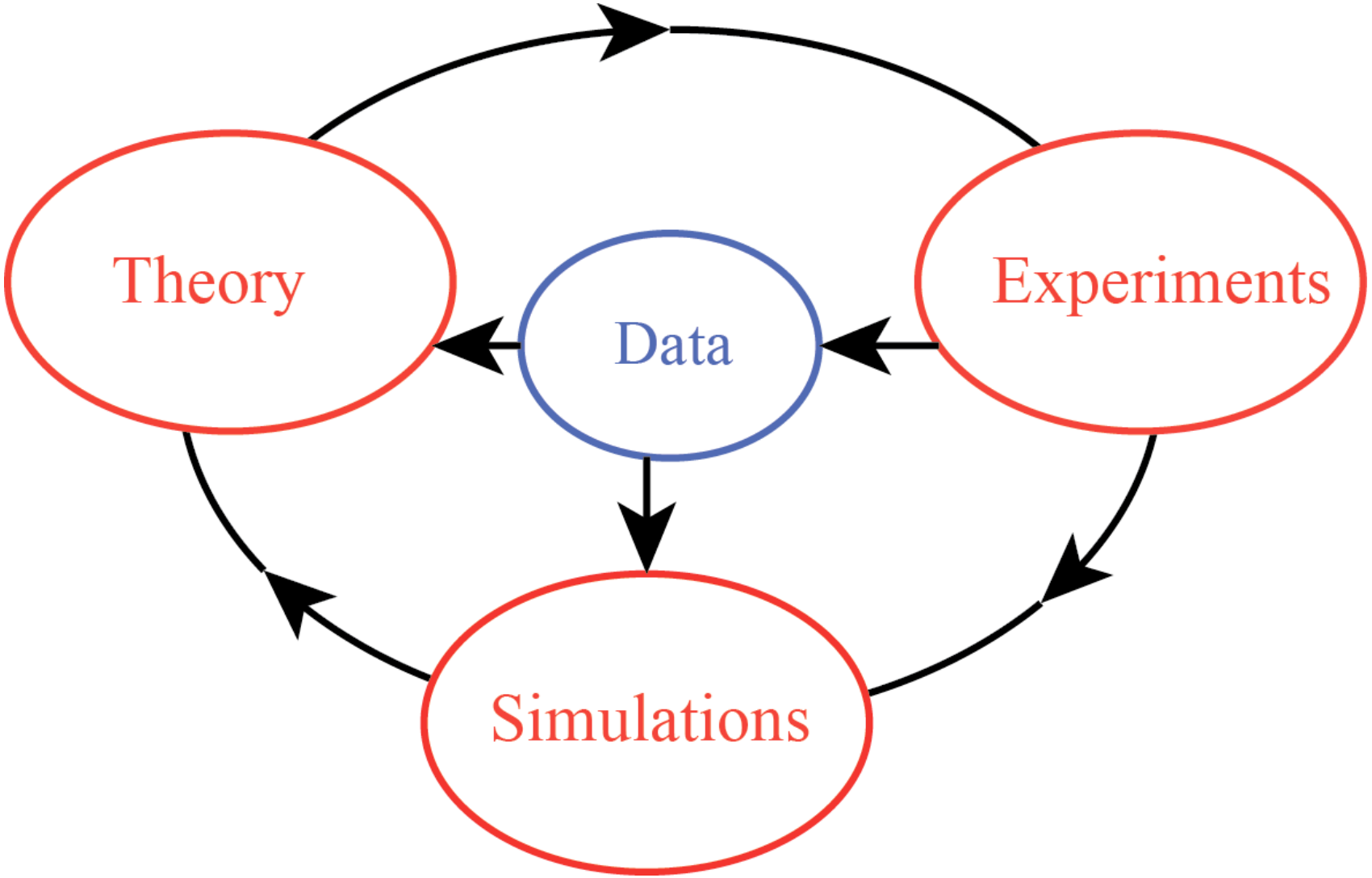
Normal Red Blood Cell

Sickle Cell

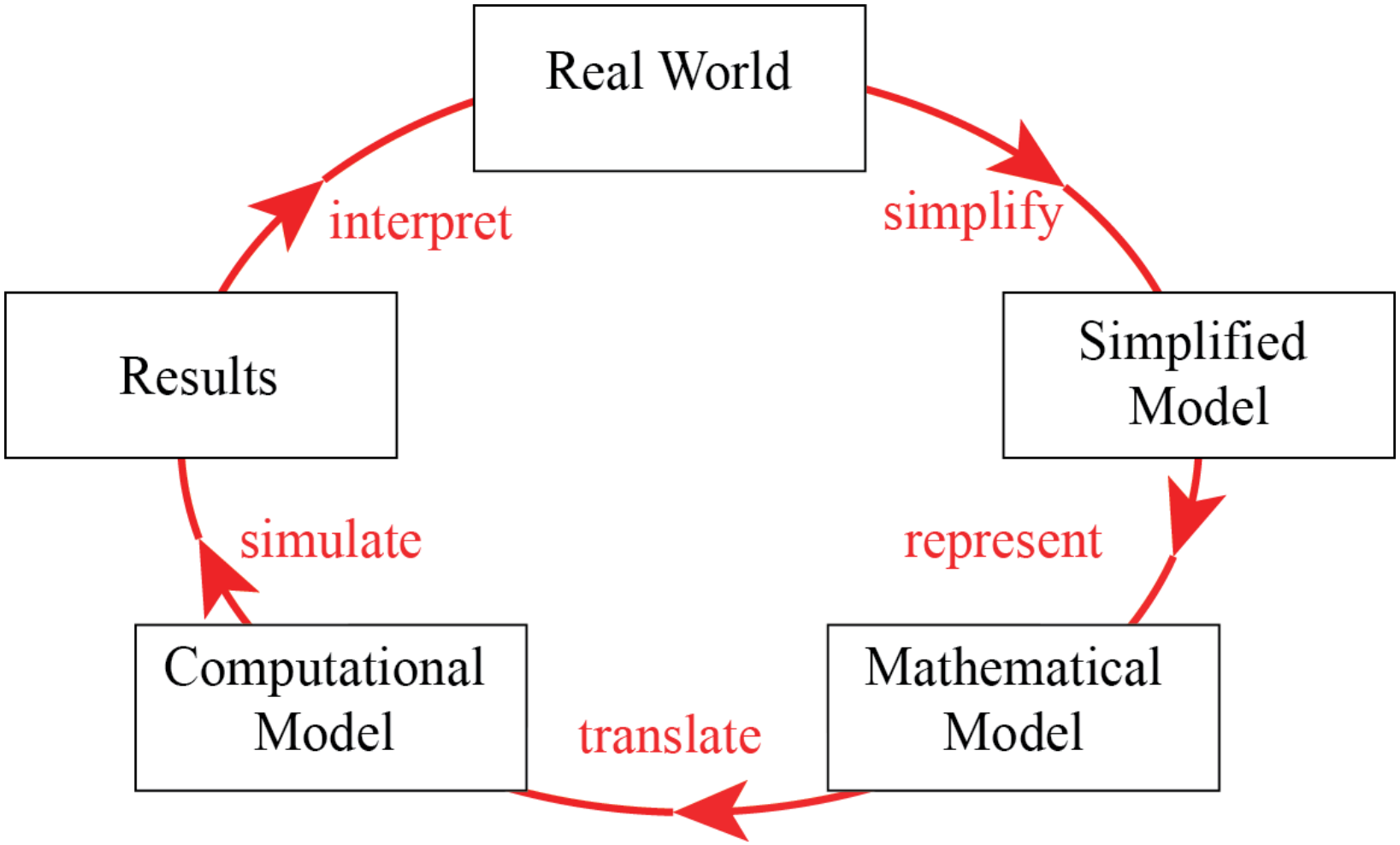




Biology = Data Science ….

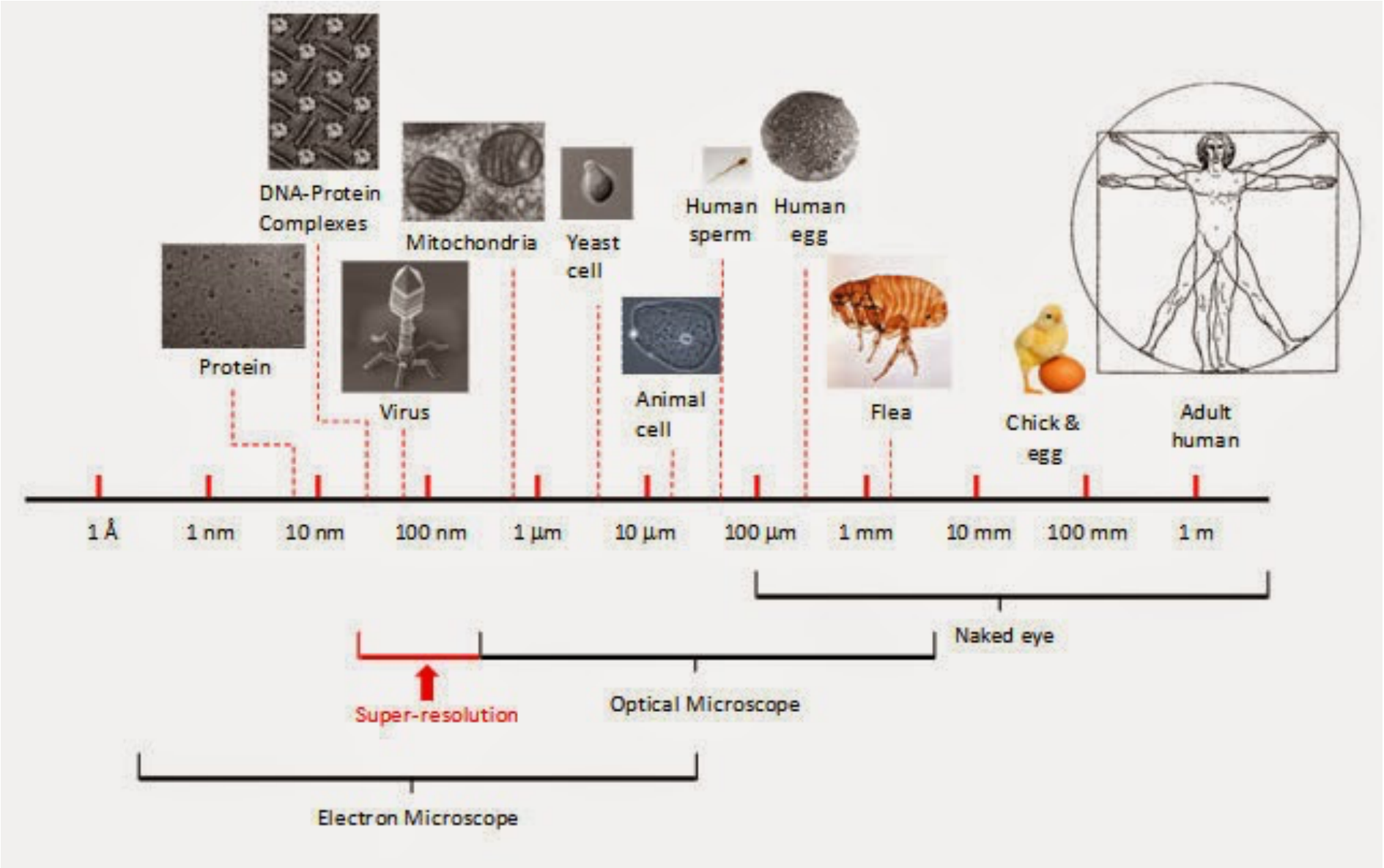


Biology = Quantitative Science ….



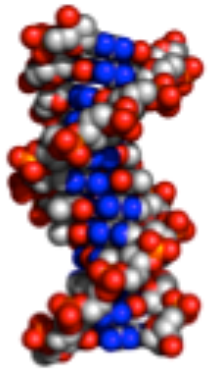
Mathematical Modeling

- Is often used in place of experiments when they are *too large, too expensive, too dangerous, or too time consuming*.
- Can be useful in "what if" studies; e.g. to investigate the use of *pathogens* (viruses, bacteria) to control an insect population.
- Is a modern tool for *scientific investigation.*


DNA-Protein Complexes
Mitochondria
Yeast cell
Human sperm
Human egg
Protein
Virus
Animal cell
Flea
Chick & egg
Adult human
1 Å
1 nm
10 nm
100 nm
1 µm
10 µm
100 µm
1 mm
10 mm
100 mm
1 m
Naked eye
Super-resolution
Optical Microscope
Electron Microscope

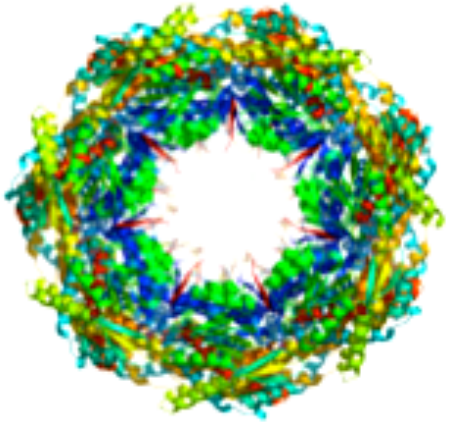
The Cell

Blueprint: parts + design



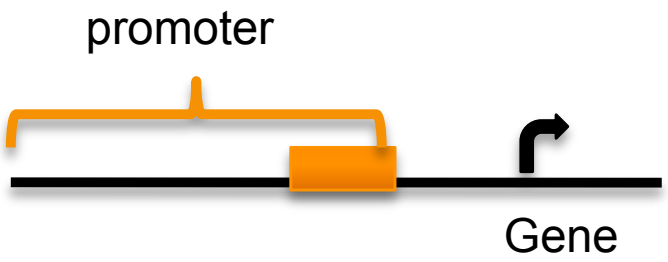
DNA, genes

Assembly



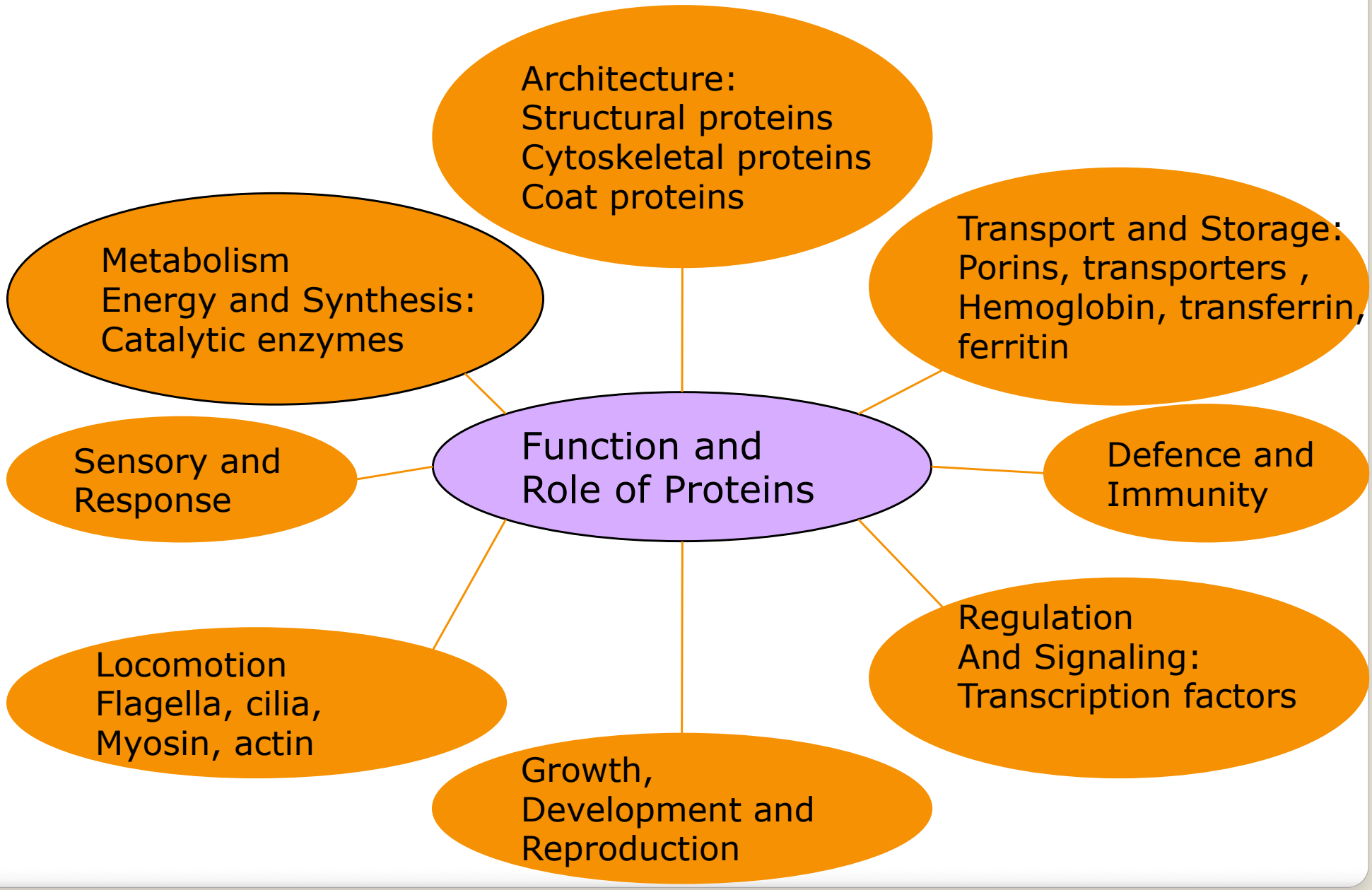
Proteins, nanomachines

Operation “manual”



Processing: logic

Why Proteins?



Protein Structure

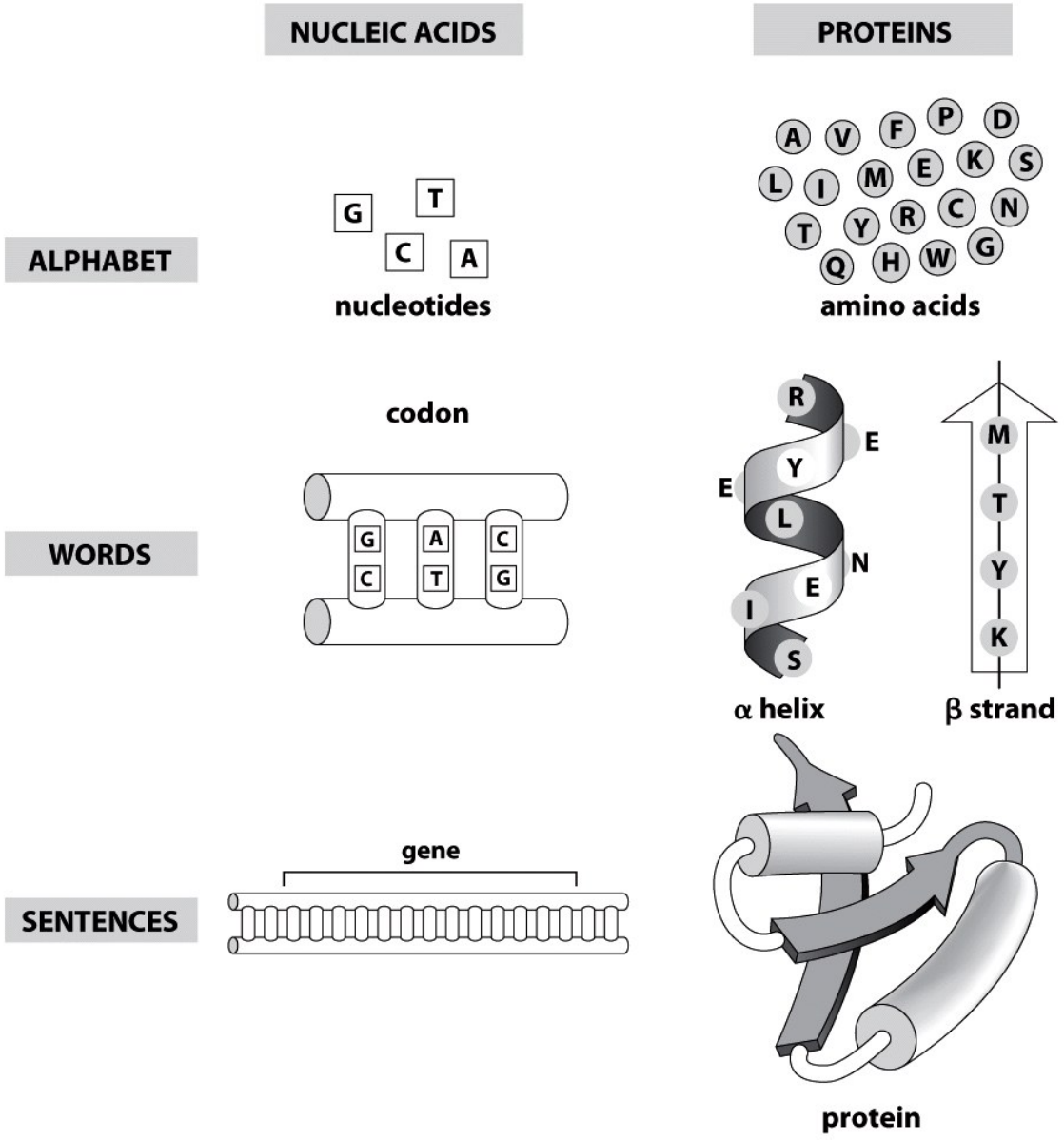


Figure 1.2 Physical Biology of the Cell (© Garland Science 2009)

Observing biomolecules

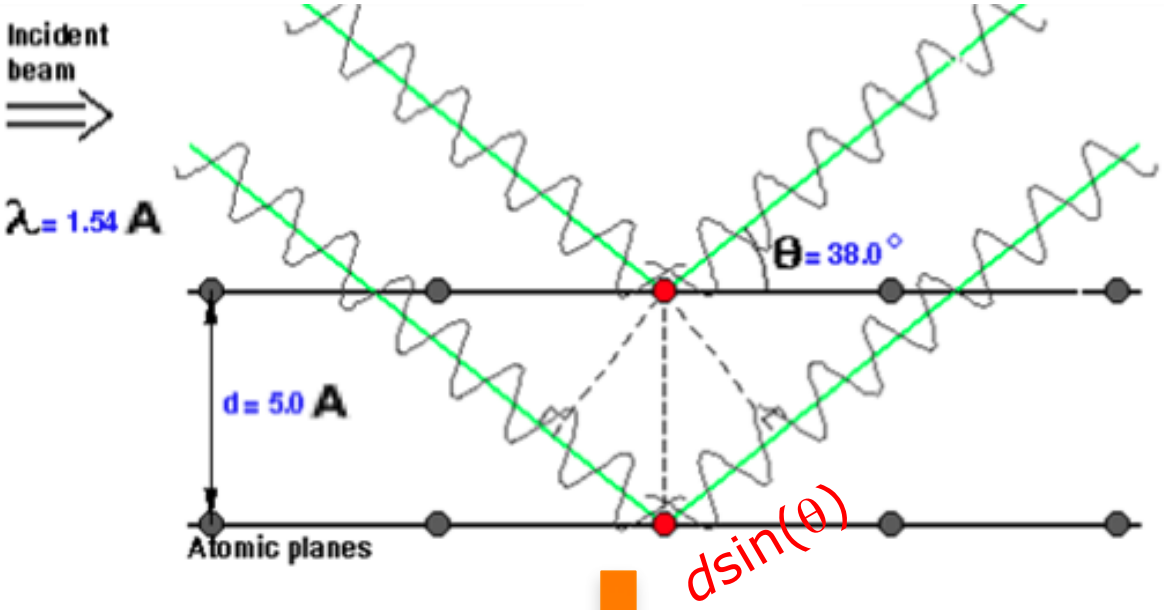
Methods for finding the 3D structure of a biomolecule:

-X-ray crystallography
(high resolution; finds structure of a protein in a crystal)

- NMR spectroscopy
(high resolution; finds structure of a protein in solution)

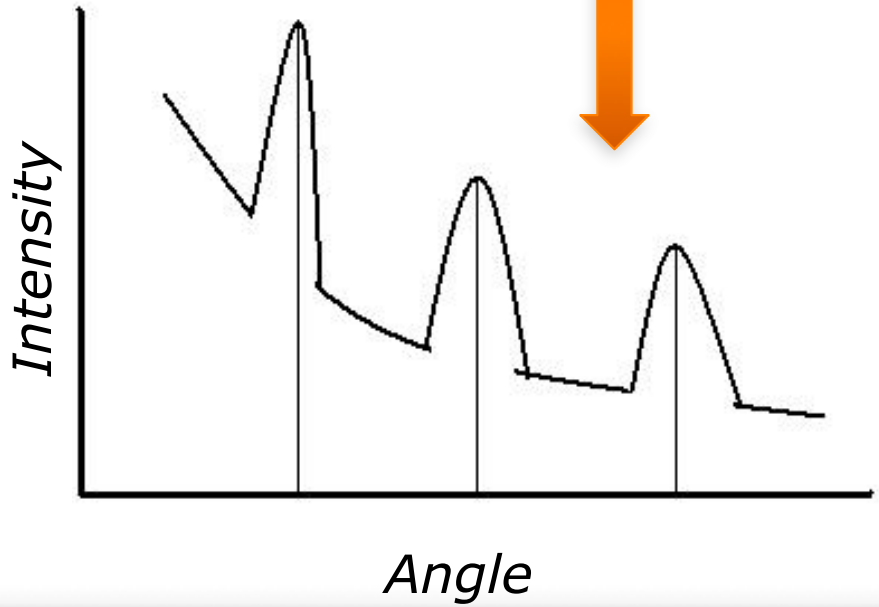
- Cryo EM
(medium -high resolution; finds structure of large systems

Proteins: X-ray Crystallography

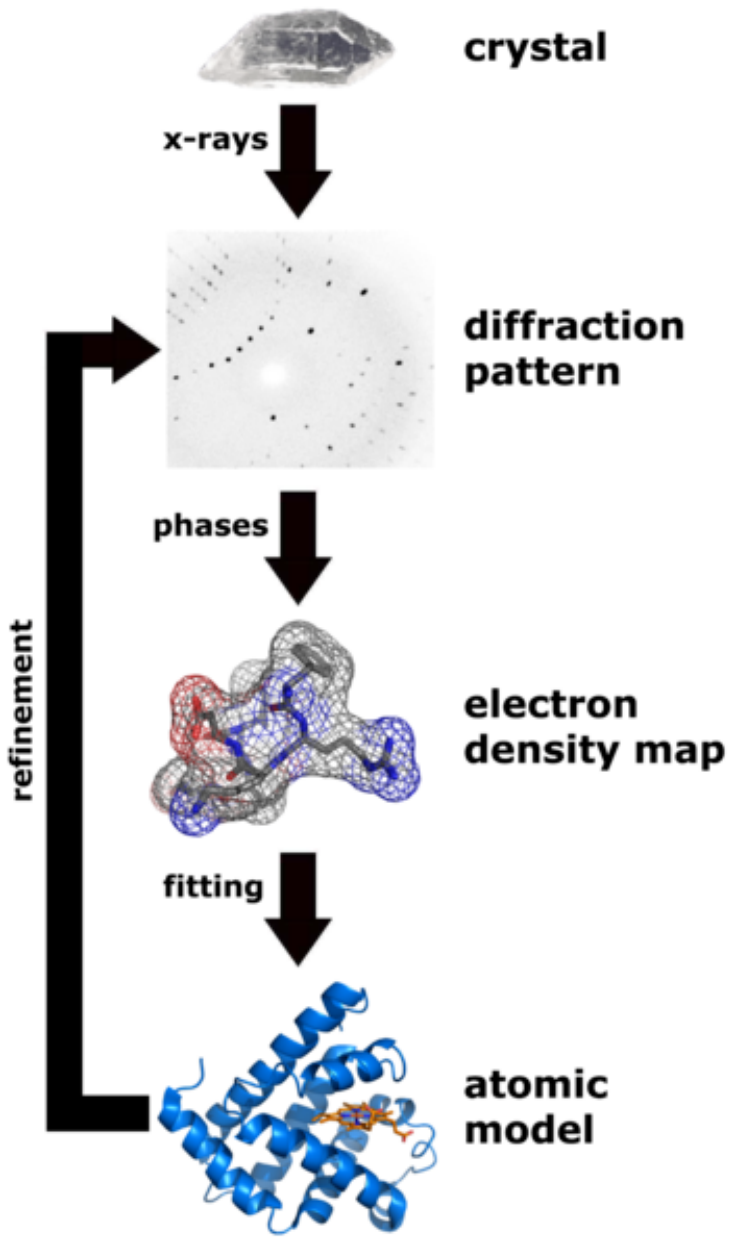


Bragg's Law:

$$2d\sin(\theta) = n\lambda$$



From the "pattern of diffraction", i.e. the maximum of intensities observed, we can find the angles of diffraction and for each angle we get the corresponding d using Bragg's law



General principle of X-ray crystallography applied to proteins:

1) We need a crystal

2) From the diffraction pattern, we get the crystal organization

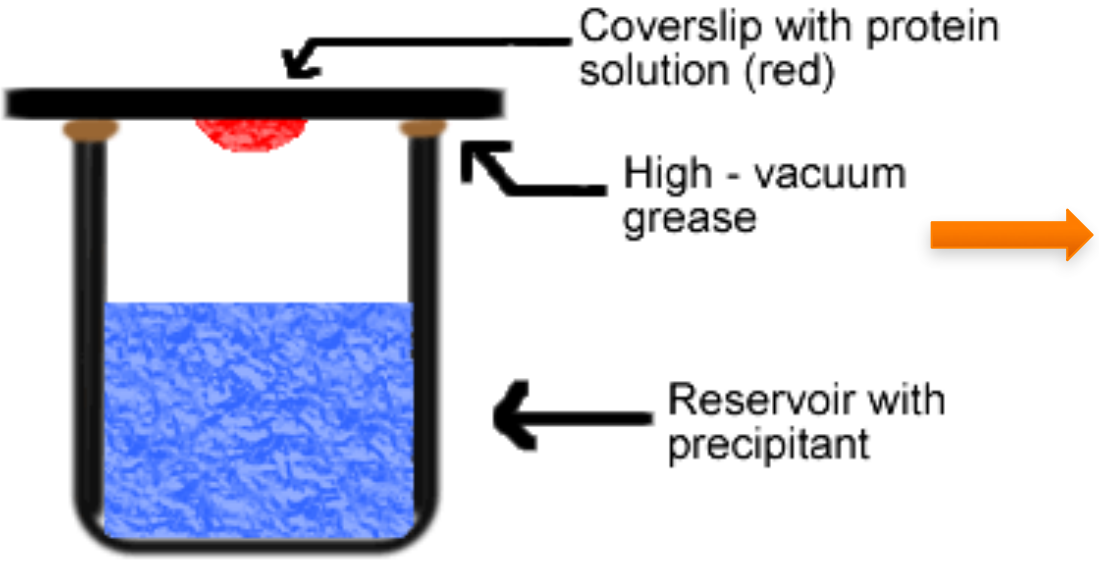
3) From the diffraction intensities, we get the electron densities

4) Once the electron density map we fit a structure that matches with this density

5) From the atomic model, we can compute a theoretical diffraction map; if it matches with the experimental one, we are done; otherwise refine

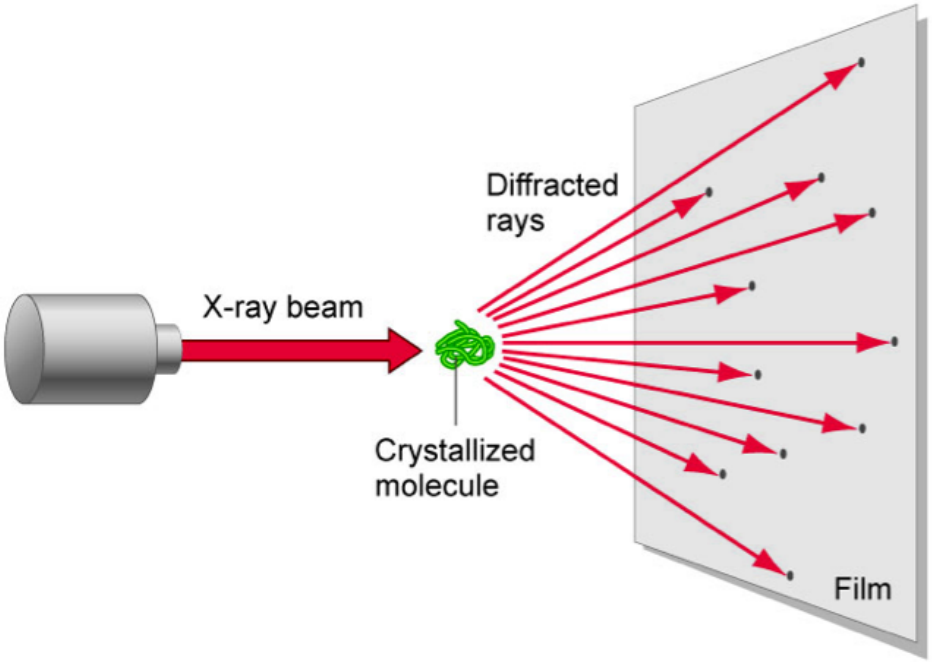
Getting a crystal

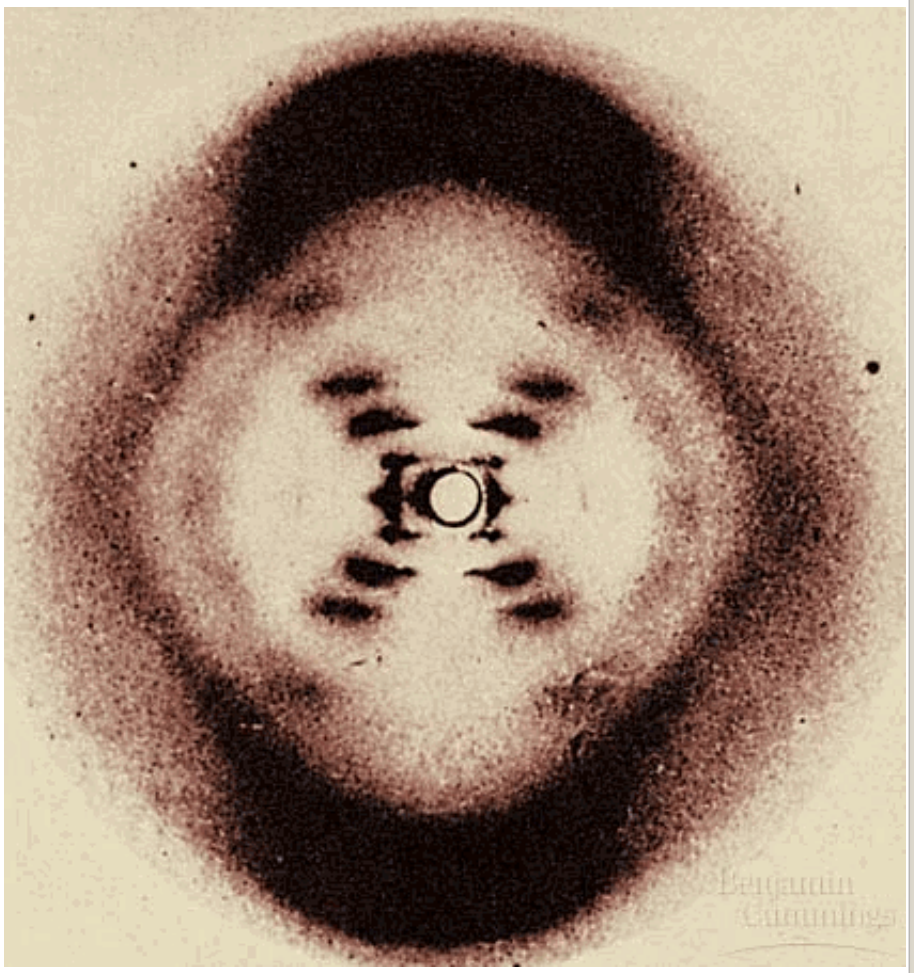
The “hanging drop” method



(http://www.molbio1.princeton.edu/macro/about.html)

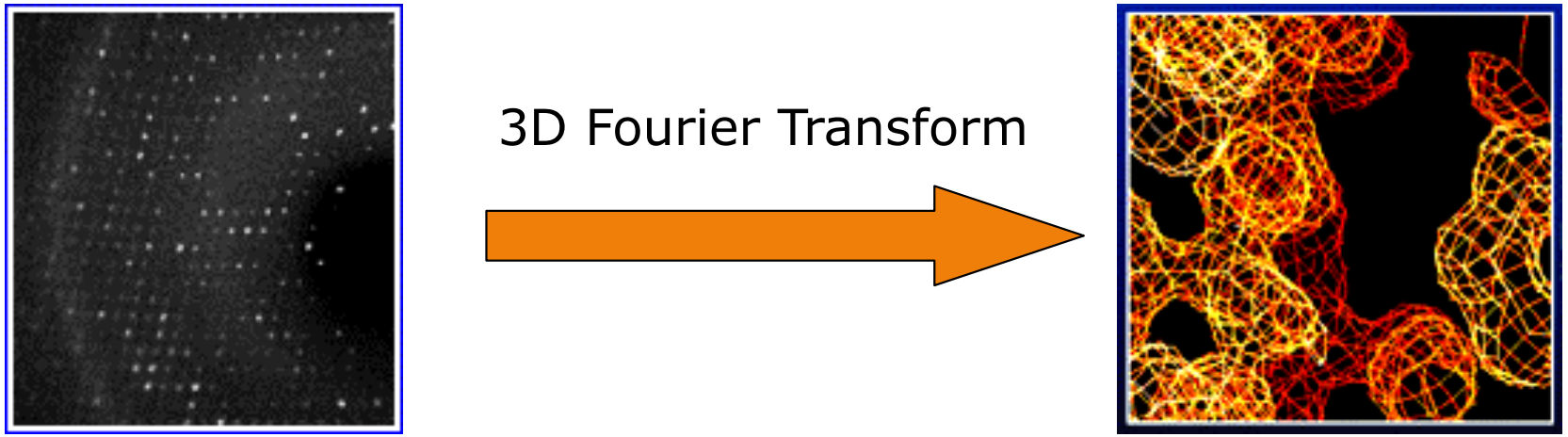
Getting the Diffraction Pattern





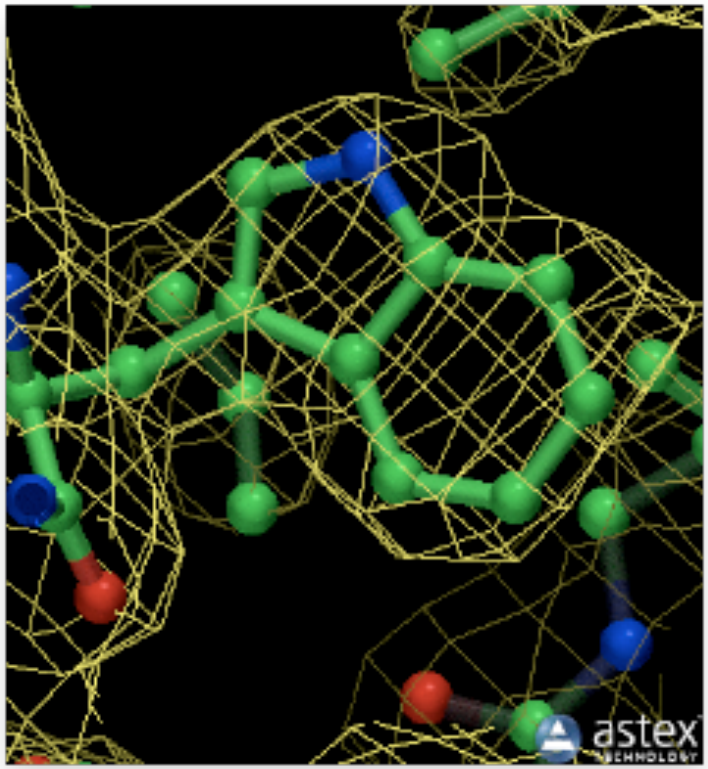
Rosalyn Franklin:
Diffraction pattern for DNA

From Diffraction to Electron Density Map

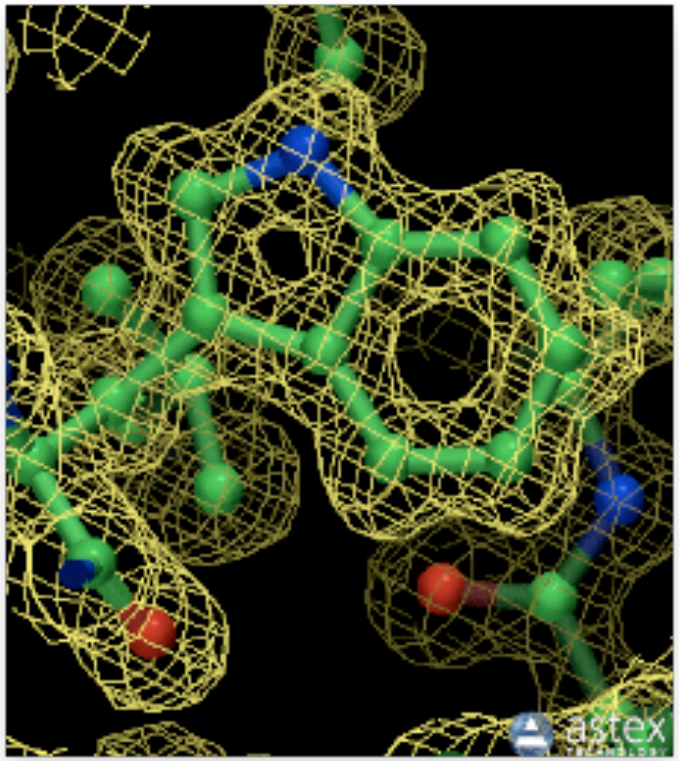


One hidden problem: diffraction patterns provide intensities; for Fourier transform, need intensity and phase. A significant step in X-ray crystallography is the solve the "phase problem".

Fitting the structure: Influence of the resolution



2.6 Å resolution



1.2 Å resolution

Resolution of X-ray structures

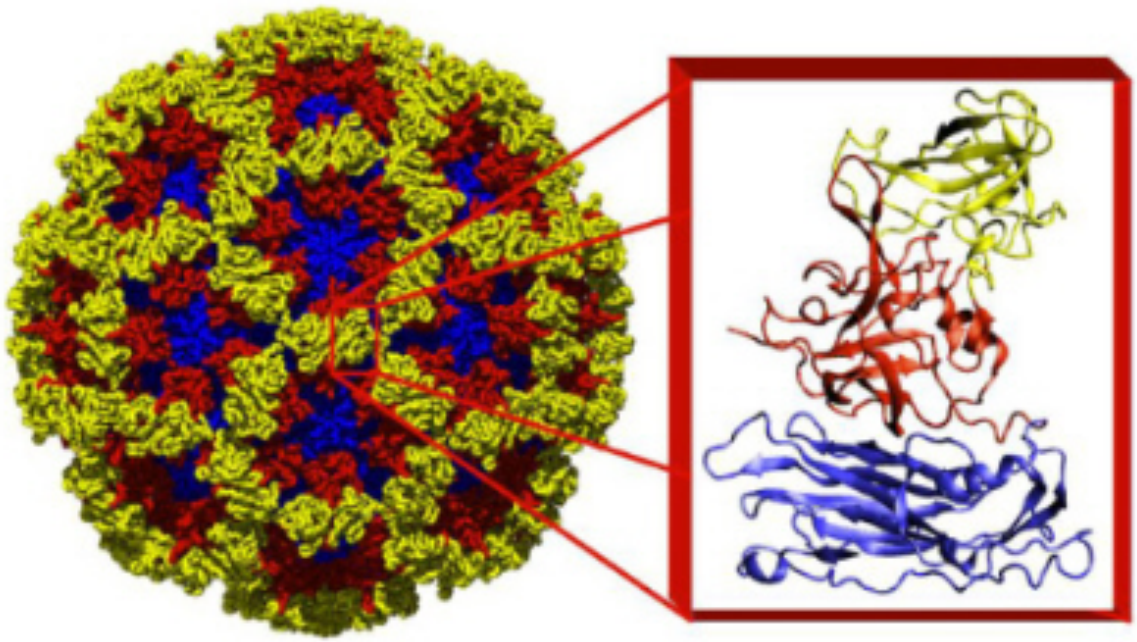
Resolution (Å)	Meaning
>4.0	Individual coordinates meaningless
3.0 - 4.0	Fold possibly correct, but errors are very likely.
2.5 - 3.0	Fold likely correct except that some surface loops might be mis-modelled.
2.0 - 3.0	Many small errors can normally be detected. Fold normally correct and number of errors in surface loops is small. Water molecules and small ligands become visible.
1.5 - 2.0	Many small errors can normally be detected. Folds are extremely rarely incorrect, even in surface loops.
0.5 - 1.5	In general, structures have almost no errors at this resolution. geometry studies are made from these structures.

(http://en.wikipedia.org/wiki/Resolution_(electron_density)

Large molecular assemblies: X-ray crystallography and Cryo-EM

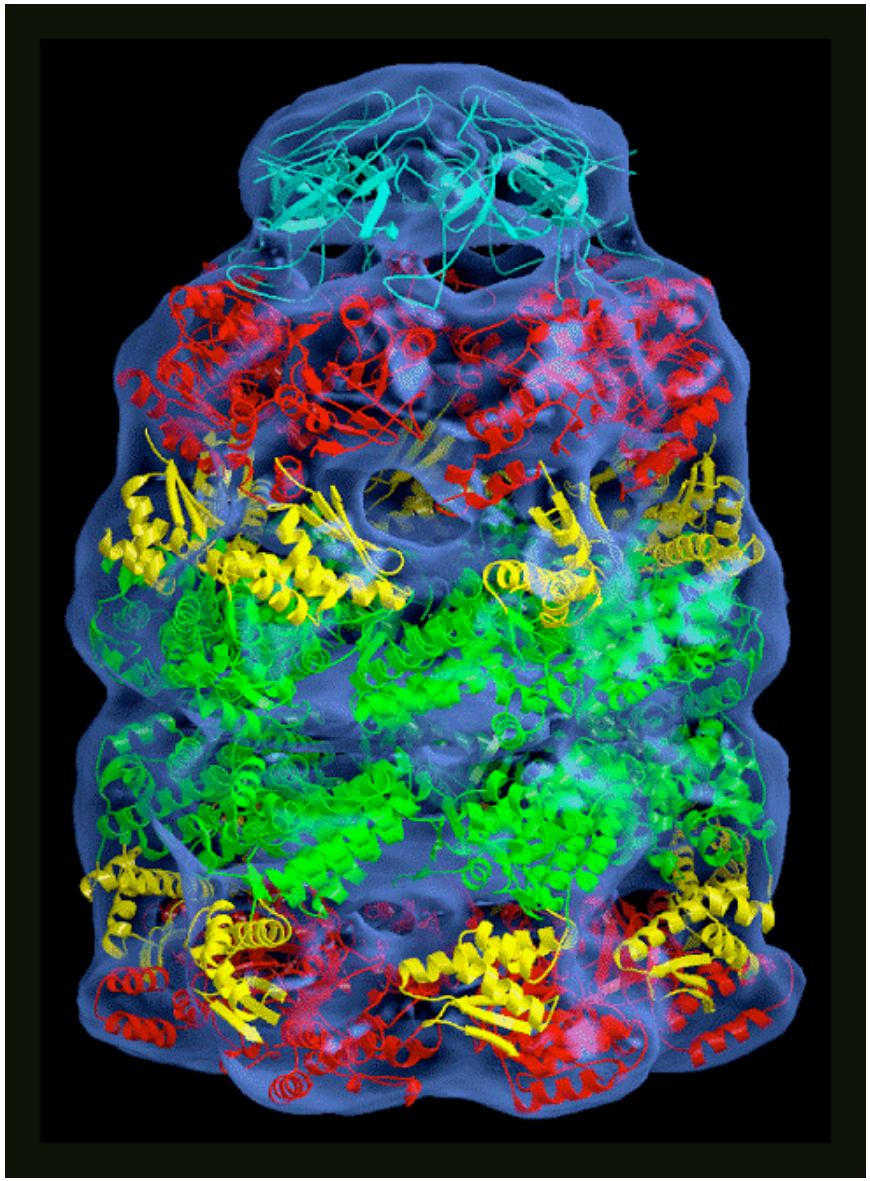
X-ray structure

(180 copies of
the same protein)



(Norwalk virus: http://www.bcm.edu/molvir/norovirus)

Large molecular assemblies: X-ray crystallography and Cryo-EM

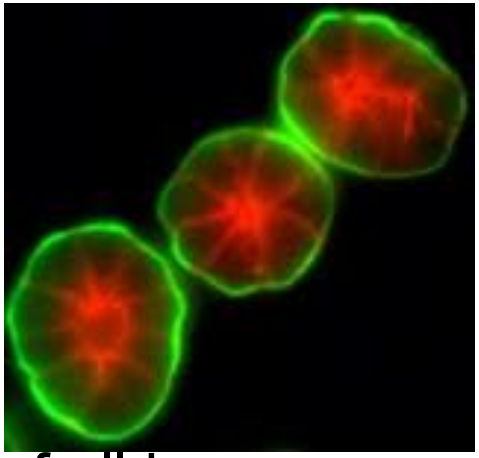


Cryo-EM:

-Microscopy technique;
as such, do not need crystal
(closer to physiological
conditions)

-Not high-resolution enough
to provide atomic details;
used in combination with
modeling

The natural world of the cell



-The cell is the basic structural and functional unit of all known living organisms.

-The cell is the smallest unit of life

-The biological information contained in an organism is encoded in its DNA sequence

-Biological information encodes for diversity and function:

Information = Work

-Cells are self-replicating: mother cell “generates” identical daughter cells

The Cell

1. **Quantum of life:**
from 1 cell (bacteria), to 10^{13} in a human
(trivia: there are approx. 10^{14} bacteria in our guts!
- we generate approx. 10^{16} cells during our life time)

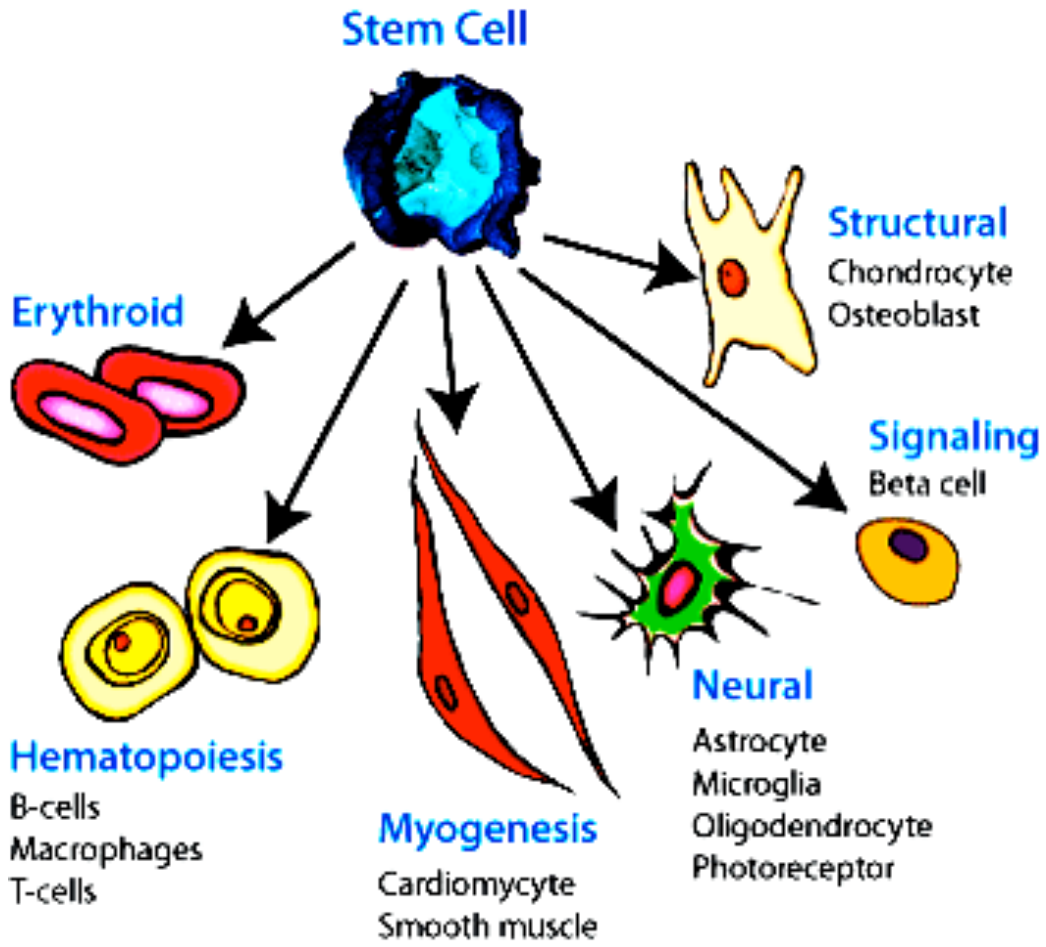
2. Cells are "machines":
They can produce chemical or mechanical work. They take energy from their environment.

3. Cells self-replicate

4. Their blueprint is the DNA they contain

The Cell

All cells of an organism contain the same information…however They may differ in aspect….and functions.



Observing cells

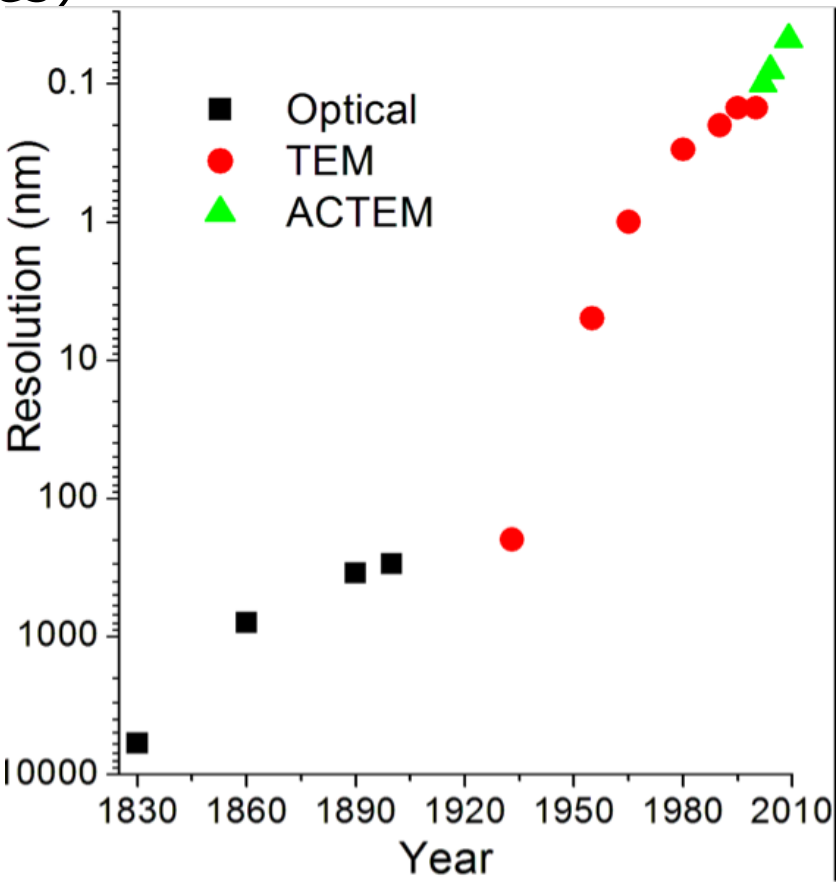
Methods for finding the geometry of cells:

-Light microscopy
(resolution 0.2 μm; see organelles)

- Fluorescence microscopy
(see individual molecules inside the cells)

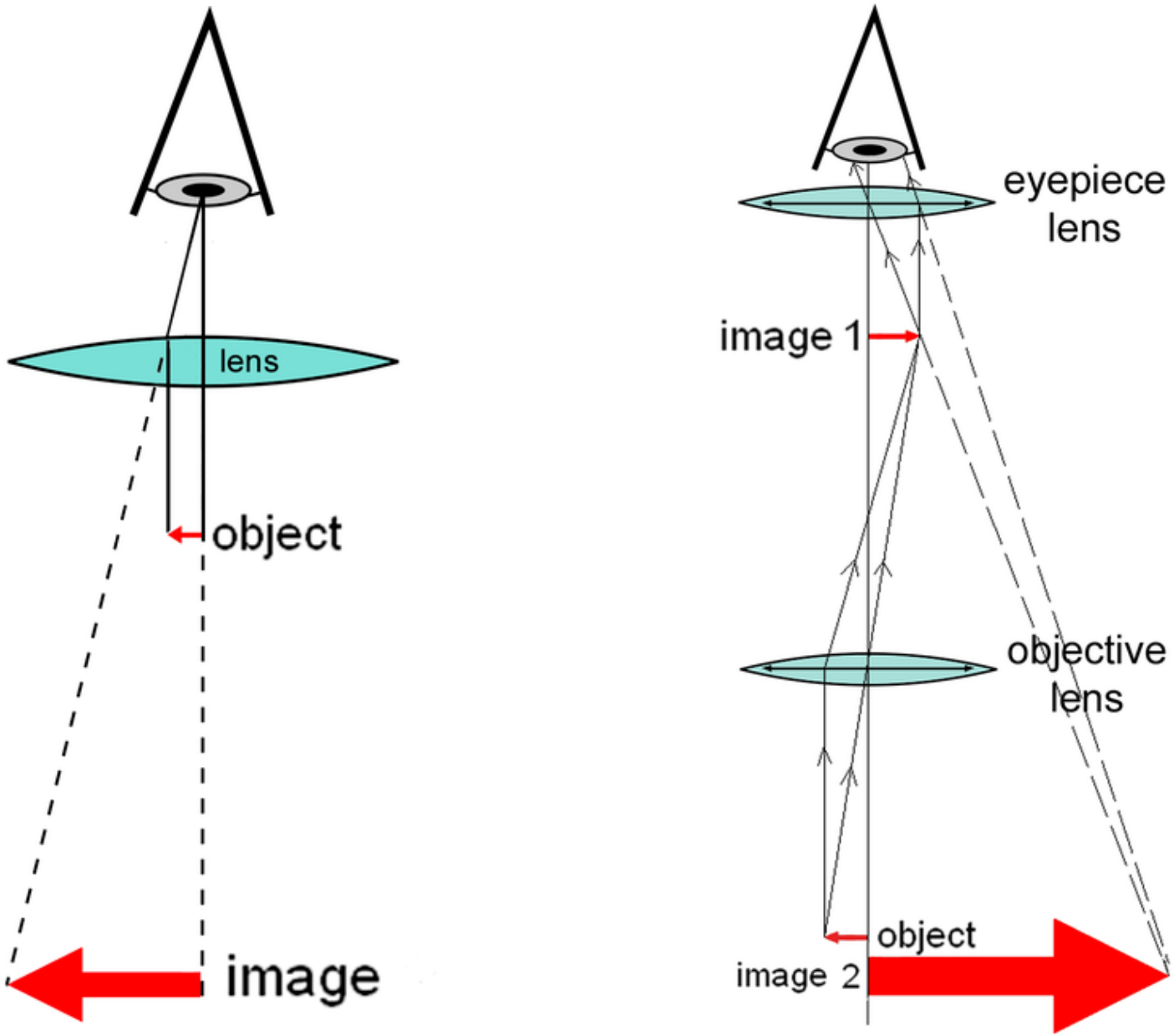
- Electron microscopy
(high resolution)

Optical: light microscopy
TEM: Transmission EM
ACTEM: Aberration Corrected TEM

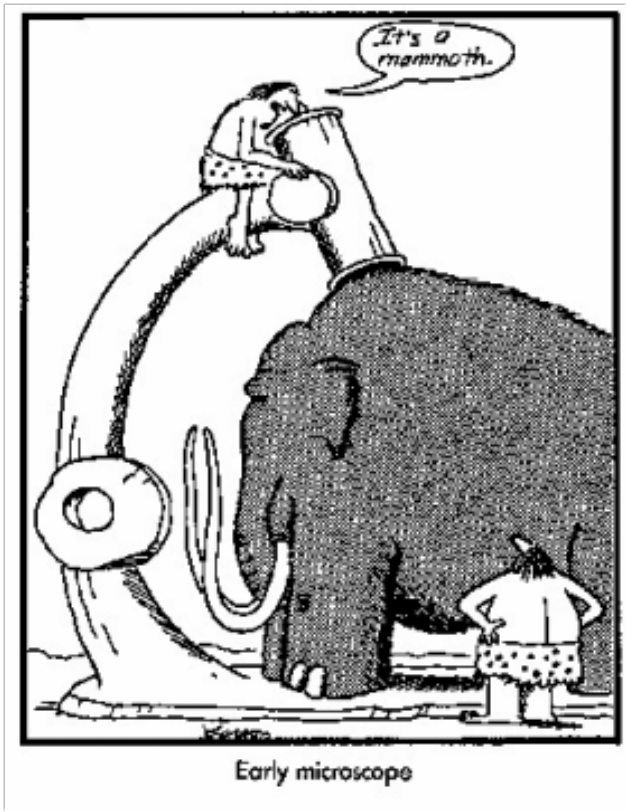


(from https://en.wikipedia.org/wiki/Microscope)

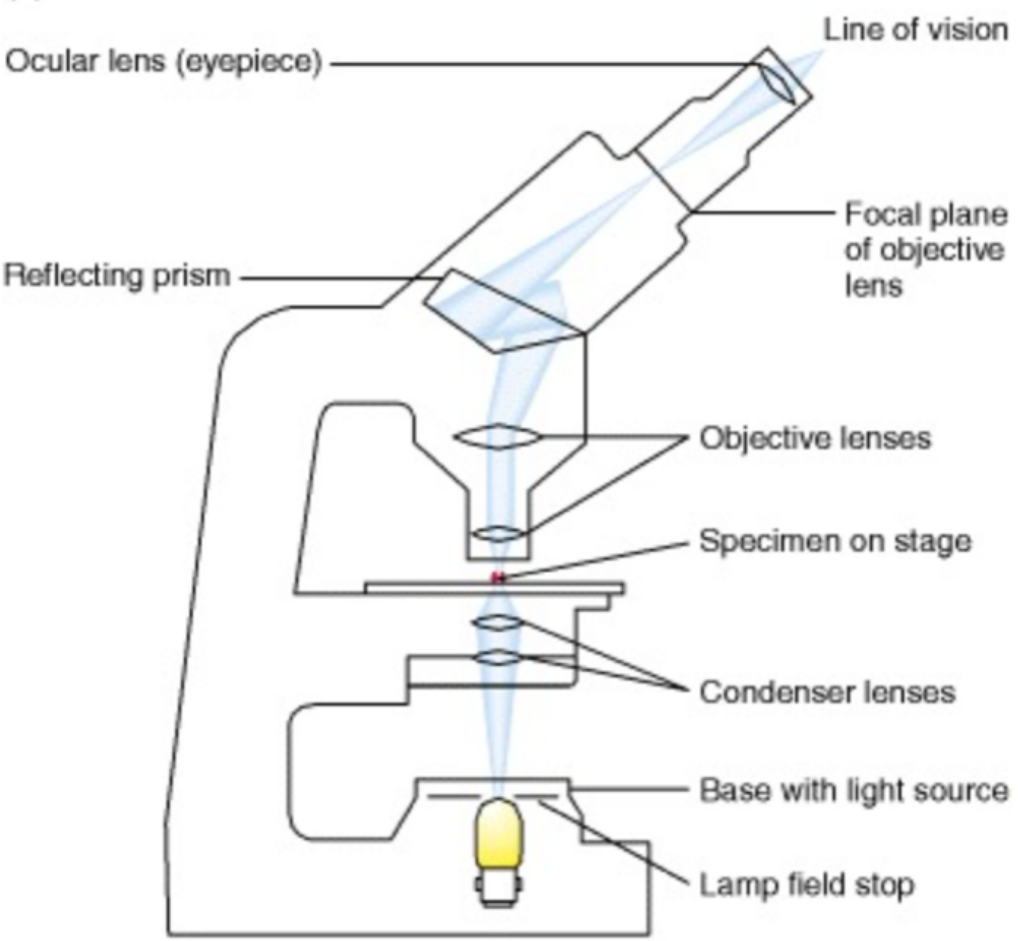
Light Microscopes



Light Microscopes



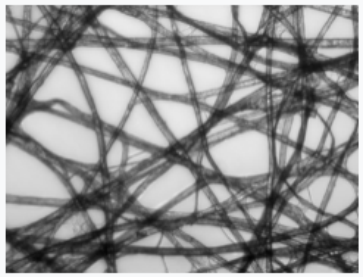
Early microscope



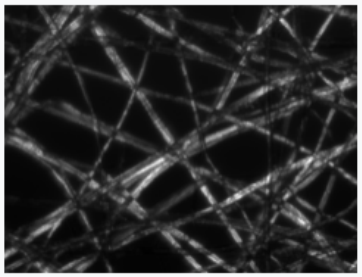
Gary Larson, "Far Side"

(from https://www.ncbi.nlm.nih.gov/books/NBK21629/)

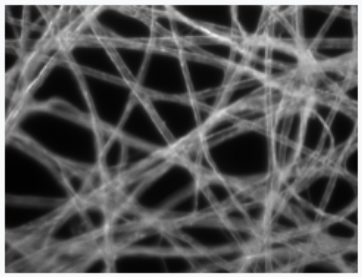
Light Microscopes



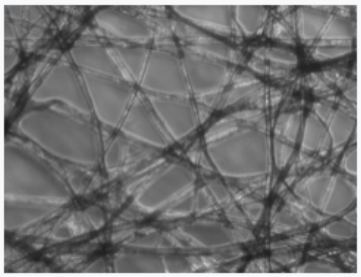
Bright field illumination, sample contrast comes from absorbance of light in the sample.



Cross-polarized light illumination, sample contrast comes from rotation of polarized light through the sample.



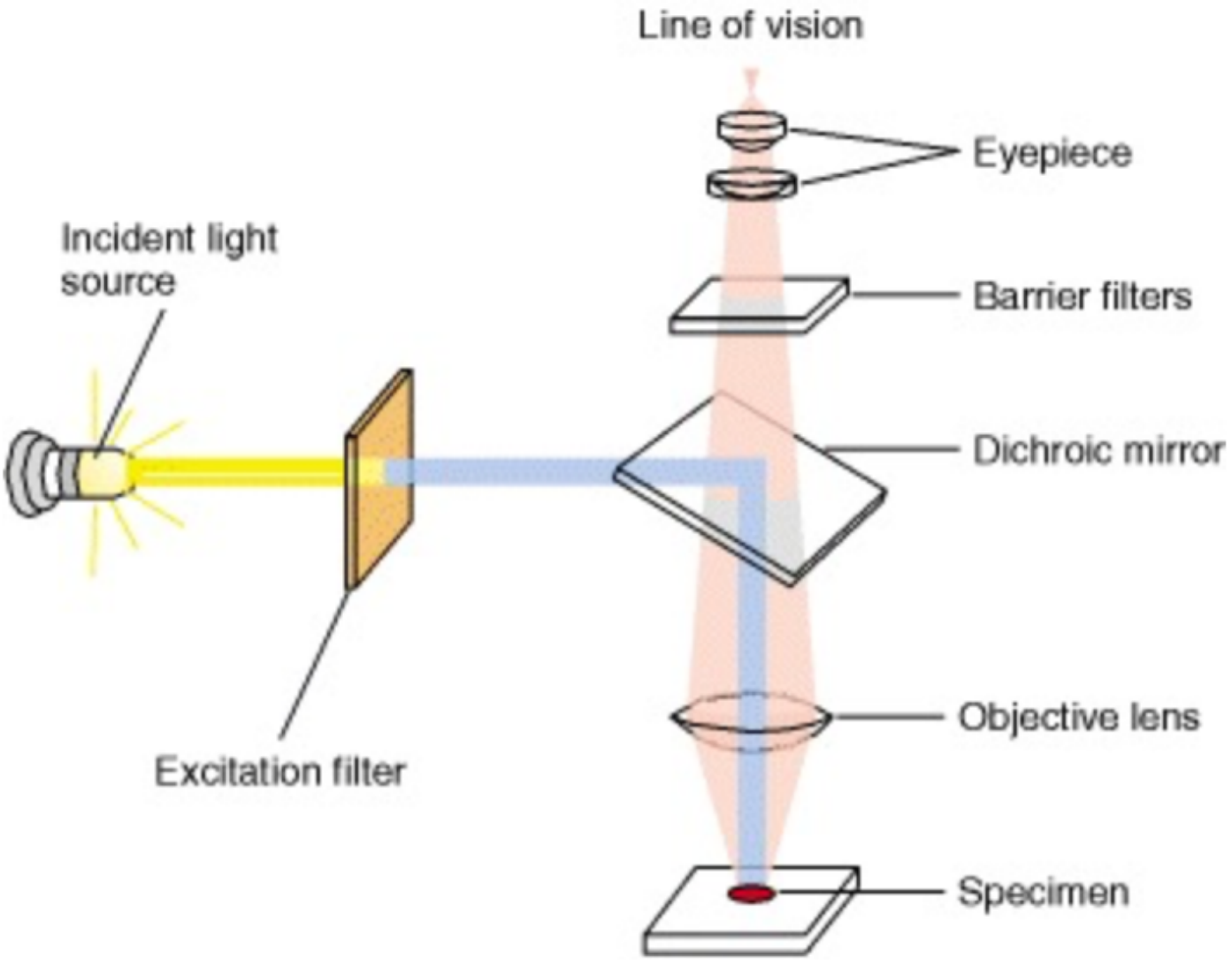
Dark field illumination, sample contrast comes from light scattered by the sample.



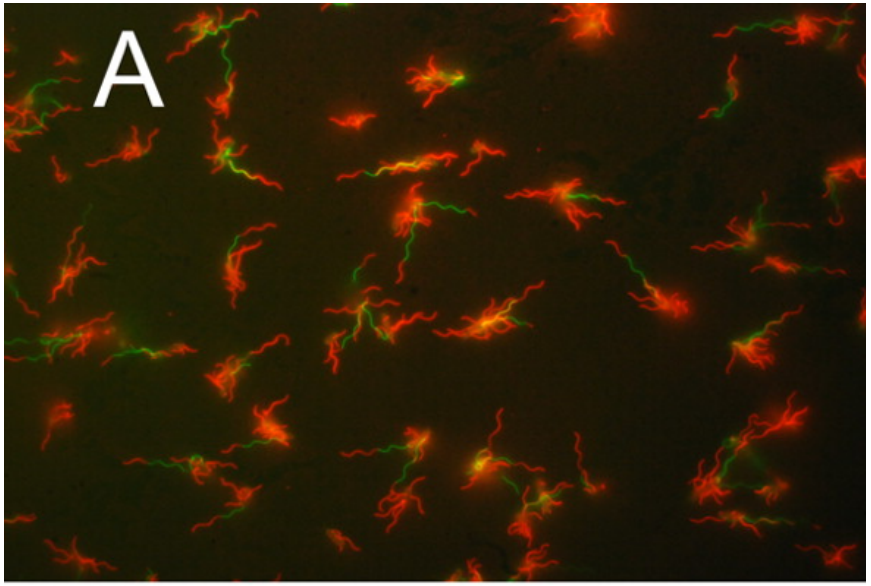
Phase contrast illumination, sample contrast comes from interference of different path lengths of light through the sample.

(from https://en.wikipedia.org/wiki/Microscopy)

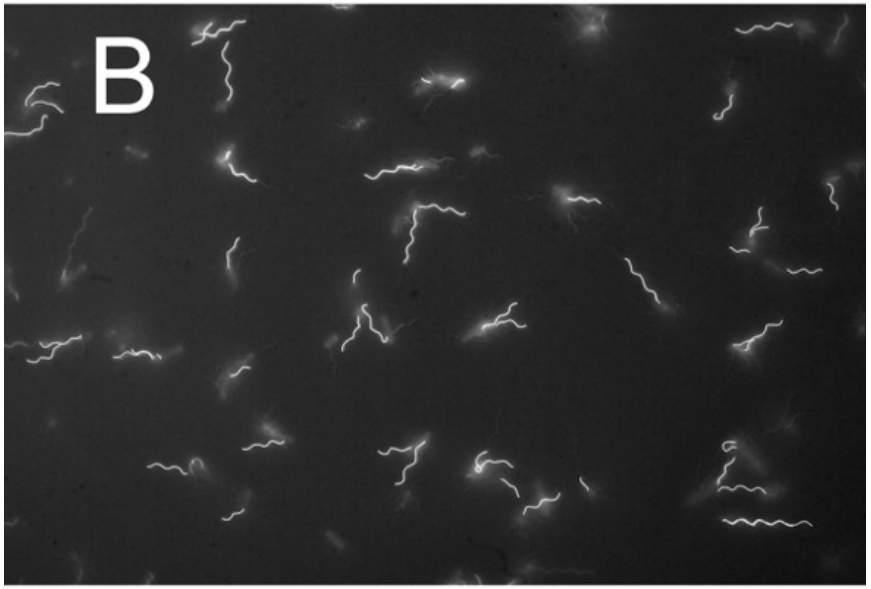
Fluorescence Microscopy



(from https://www.ncbi.nlm.nih.gov/books/NBK21629/)

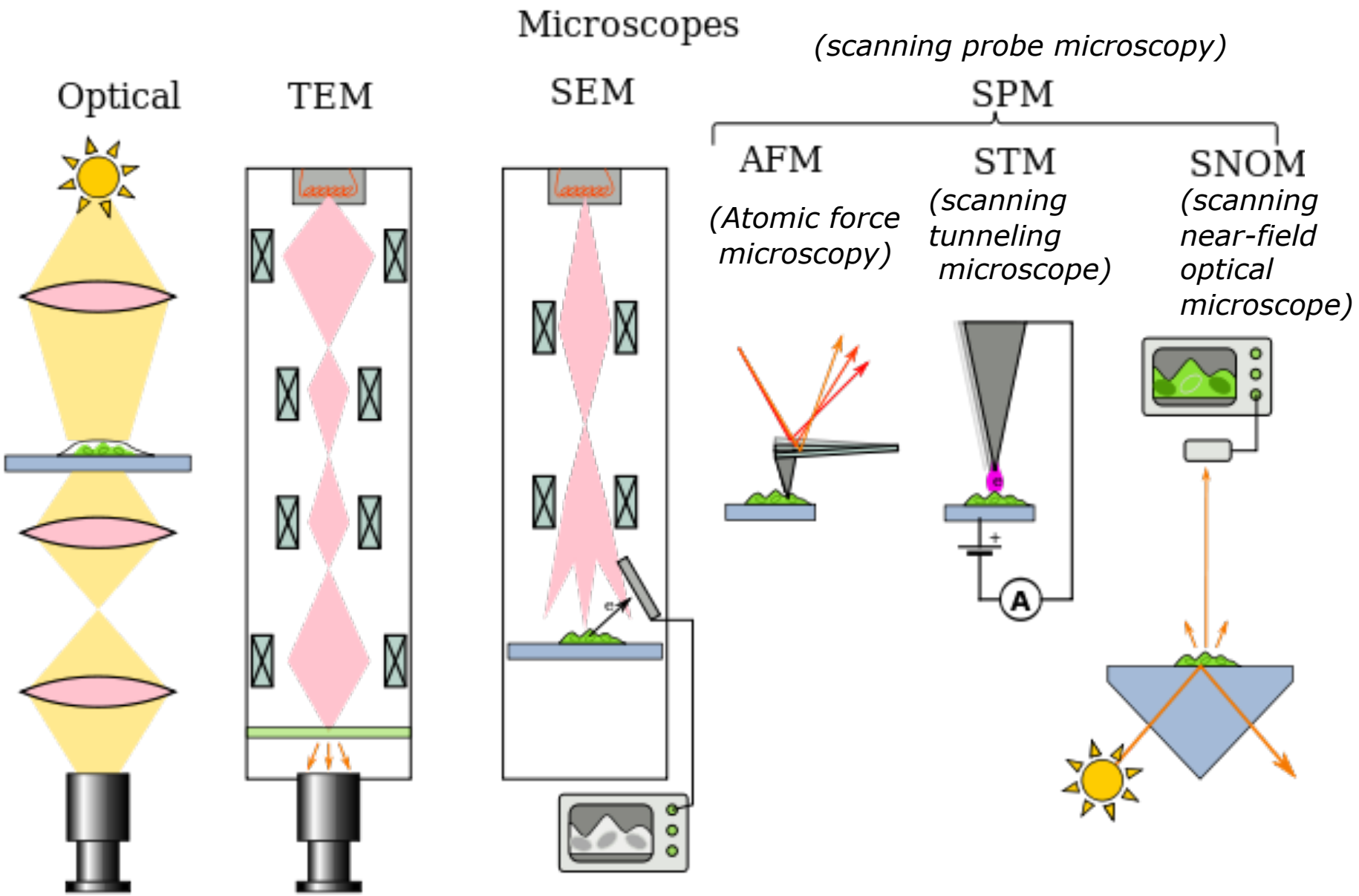


Flagellar filament (green) from Ecoli (red)



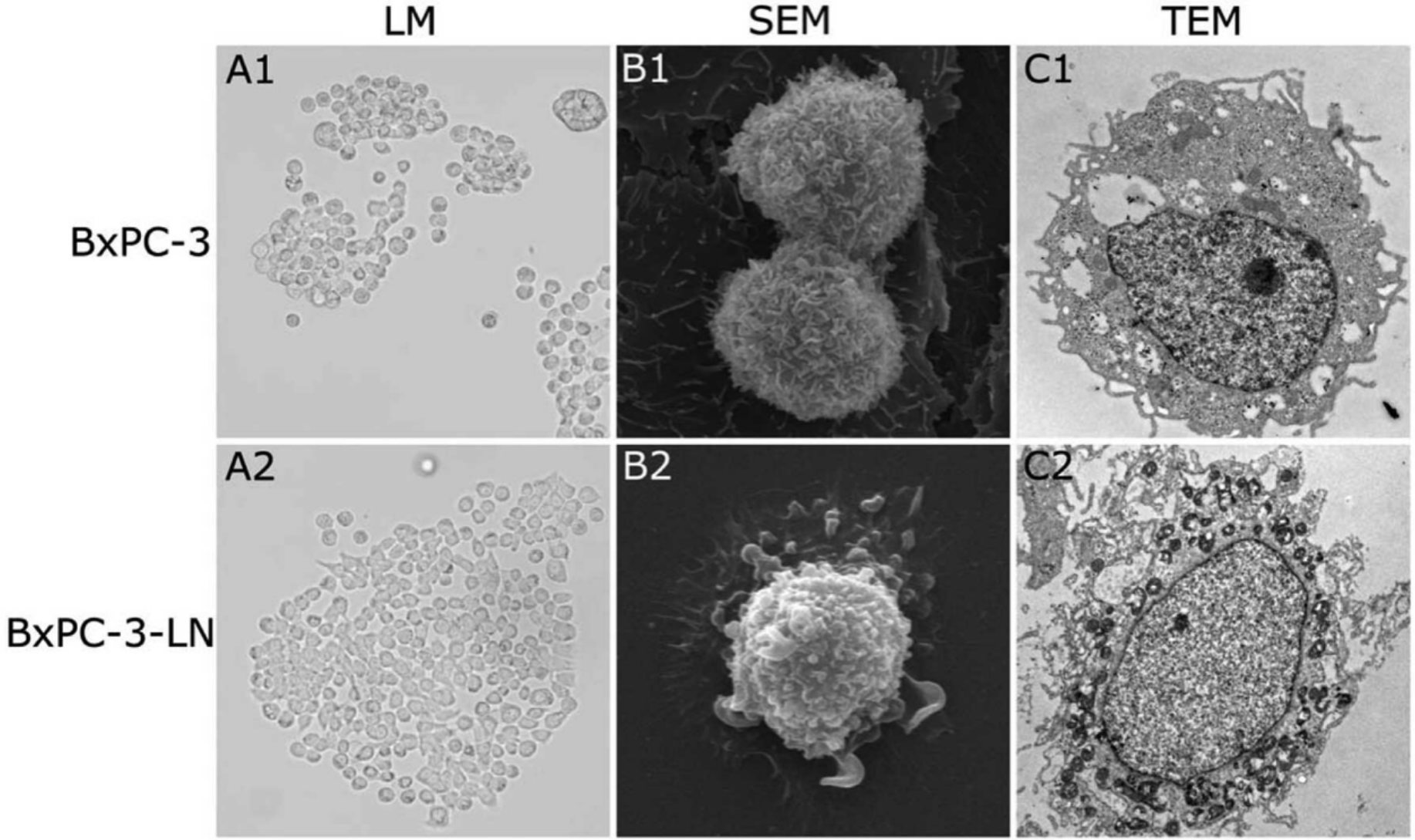
Same image, with green shown in white and other colors in black

(from http://jb.asm.org/content/194/10/2437/F1.expansion.html)



(from https://en.wikipedia.org/wiki/Microscope)

Differences between Light Microscopes and Electron Microscopes



(from https://www.spandidos-publications.com/10.3892/ijo.2012.1613)

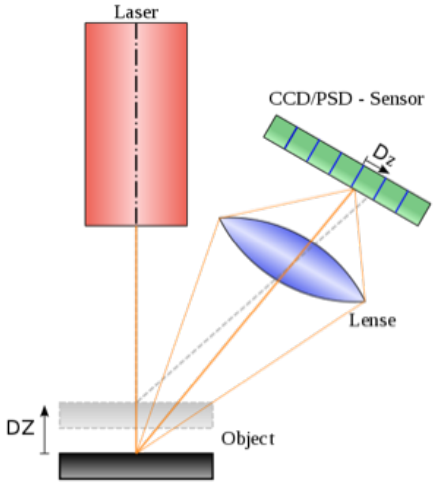
Imaging larger objects: scanners

1)Surface scanners

A) Contact scanners

B) Laser scanner: time of flight

C) Laser scanner: triangulation

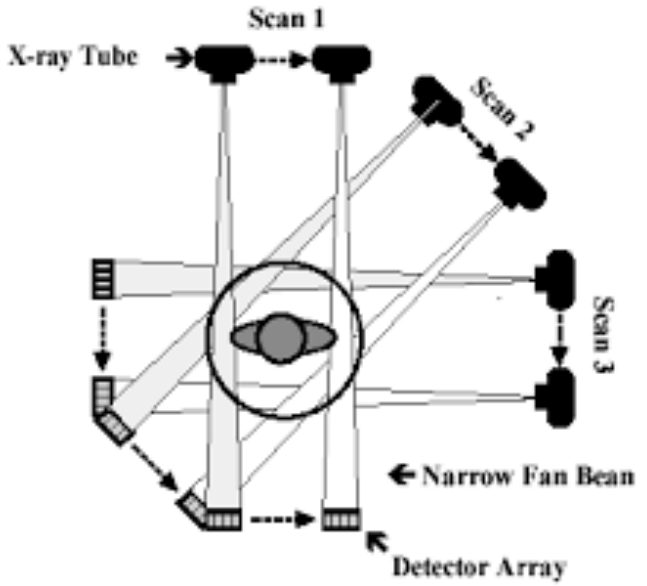


(from https://en.wikipedia.org/wiki/3D_scanner)

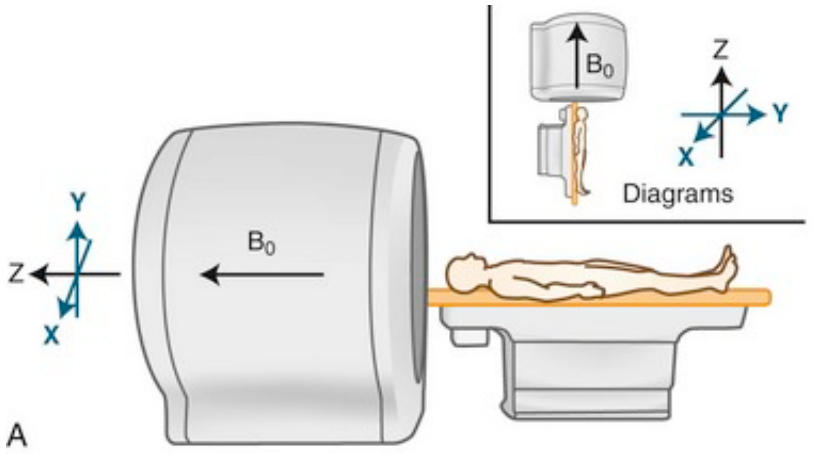
Imaging larger objects: scanners

1)Volumetric Scanners

X-ray computer tomography (CT scanners)



Magnetic Resonance Imaging (MRI)



Geometry and Topology in Biology

1)Phylogenetic trees

2)Geometry of Biomolecules

3)Morphometrics

Geometry and Topology in Biology

1)Phylogenetic trees

2)Geometry of Biomolecules

3)Morphometrics

Similarity: Homology vs Analogy

Homology: **Similarity in characteristics resulting from shared ancestry.**

Analogy: **The similarity of characteristics between two species that are not closely related; attributable to convergent evolution.**

Two sisters: homologs

Two "Elvis": analogs

Homology: Orthologs and Paralogs

Homology:
Similarity in characteristics resulting from shared ancestry.

Paralogy:

Homologous sequences are paralogous if they were separated by a gene duplication event

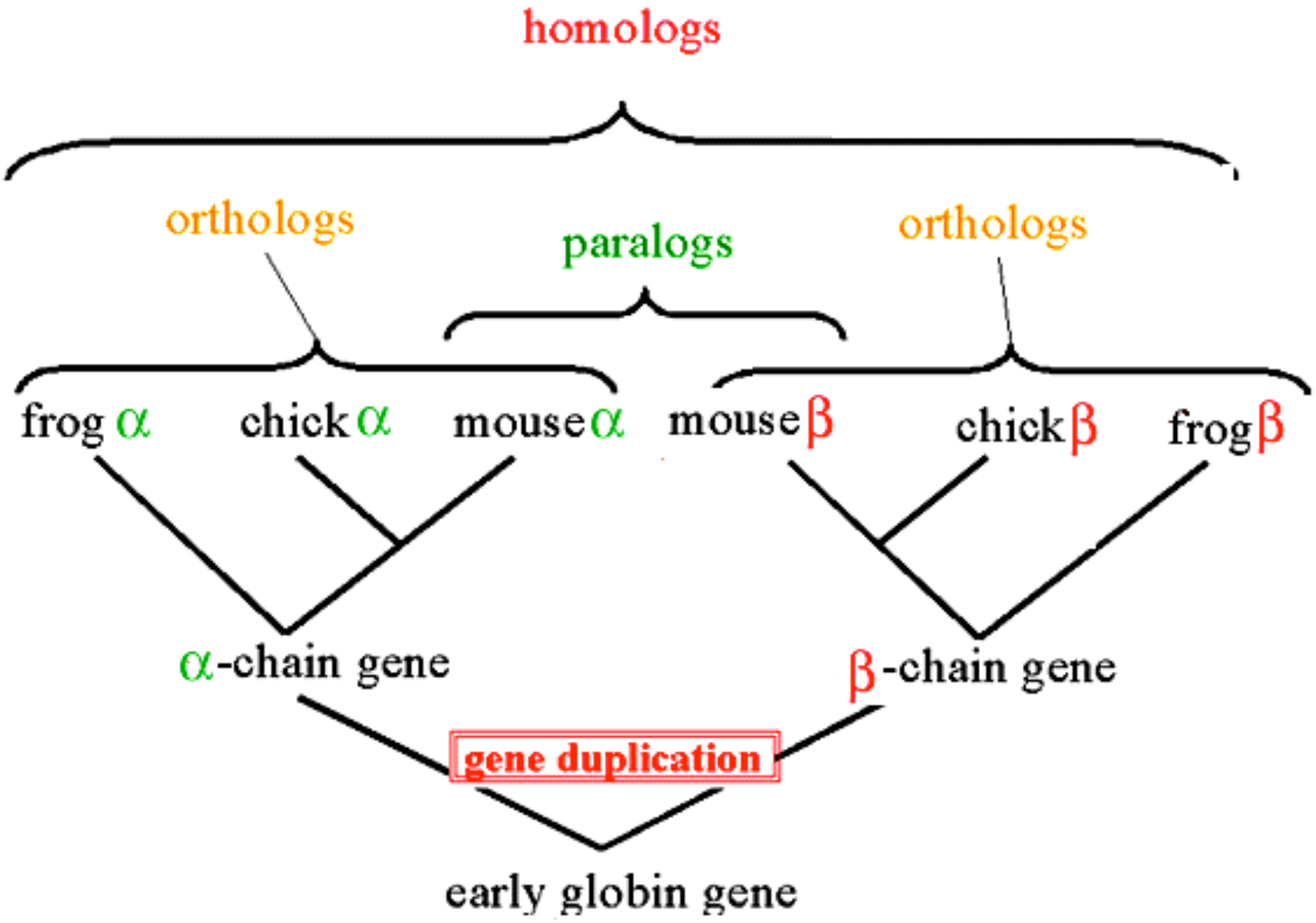
Orthology:

Homologous sequences are orthologous if they were separated by a speciation event

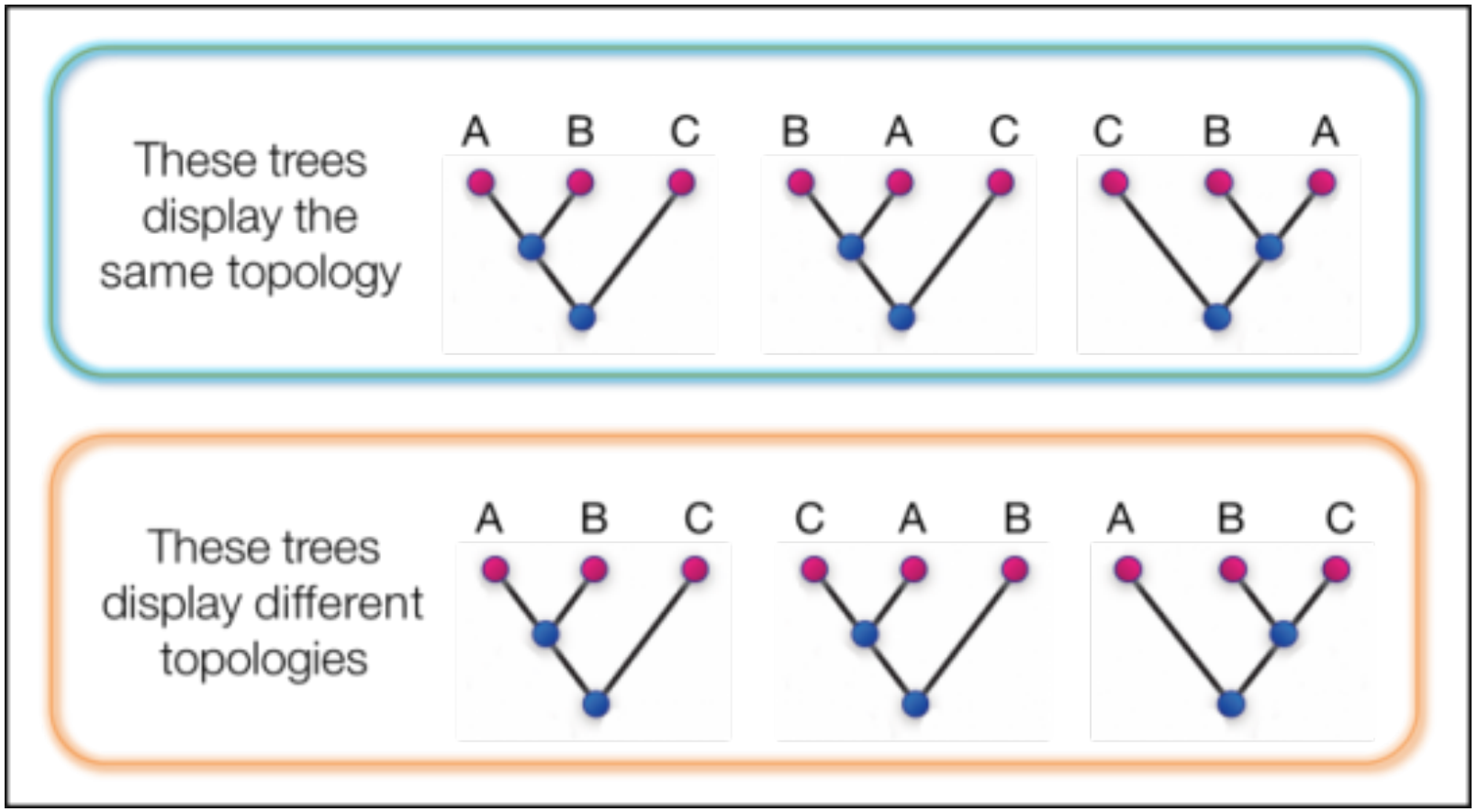
Further reading:

Koonin EV (2005). "Orthologs, paralogs, and evolutionary genomics". Annu. Rev. Genet. 39:309-338.

Homology: Orthologs and Paralogs



http://tolweb.org


These trees display the same topology
A B C
B A C
C B A
These trees display different topologies
A B C
C A B
A B C

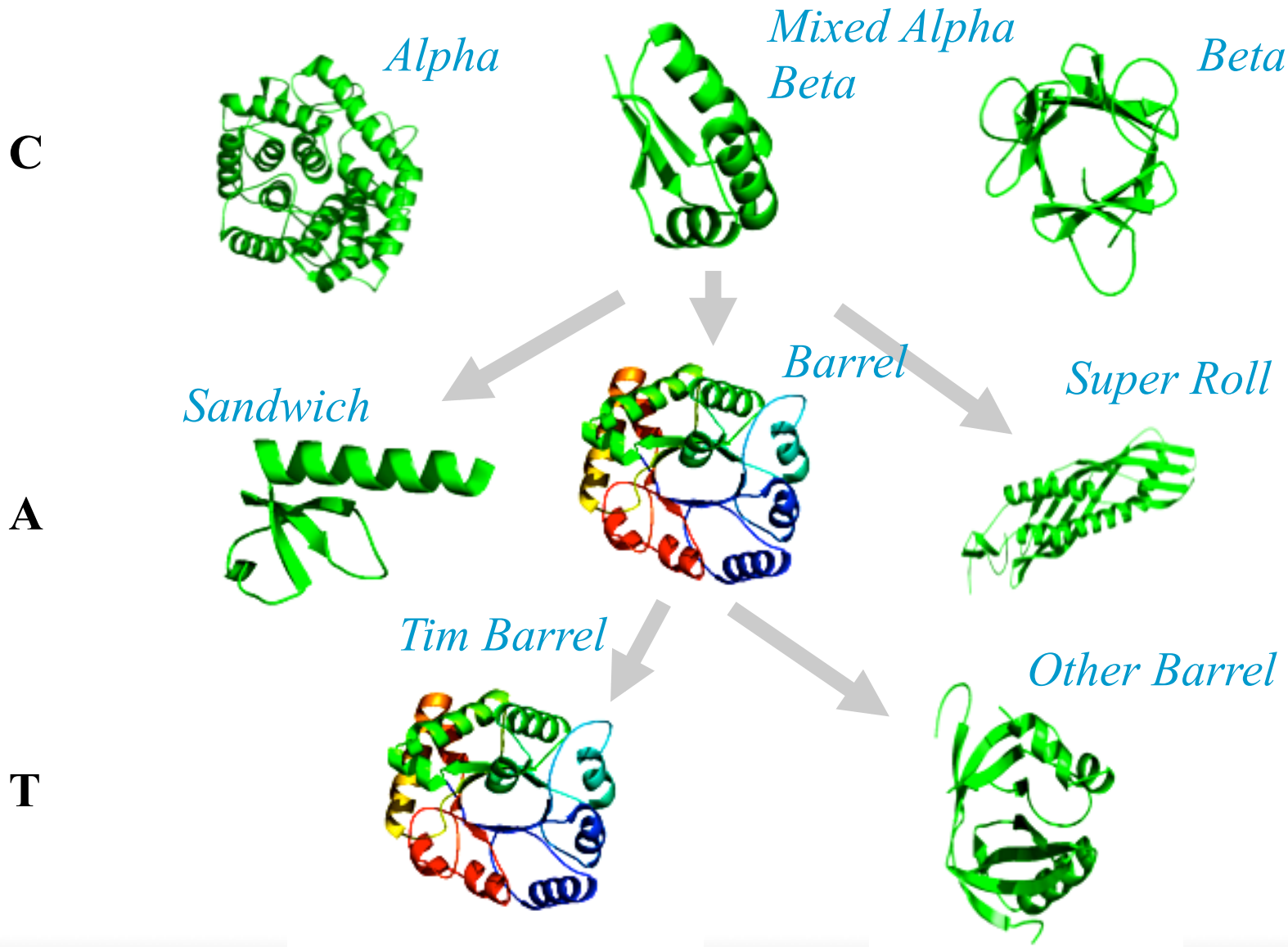
Geometry and Topology in Biology

1)Phylogenetic trees

2)Geometry of Biomolecules

3)Morphometrics

Protein Structures



Protein Structure Space

$$D = \begin{bmatrix} 0 & \dots & d_{1N} \\ \dots & 0 & \dots \\ d_{N1} & \dots & 0 \end{bmatrix} \rightarrow G = X^T X \rightarrow X$$

Distance Matrix

Metric Matrix

Points in Space

Protein Structure Comparison

Given two "shapes" or structures A and B, we are interested in defining a distance, or similarity measure between A and B.

- *Visual comparison*
- *Dihedral angle comparison*
- *Distance matrix*
- *RMSD (root mean square distance)*

Is the resulting distance (similarity measure) D a metric?

$$D(A,B) \leq D(A,C) + D(C,B)$$

Protein Structure Comparison

To compare two sets of points (atoms) $A=\{a_1, a_2, \ldots a_N\}$ and $B=\{b_1, b_2, \ldots, b_N\}$:

-Define a 1-to-1 correspondence between A and B

for example, a_i corresponds to b_i, for all i in [1,N]

-Compute RMS as:

$$RMS(A,B) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} d(a_i,b_i)^2}$$

$d(A_i,B_i)$ is the Euclidian distance between a_i and b_i.

Protein Structure Comparison

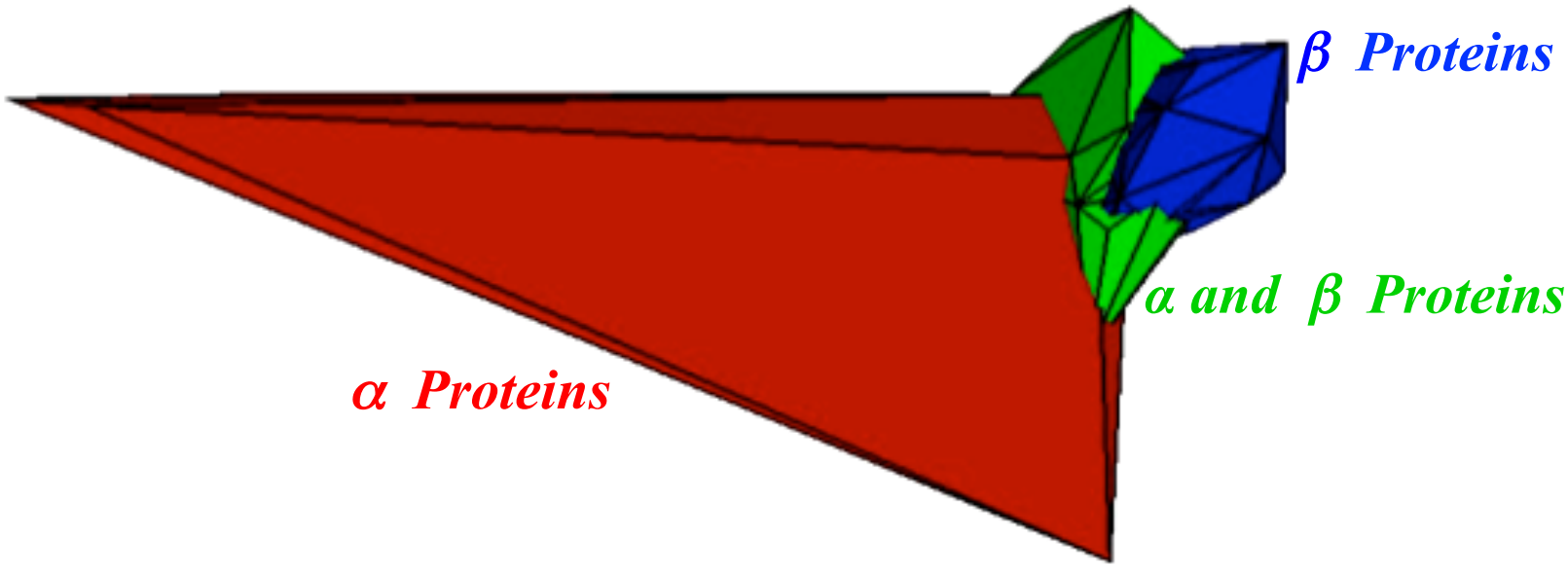
- Simplified problem: we know the correspondence between set A and set B
- We wish to compute the rigid transformation T that best align a_1 with b_1, a_2 with b_2, …, a_N with b_N
- The error to minimize is defined as:

Old problem, solved in Statistics, Robotics, Medical Image Analysis,

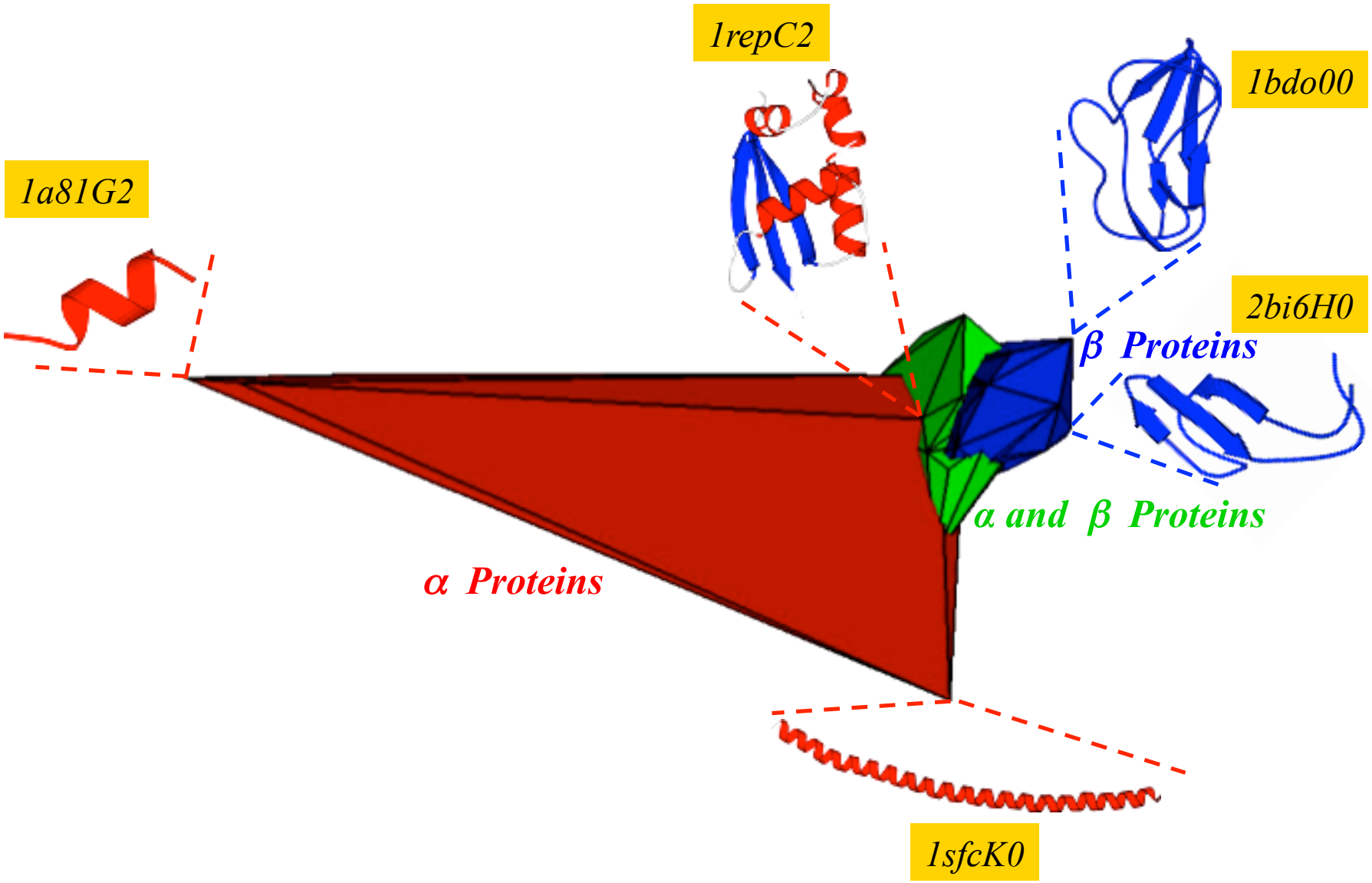
…

$$\varepsilon = \min_{T} \sum_{i=1}^{N} \left\| T(a_i) - b_i \right\|^2$$

A Picture of the Protein Structure Space



A Picture of the Protein Structure Space



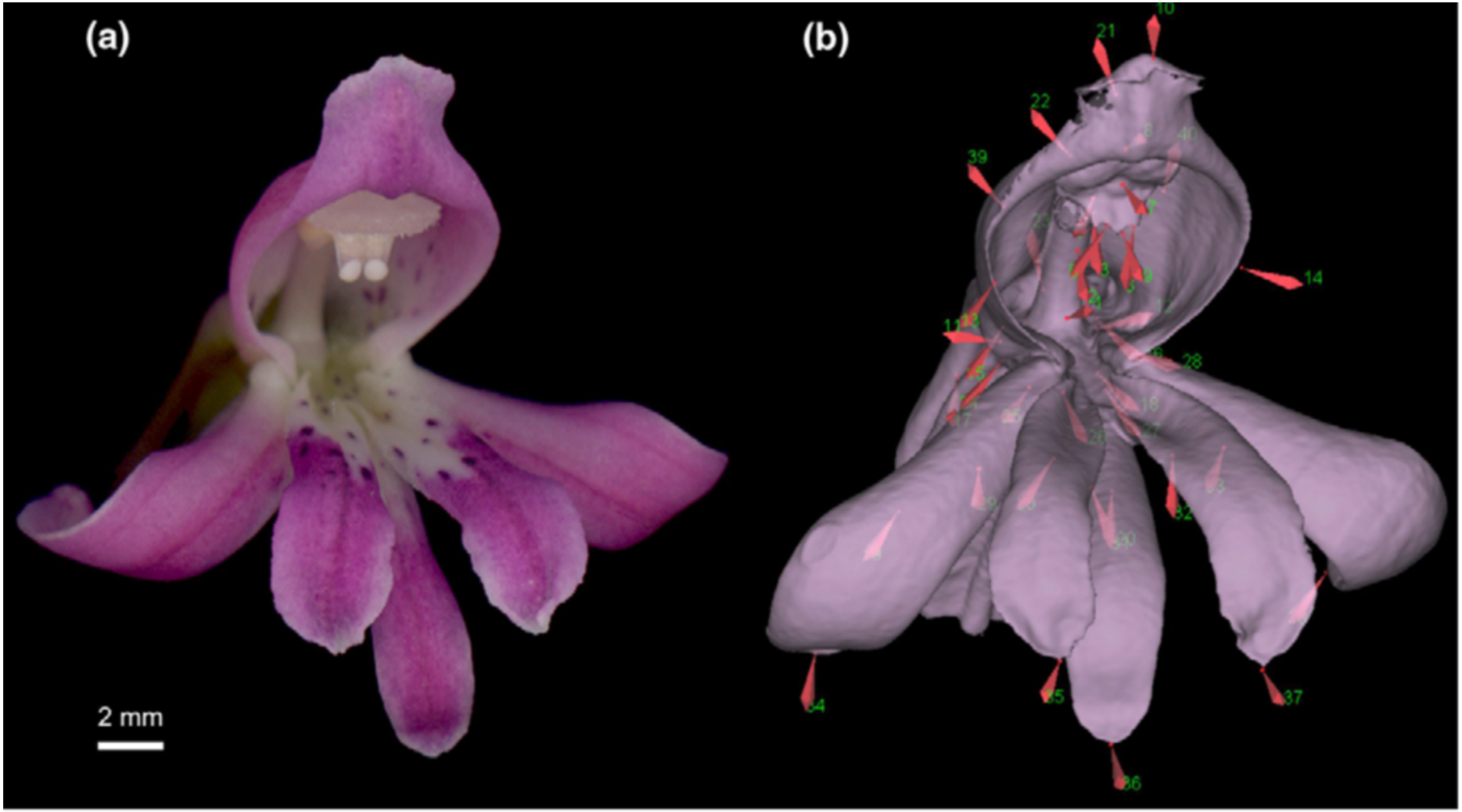
Geometry and Topology in Biology

1)Phylogenetic trees

2)Geometry of Biomolecules

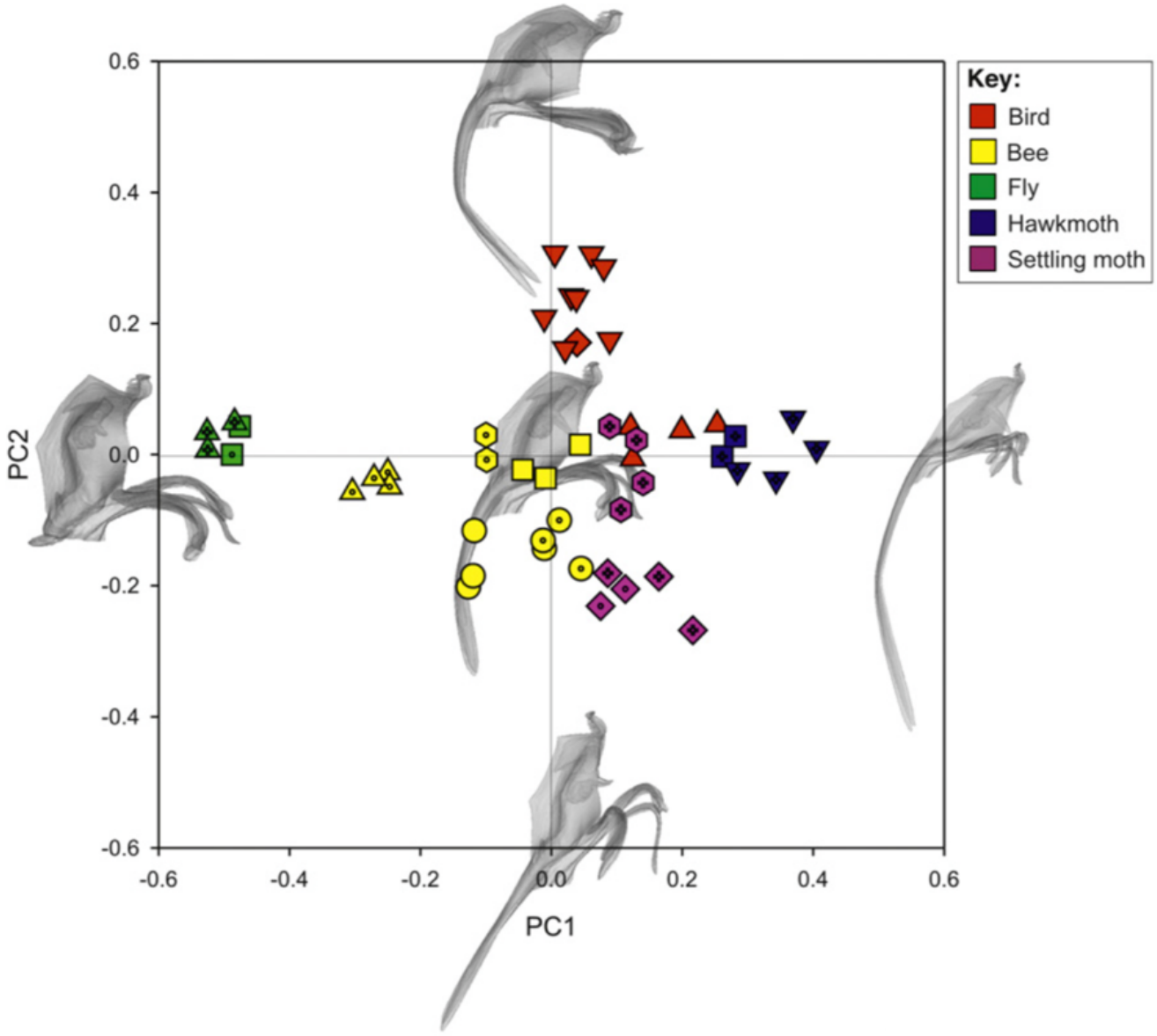
3)Morphometrics

3D Morphometrics of Orchids



(from http://www.sciencedirect.com/science/article/pii/S136013851000981)

3D Morphometrics of Orchids



(from http://www.sciencedirect.com/science/article/pii/S1360138510000981)