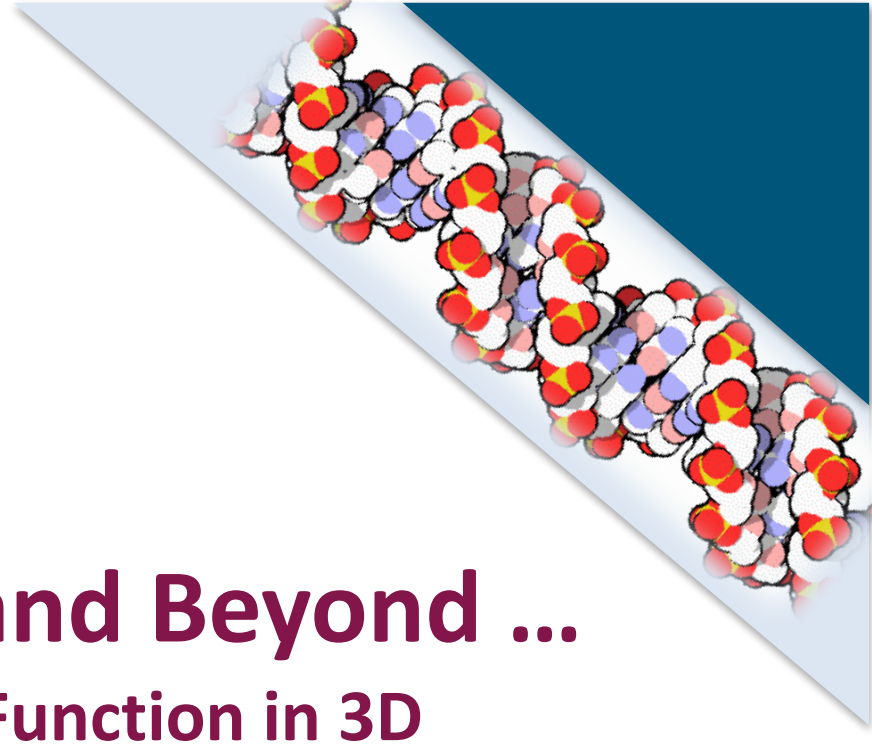
RCSB PDB
PROTEIN DATA BANK



# To the Double Helix and Beyond …

## Exploring DNA Structure and Function in 3D

Shuchismita Dutta, Ph.D.

Stephen K. Burley, M.D., D.Phil.

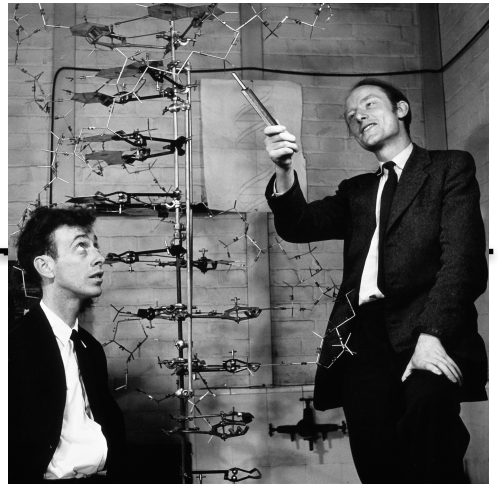
RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

UCSD SDSC SKAGGS SCHOOL of PHARMACY and PHARMACEUTICAL SCIENCES

July 24, 2017, Making Meaning through Modeling: Problem solving in Biology

# Learning Objectives

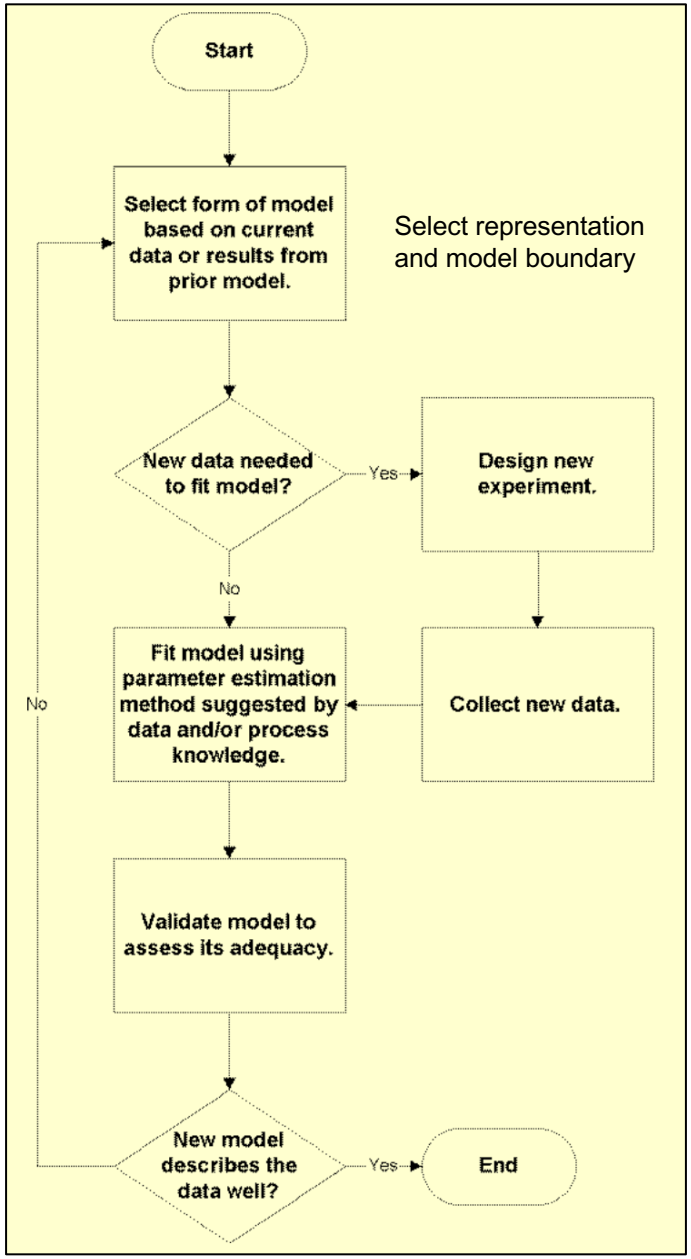
- Modeling in 3D
- Modeling the Double Helix
- Functions of the Double Helix
- Designing with the Double Helix



markuslibrary.rockefeller.edu/assets/image/DNA/12.JPG

# Modeling in 3D

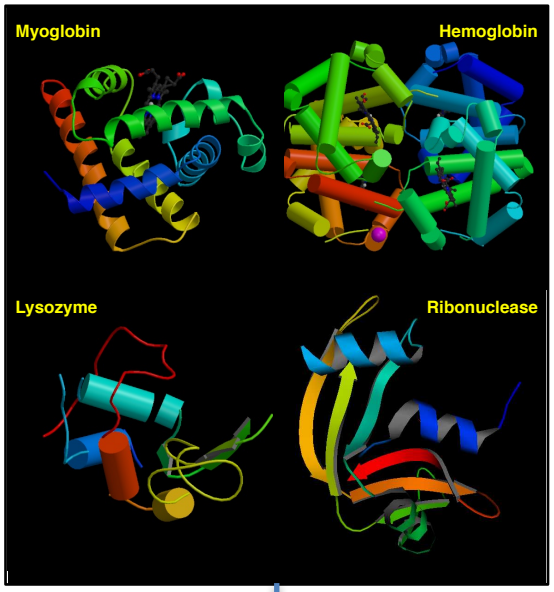
- **Visualize** structures at a molecular level
- **Understand** (model and/or interpret) different scenarios
- **Design** change(s) in the model to test and/or incorporate new properties

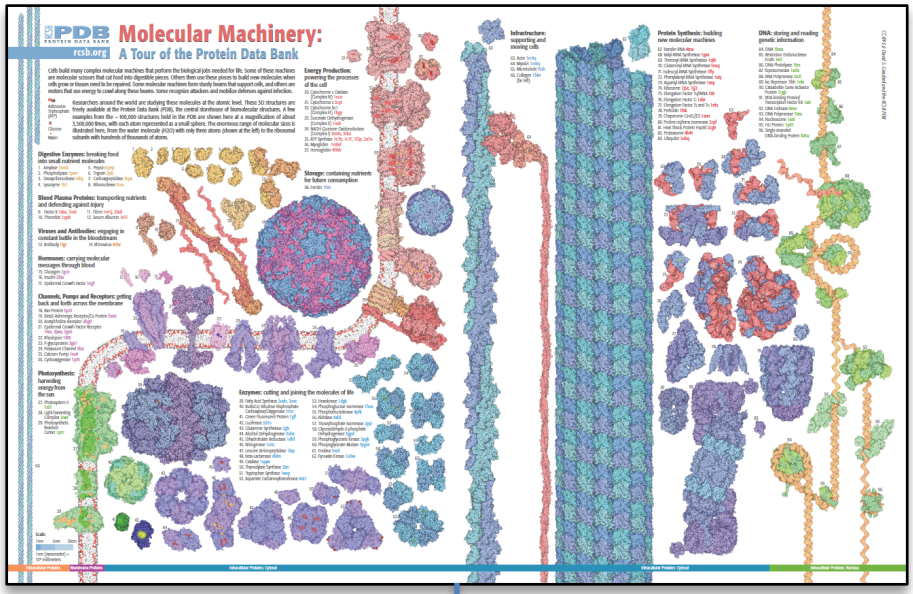



www.itl.nist.gov/div898/handbook/pmd/section4/pmd41.htm

# Protein Data Bank: History

1971, 7 structures




Current: >130,000 structures

1971 — 2003 — Current

WORLDWIDE PDB PROTEIN DATA BANK

RCSB PDB PROTEIN DATA BANK

PDBe Protein Data Bank in Europe

PDBj Protein Data Bank Japan

BMRB

# RCSB PDB: Bicoastal Organization

- Managed since 1999 by RUTGERS / UC San Diego
- Collaborate with Worldwide PDB to serve more than 1 million Data Producers/Consumers
- Support User query, analysis, visualization, and download of PDB data without usage restrictions
- Global reach to diverse stakeholder communities
- Funded jointly by NSF, NIH, DOE

# What is Archived in the PDB?

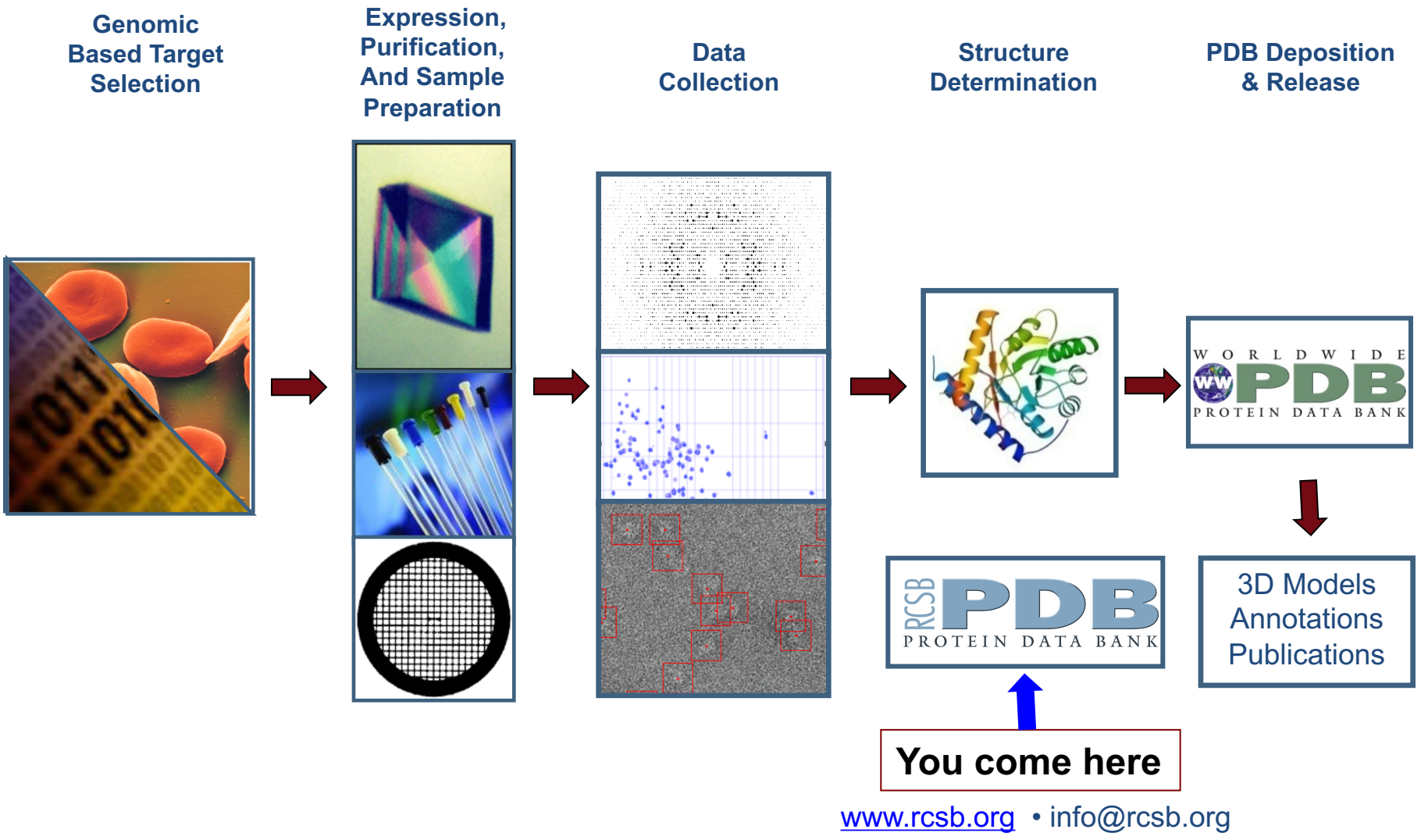
- Coordinates and experimental data
- Sample preparation, data collection and structure solution details
- Sequence(s) of polymers (proteins and nucleic acids) in the structure
- Information about ligands in the structure
- Links to various resources that describe sequence, function, and other properties of the molecule.
- Classification of structures by sequence, structure, function and other criteria

```
HEADER    OXYGEN TRANSPORT                        07-MAR-84   4HHB
TITLE     THE CRYSTAL STRUCTURE OF HUMAN DEOXYHAEMOGLOBIN AT 1.74 ANGSTROMS
TITLE    2 RESOLUTION
COMPND     MOL_ID: 1;
COMPND    2 MOLECULE: HEMOGLOBIN (DEOXY) (ALPHA CHAIN);
COMPND    3 CHAIN: A, C;
COMPND    4 ENGINEERED: YES;
COMPND    5 MOL_ID: 2;
COMPND    6 MOLECULE: HEMOGLOBIN (DEOXY) (BETA CHAIN);
COMPND    7 CHAIN: B, D;
COMPND    8 ENGINEERED: YES
SOURCE     MOL_ID: 1;
SOURCE    2 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE    3 ORGANISM_COMMON: HUMAN;
SOURCE    4 ORGANISM_TAXID: 9606;
SOURCE    5 MOL_ID: 2;
SOURCE    6 ORGANISM_SCIENTIFIC: HOMO SAPIENS;
SOURCE    7 ORGANISM_COMMON: HUMAN;
SOURCE    8 ORGANISM_TAXID: 9606
KEYWDS     OXYGEN TRANSPORT
EXPDTA     X-RAY DIFFRACTION
AUTHOR     G.FERMI,M.F.PERUTZ
REVDAT    5   13-JUL-11 4HHB     1       VERSN
REVDAT    4   24-FEB-09 4HHB     1       VERSN
REVDAT    3   01-APR-03 4HHB     1       JRNL
REVDAT    2   15-OCT-89 4HHB     3       MTRIX
REVDAT    1   17-JUL-84 4HHB     0
SPRSDE     17-JUL-84 4HHB      1HHB
JRNL        AUTH   G.FERMI,M.F.PERUTZ,B.SHAANAN,R.FOURME
JRNL        TITL   THE CRYSTAL STRUCTURE OF HUMAN DEOXYHAEMOGLOBIN AT 1.74 A
JRNL        TITL 2 RESOLUTION
JRNL        REF    J.MOL.BIOL.                   V. 175   159 1984
JRNL        REFN                     ISSN 0022-2836
JRNL        PMID   6726807
-snip-
CRYST1   63.150   83.590   53.800  90.00  99.34  90.00 P 1 21 1      4
ORIGX1      0.963457  0.136613  0.230424       16.61000
ORIGX2     -0.158977  0.983924  0.081383       13.72000
ORIGX3     -0.215598 -0.115048  0.969683       37.65000
SCALE1      0.015462  0.002192  0.003698        0.26656
SCALE2     -0.001902  0.011771  0.000974        0.16413
SCALE3     -0.001062 -0.001721  0.018728        0.75059
MTRIX1   1 -1.000000  0.000000  0.000000        0.00001    1
MTRIX2   1  0.000000  1.000000  0.000000        0.00002    1
MTRIX3   1  0.000000  0.000000 -1.000000        0.00002    1
ATOM      1  N   VAL A   1       6.204  16.869   4.854  1.00 49.05           N
ATOM      2  CA  VAL A   1       6.913  17.759   4.607  1.00 43.14           C
ATOM      3  C   VAL A   1       8.504  17.378   4.797  1.00 24.80           C
ATOM      4  O   VAL A   1       8.805  17.011   5.943  1.00 37.68           O
ATOM      5  CB  VAL A   1       6.369  19.044   5.810  1.00 72.12           C
ATOM      6  CG1 VAL A   1       7.009  20.127   5.418  1.00 61.79           C
ATOM      7  CG2 VAL A   1       5.246  18.533   5.681  1.00 80.12           C
ATOM      8  N   LEU A   2       9.096  18.040   3.857  1.00 26.44           N
```

Header-Meta data

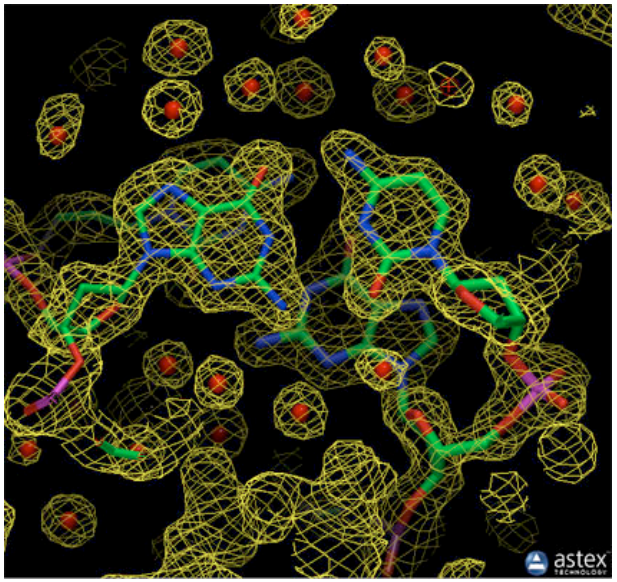
Coordinate data

# The Data Pipeline

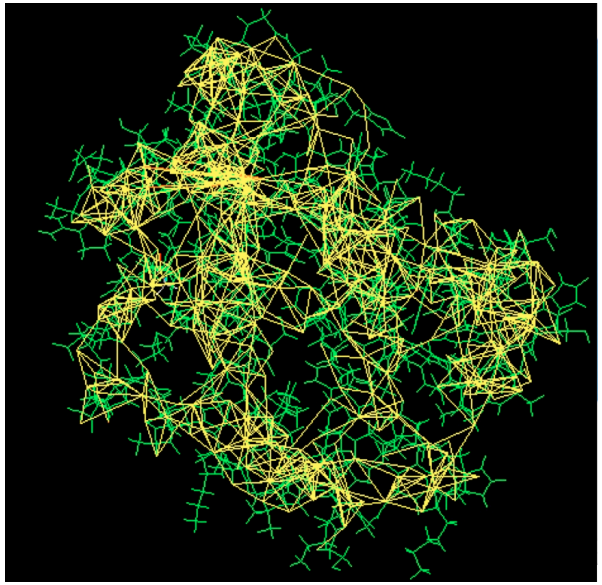



www.rcsb.org • info@rcsb.org

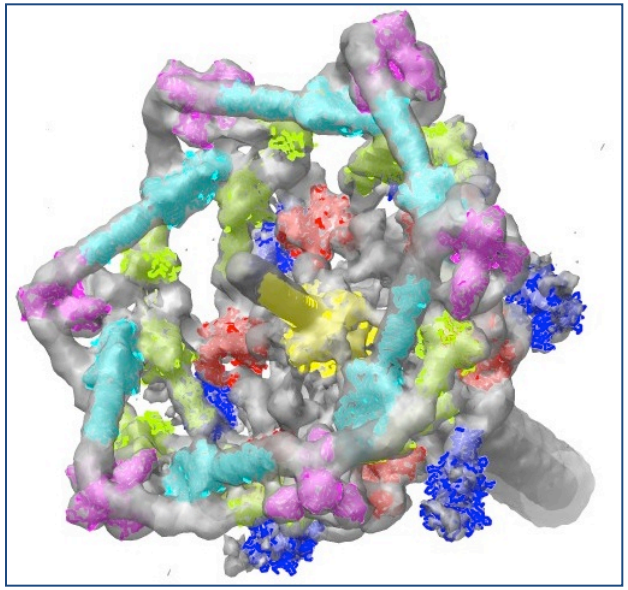
# Building X-ray, NMR, EM Models




**X-ray Crystallography**: Electron density from a structure of DNA is shown here (PDB entry 196d), along with the atomic model



**NMR**: Restraints used to solve the structure of a small monomeric hemoglobin. The protein (1vre and 1vrf) is shown in green, and restraints are shown in yellow.



**EM**: Tail of the T4 bacteriophage. Surface rendering of the EM data (emd-1048) with atomic coordinates from PDB entries 1pdf, 1pdi, 1pdl, 1pdm, 1pdp, and 2fl8.

# RCSB PDB Portal (rcsb.org)

RCSB PDB PROTEIN DATA BANK

An Information Portal to
130997 Biological
Macromolecular Structures

Search by PDB ID, author, macromolecule, sequence, or ligands

Go

Advanced Search | Browse by Annotations

Query

PDB-101 | Worldwide PDB | EMDataBank | NDB Nucleic Acid Database | StructuralBiology Knowledgebase | Worldwide Protein Data Bank Foundation

- Welcome
- Deposit
- Query
- Visualize
- Analyze
- Download
- Learn

## A Structural View of Biology

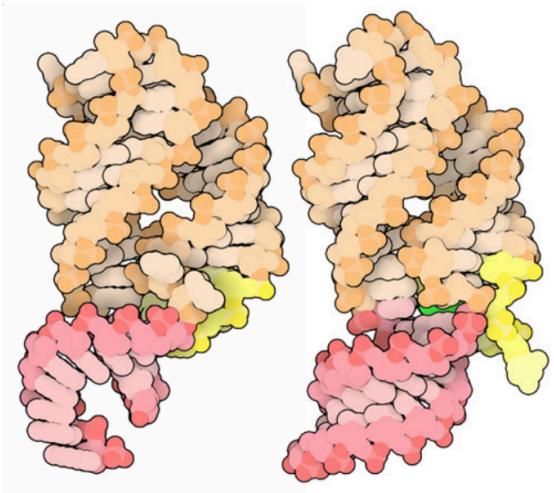
This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

As a member of the wwPDB, the RCSB PDB curates and annotates PDB data.

The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

### Zika Illustration Named People's Choice


### June Molecule of the Month



Adenine Riboswitch in Action

# PDB-101 (pdb101.rcsb.org)

Query

PDB-101 | Molecule of the Month | Browse | Learn | Global Health | Teach | Geis Archive | Events | More

RCSB PDB-101

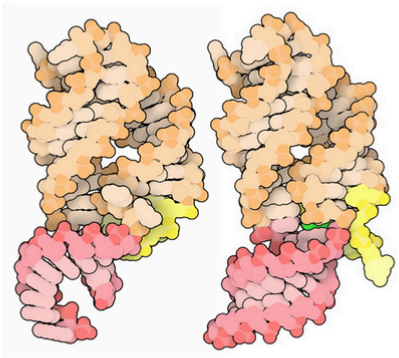
Molecular explorations through biology and medicine

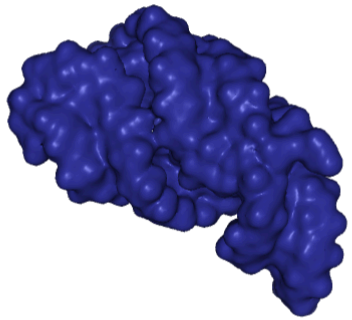
Educational portal of RCSB PDB PROTEIN DATA BANK

Search Molecule of the Month articles and more | Go

## Molecule of the Month

June 2017





### Adenine Riboswitch in Action

XFEL serial crystallography reveals what happens when adenine binds to a riboswitch

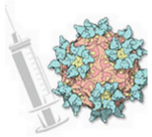
More

3D View: 5SWE

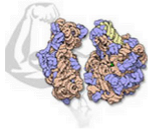
| Style | Color | Spin |
|---|---|---|
| ○ Cartoon | ○ Rainbow | ● On |
| ○ Spheres | ● Chain | ○ Off |
| ● Surface | ○ Structure | |

All articles: By Date | By Category | By Title

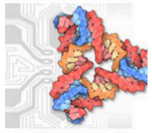
## Browse resources by category



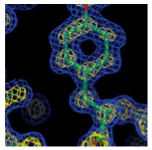
Health and Disease



Molecules of Life



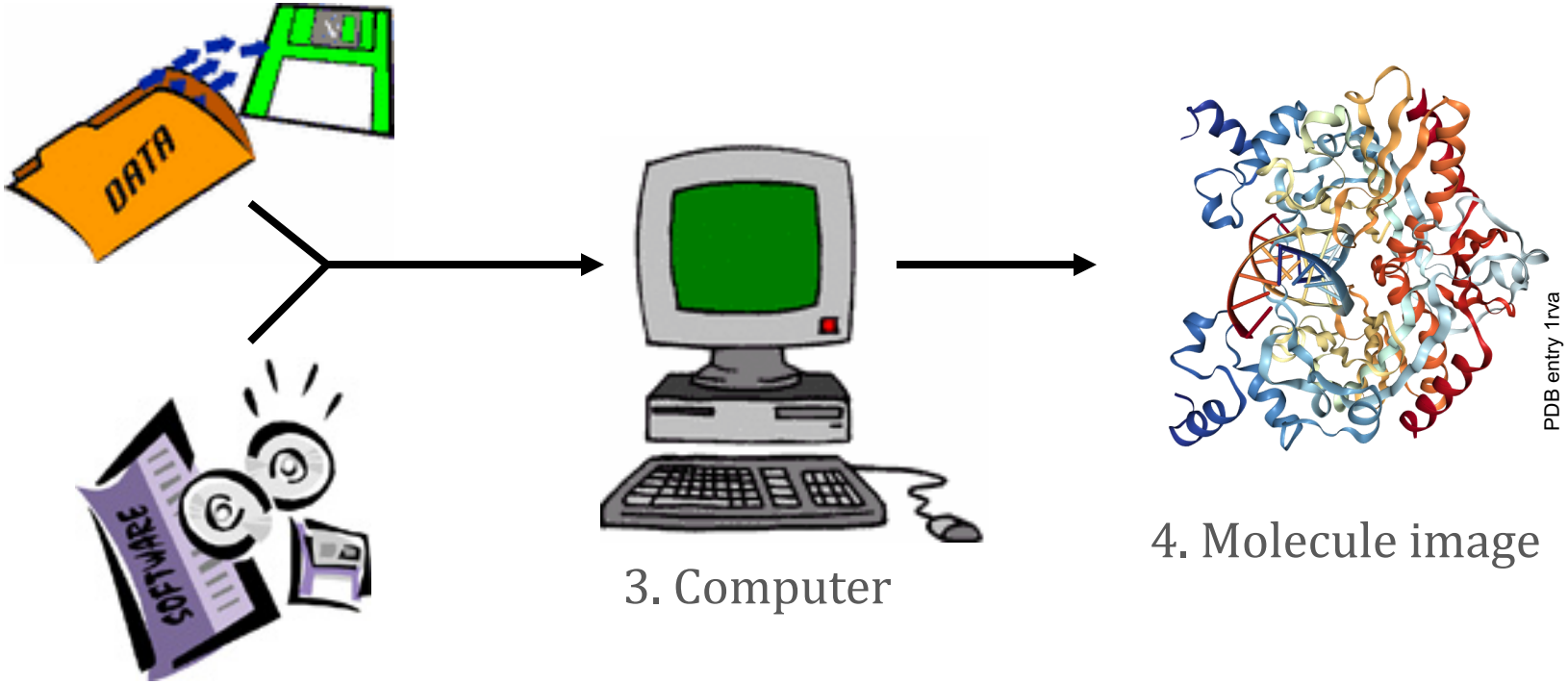
Biotech and Nanotech



Structures and Structure Determination

# Visualizing Molecules on a Device Screen

1. Coordinate file from PDB



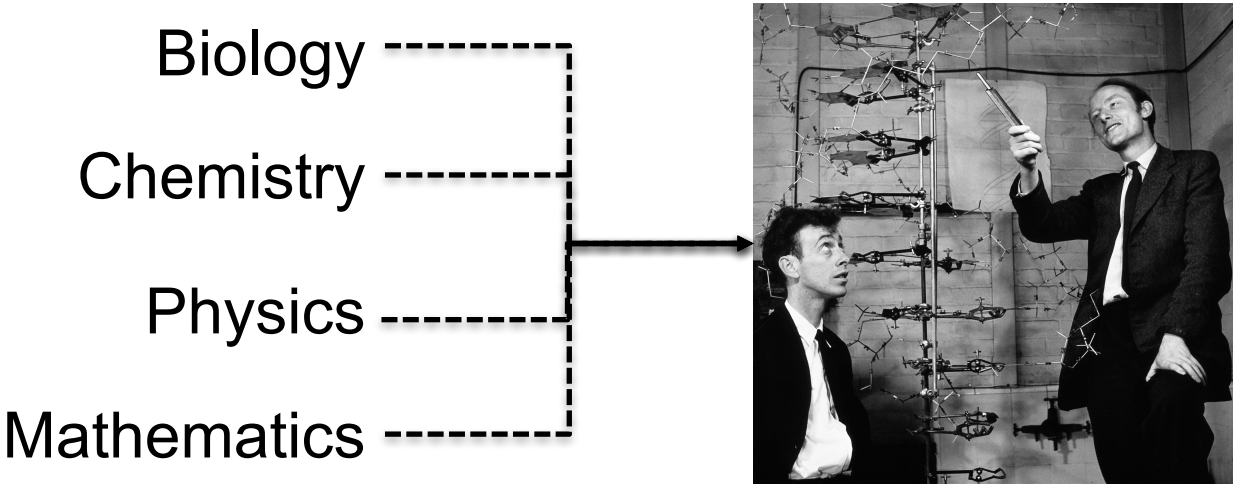

3. Computer

4. Molecule image

2. Visualization software
*e.g.*, RasMol, Chimera, NGL,
Swiss PDB Viewer

# Learning Objectives

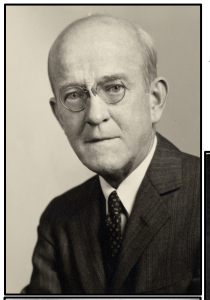
- Modeling in 3D
- Modeling the Double Helix
- Functions of the Double Helix
- Designing with the Double Helix



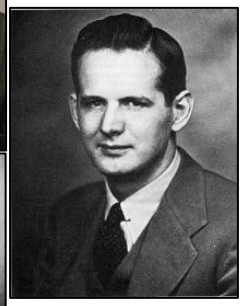

markuslibrary.rockefeller.edu/assets/image/DNA/12.JPG

# DNA: The Transforming Principle

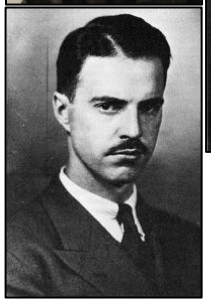
- Avery, McLeod, and McCarty experiment (1944)
- *Streptococcus pneumoniae* pass DNA among themselves
  - Rough colonies = no polysaccharide coat → non-pathogenic
  - Smooth colonies = polysaccharide coat → lethal!
  - Difference → plasmid encoded genes



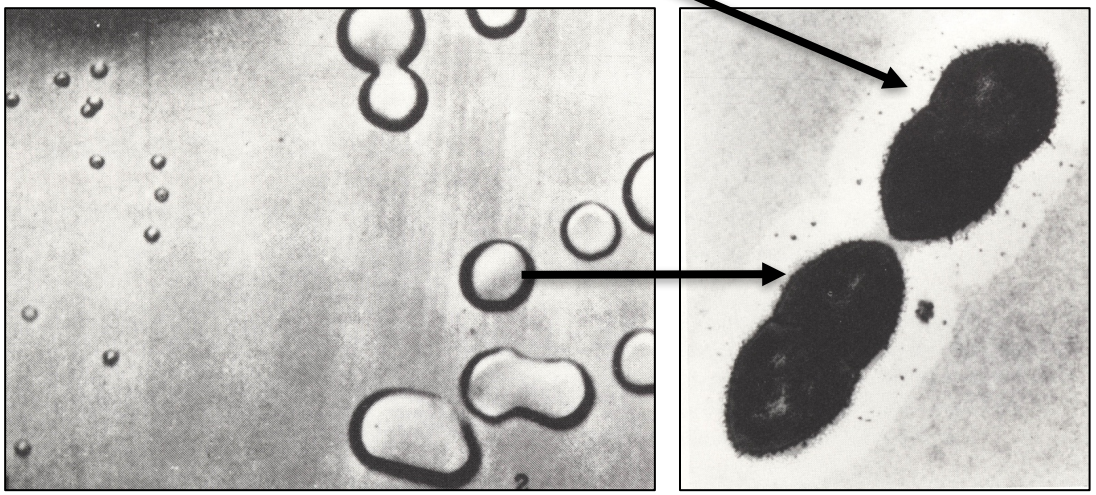
Avery



McCarty



McLeod



http://markuslibrary.rockefeller.edu/events_exhibits?page=events_exhibits_dna

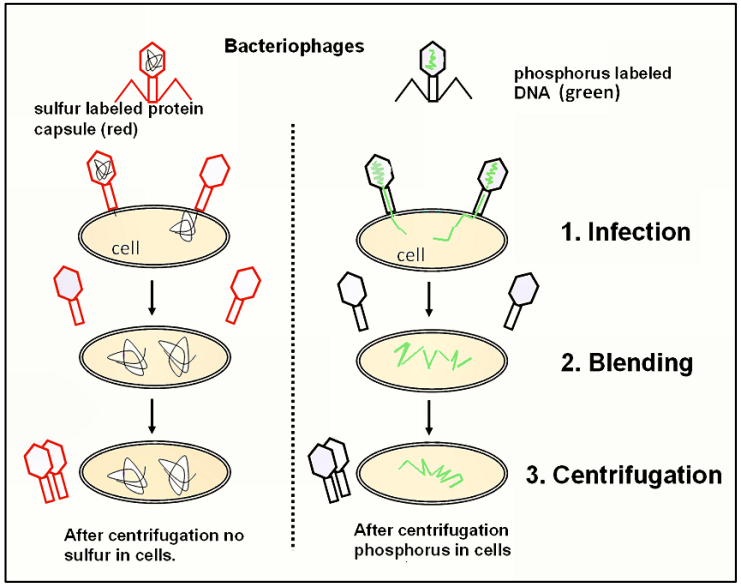
https://profiles.nlm.nih.gov/ps/retrieve/narrative/cc/p-nid/158/p-visuals/true

# DNA: The Transforming Principle – Cont.

- Implications of the Avery, McLeod, McCarty experiment → not widely accepted in 1944
- In 1952, Alfred Hershey and Martha Chase used $^{35}S$ and $^{32}P$ isotope laden bacterial DNA viruses

Hershey and Chase

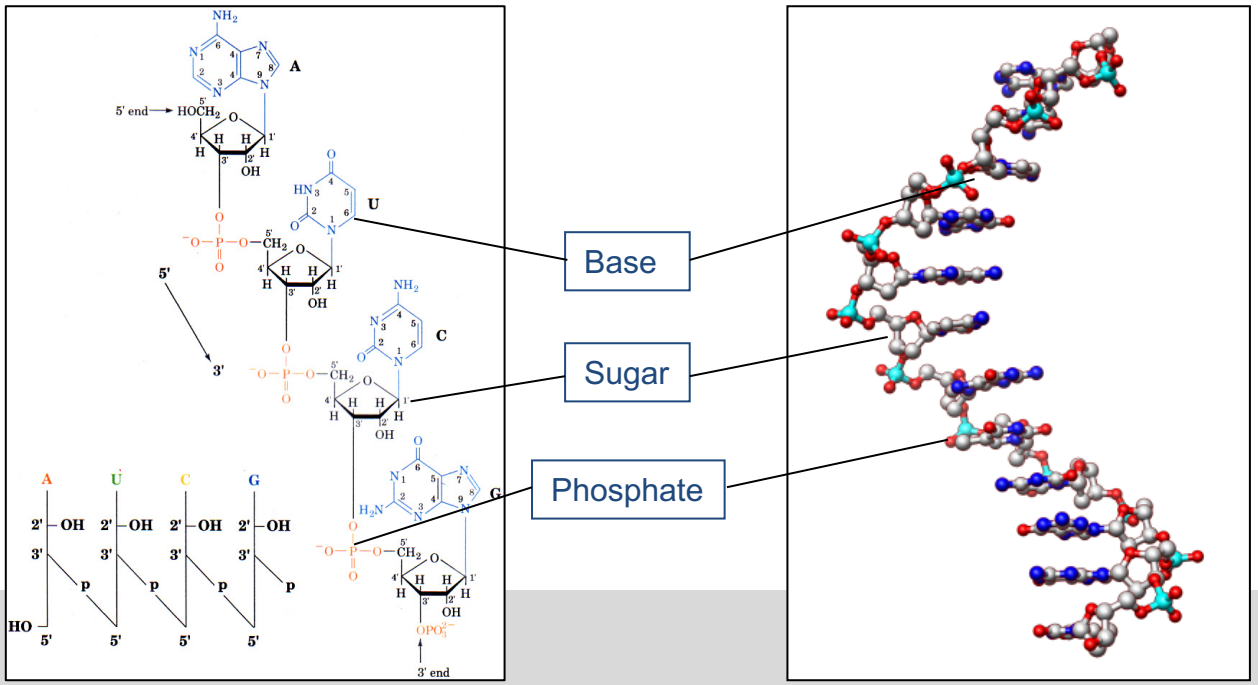
labdish.cshl.edu/2016/06/15/base-pairs-episode-1-from-phages-to-faces/




https://en.wikipedia.org/wiki/Hershey%E2%80%93Chase_experiment#/media/File:Hershey_Chase_experiment.png

# DNA: The Chemical Composition

- DNA is a polymer composed of a sugar phosphate backbone and bases
- There are two types of bases –
  - Purines (Adenine or A, and Guanine or G)
  - Pyrimidines (Thymine or T, and Cytosine or C)

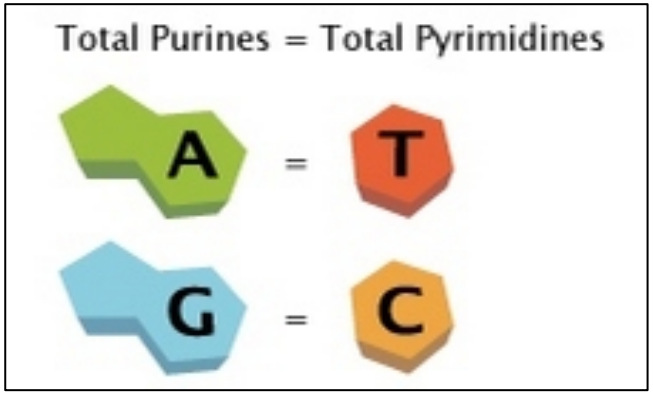


# DNA: The Chemical Composition – Cont.

- Chargaff's Rules:
  1. Amounts of A= T and C= G
  2. Ratio of A/T *vs* C/G varies among different organisms (makes sense for heredity)
- Chargaff met with Watson and Crick in 1952

Chargaff

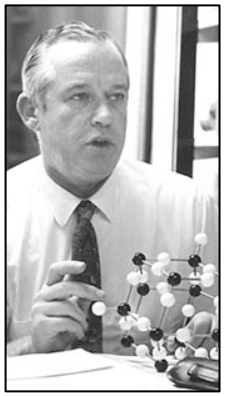
markuslibrary.rockefeller.edu /assets/image/DNA/11.JPG




www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397

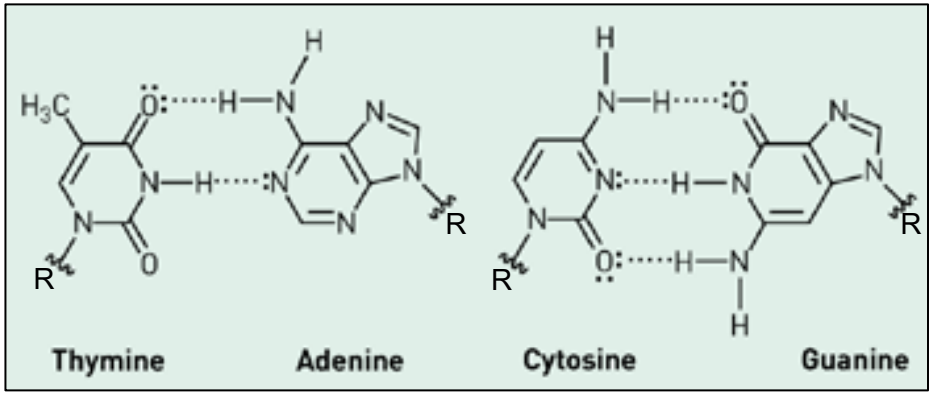
# DNA: Base Pairing in Keto *vs.* Enol

- Jerry Donohue shared office with Watson and Crick
- Keto form of bases is dominant (based on Furberg's cytidine 1951 X-ray structure)
- Explained isosteric base pairing – R---R distance for A=T and G≡C is the same → molecular basis for Chargaff's Rule No. 1)



Donohue

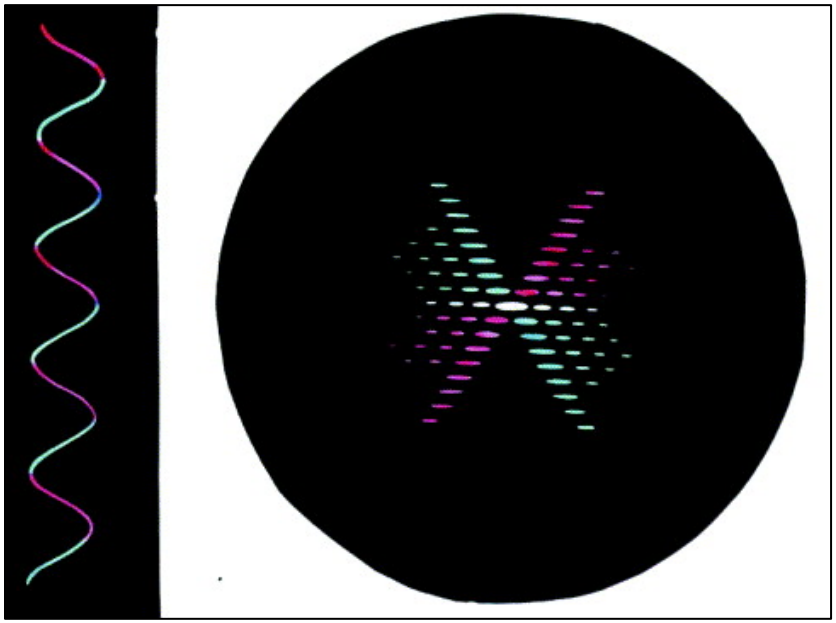
www.archives.upenn.edu/img/20011023003x200.jpg




http://pubs.acs.org/cen/coverstory/8110/8110dna2.html

# DNA: is a Helix

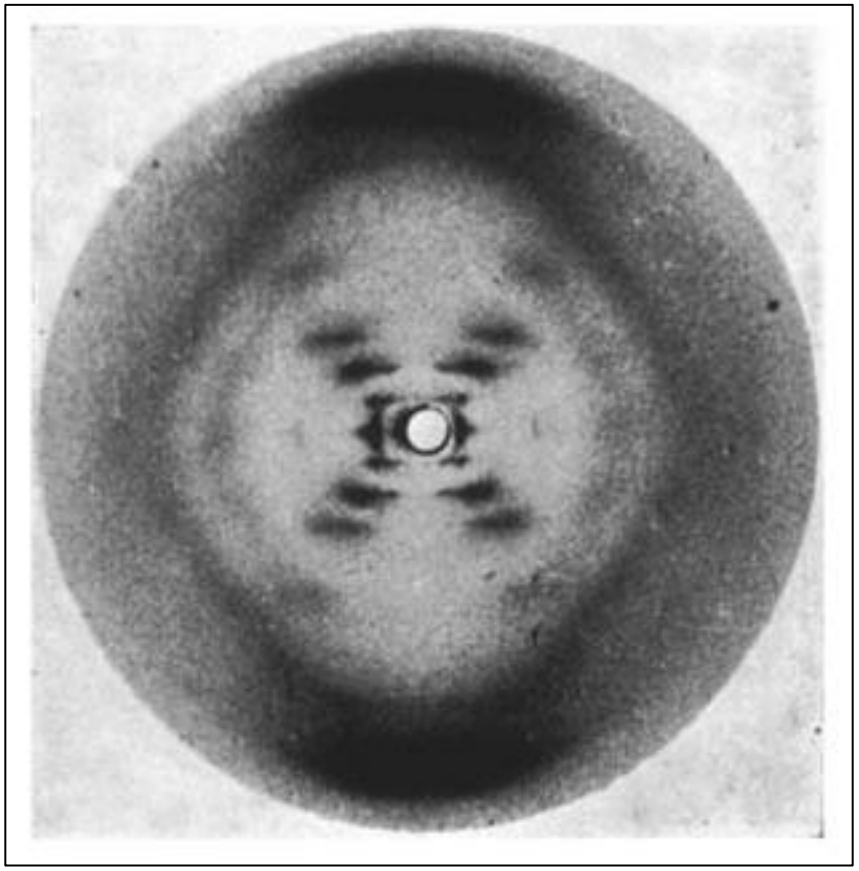
Diffraction of a Helix



http://ars.els-cdn.com/content/image/1-s2.0-S0022283603014128-gr10.jpg

- Cochran, Crick and Vand published theoretical paper on X-ray diffraction from helical polymers (1952) *Acta Crystallographica 5*, 5811
- Motivation Pauling's α-helix

Franklin's Photograph 51

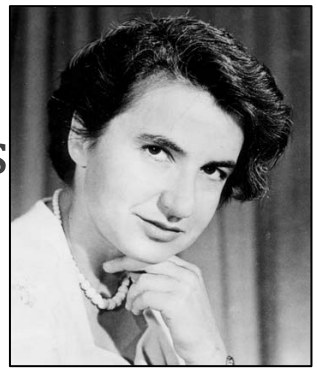


http://ars.els-cdn.com/content/image/1-s2.0-S0022283603014128-gr16.jpg

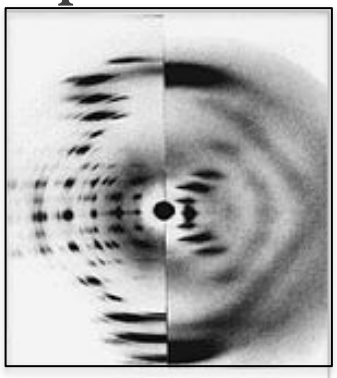
**Diffraction Demo**

# DNA: Fiber Diffraction

- Rosalind Franklin (King's College London) recorded X-ray diffraction from DNA fibers
- Franklin and Wilkins identified 2 DNA forms:
  - A-form: Less hydrated (better ordered)
  - B-form: More hydrated
- A- (and B-) form DNA diffraction show → space groups was C2
  - Clue → antiparallel DNA strands



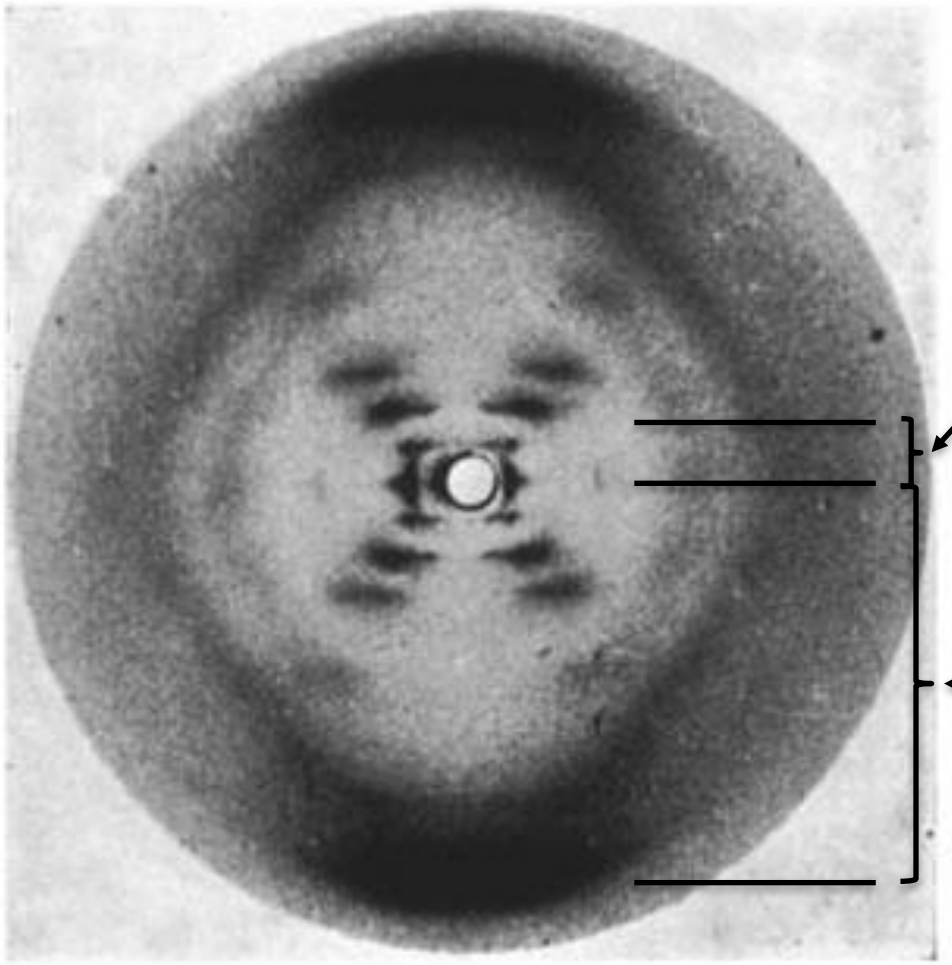
Franklin



A-DNA B-DNA

http://www.intellectualventureslab.com/invent/scientists-in-history-rosalind-franklin

# Details of Franklin's Photograph 51



- Helical cross!
  - DNA is a helical polymer
- Regular layer line spacing ($1/34Å^{-1}$)
  - DNA is composed of substructures ~3.4Å apart
- Meridional reflection with smaller spacing ($1/3.4Å^{-1}$)
  - Overall structure of DNA repeats every ~34Å (or every 10 substructures)

# Building a 3D Model for DNA

## Pauling and Corey (1953)

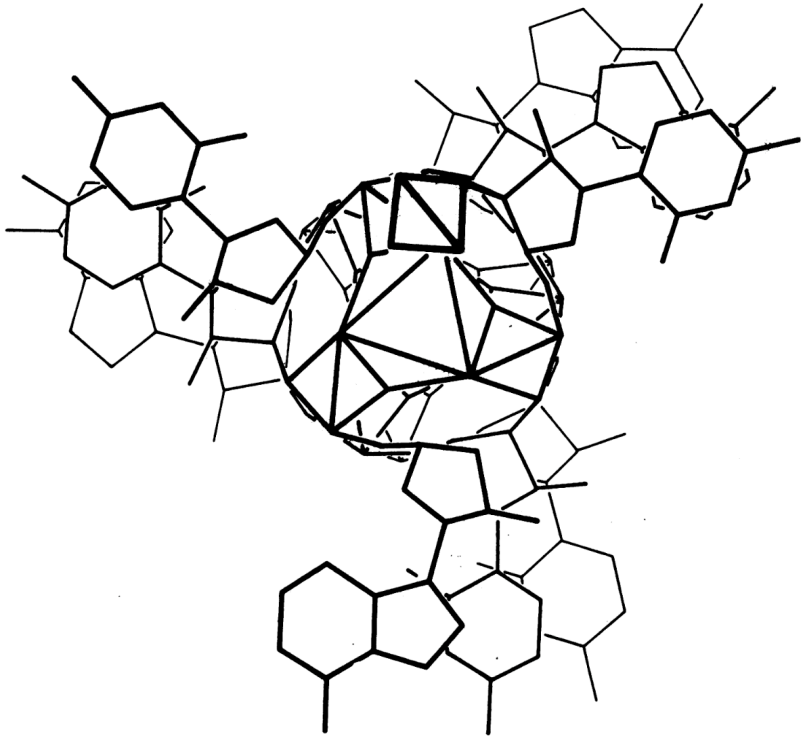


FIGURE 6

Plan of the nucleic acid structure, showing several nucleotide residues.

http://www.pnas.org/content/39/2/84.full.pdf

## Fraser (1952), unpublished

diffraction, and physico-chemical properties. The task of integrating these observations to give a detailed picture of the structure of the nucleic acids is a formidable one, and is by no means complete. However, in view of the letter by Pauling and Grey it seems worth describing a type of structure that we have considered, which, although it involves three intertwined helical polynucleotide chains, differs considerably from that formulated by them.

-snip-

(a) electrostatic attractions between the negatively charged phosphate groups and the sodium ions,

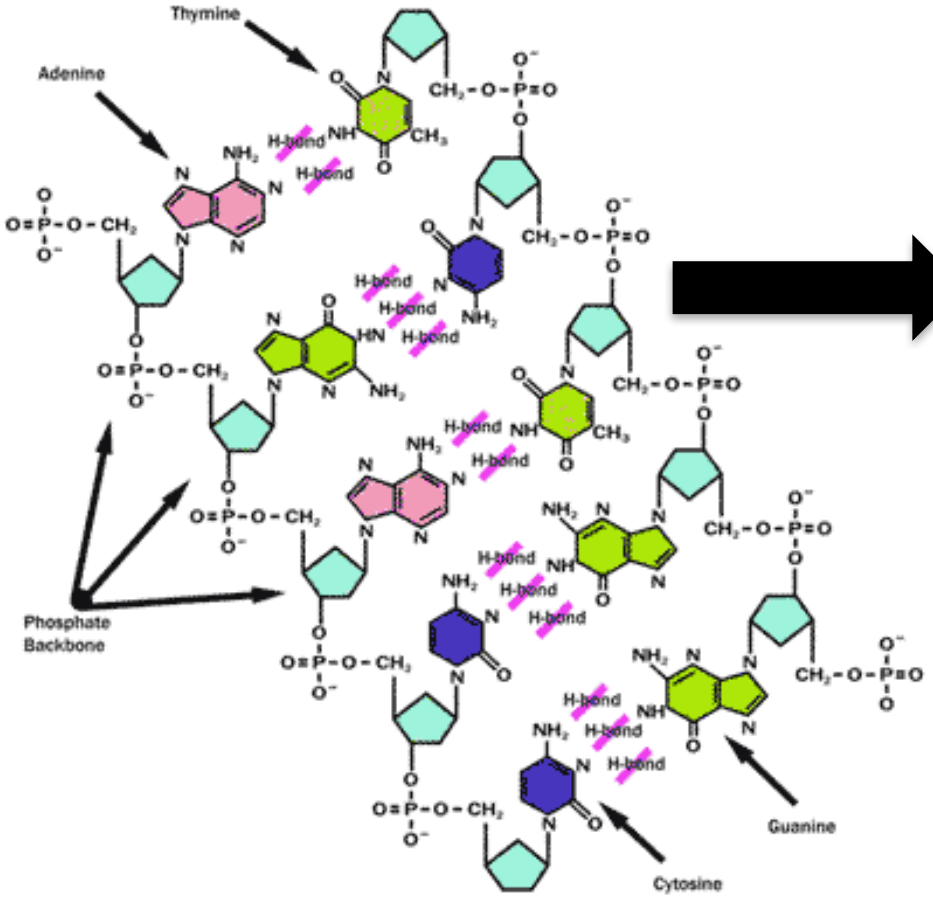
(b) Van der Waals attractions between the planar purine and pyrimidine residues,

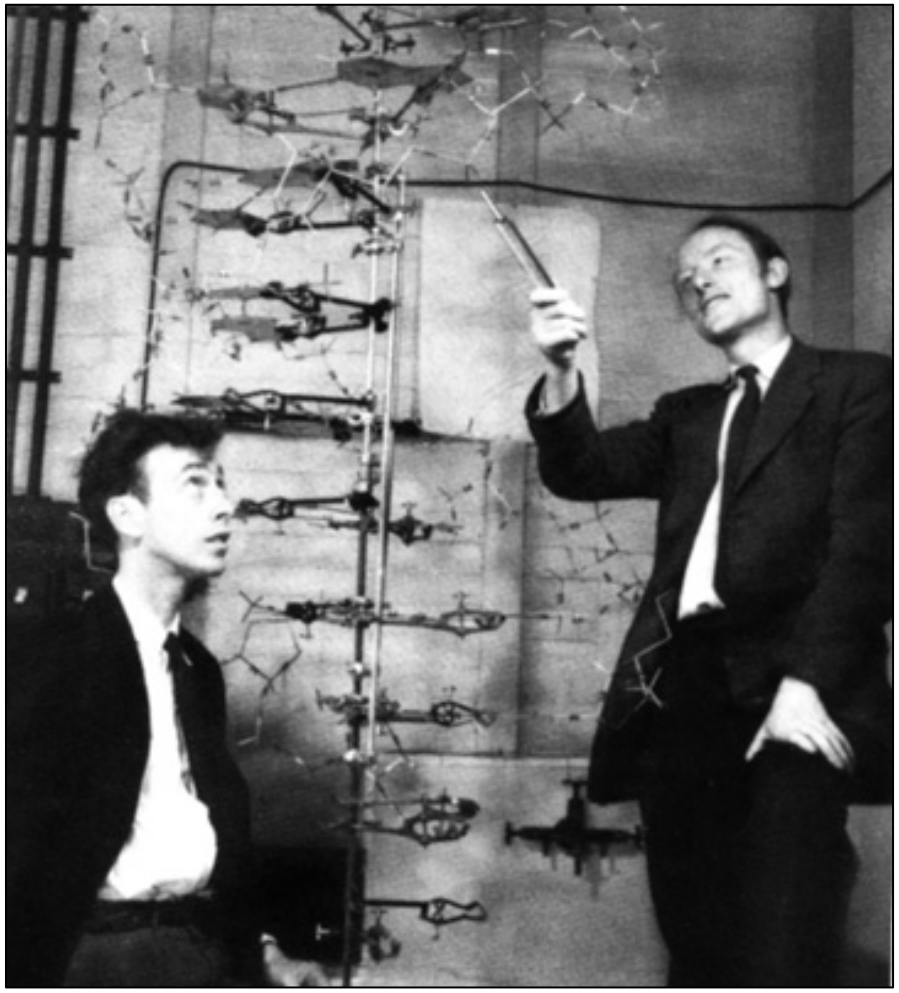
(e) hydrogen bonds formed between the C=O, NH, NH and OH groups of the purine and pyrimidine residues.

Models of polynucleotide chains were built assuming that the β-deoxyribofuranosides were joined by 3',5' phosphate diester linkages, as in Figure 1.

http://scarc.library.oregonstate.edu/coll/pauling/dna/notes/fraser-structure-01-large.html

# Building a 3D model for DNA – Cont.



# Publish or Perish!

- We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.) …

- It has not escaped our notice that the specific pairing we have postulated immediately suggests a copying mechanism for the genetic material.



equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.


[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1949).

[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).

[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1956).

[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik.* (Stockholm), **2** (11) (1906).


## MOLECULAR STRUCTURE OF NUCLEIC ACIDS

### A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[5]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3′,5′ linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the *z*-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical *z*-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

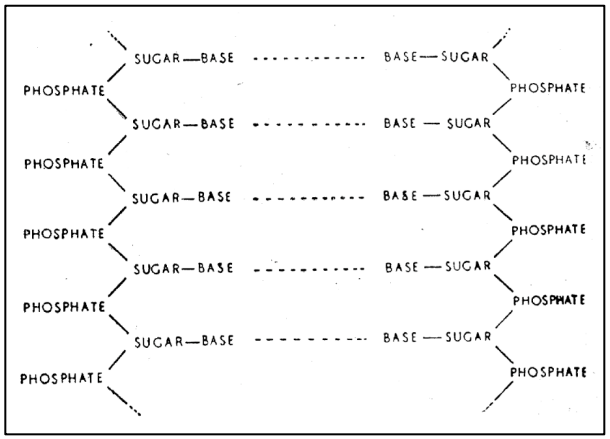
We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

# GENETICAL IMPLICATIONS OF THE STRUCTURE OF DEOXYRIBONUCLEIC ACID

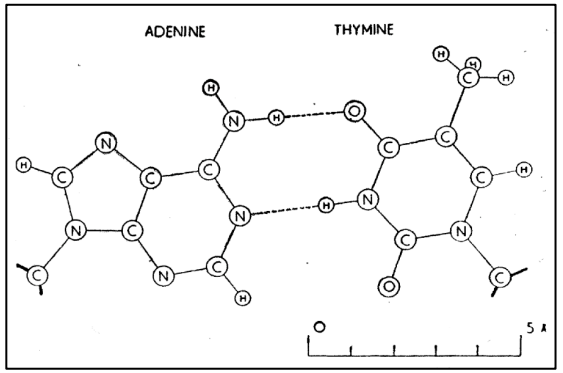
By J. D. WATSON and F. H. C. CRICK

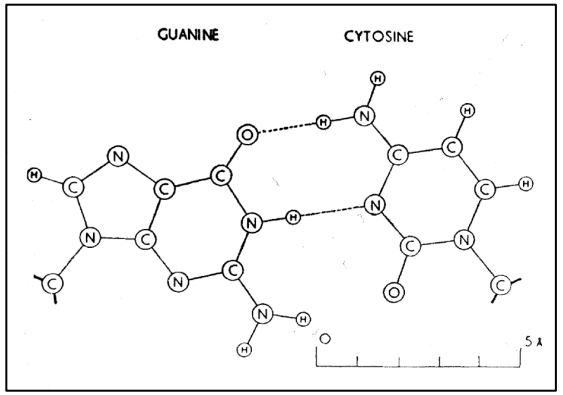
Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, Cavendish Laboratory, Cambridge

No. 4361 May 30, 1953 NATURE




Watson-Crick Base Pairs



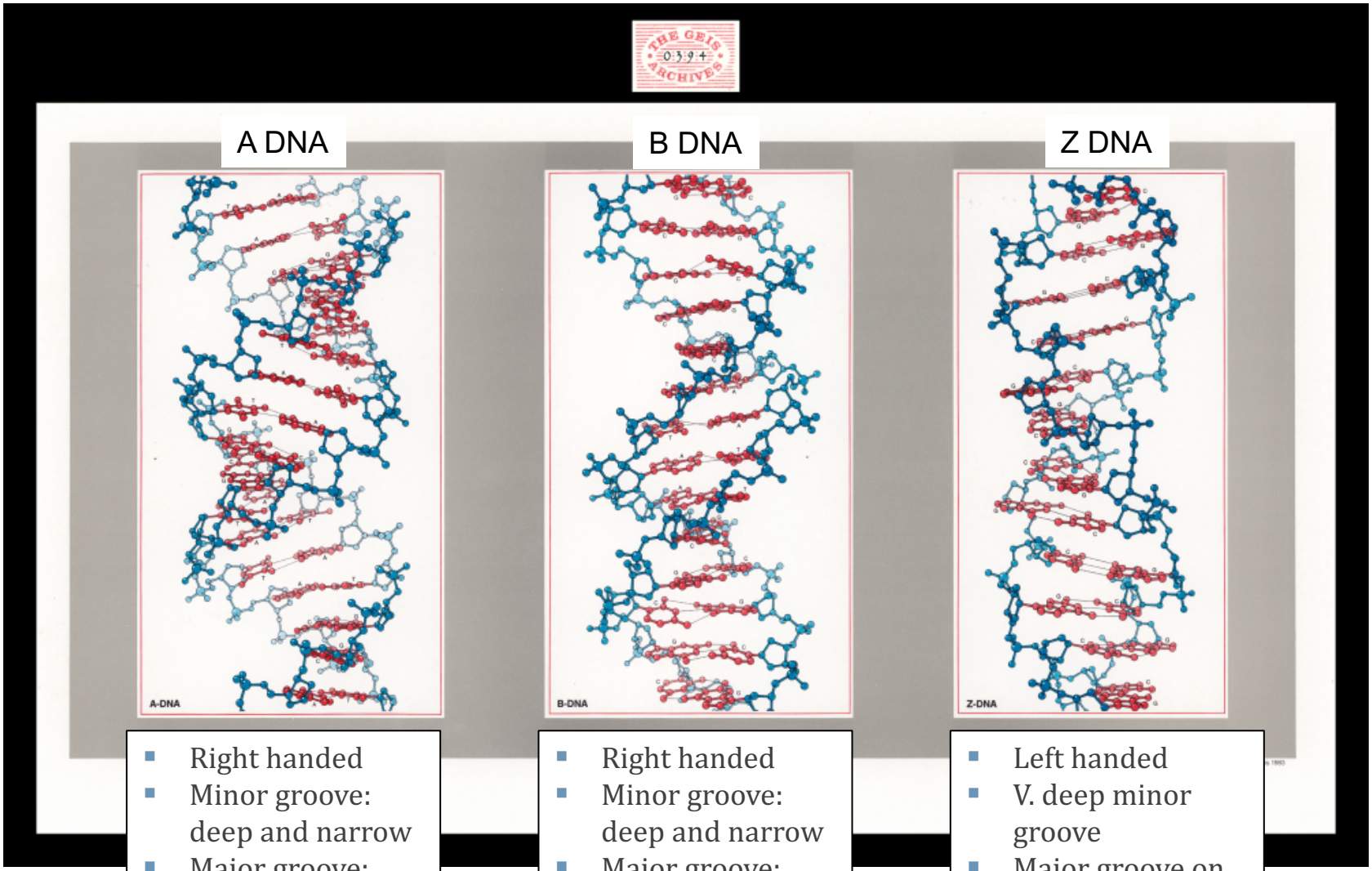




For the moment, the general scheme we have proposed for the reproduction of deoxyribonucleic acid must be regarded as speculative. Even if it is correct, it is clear from what we have said that much remains to be discovered before the picture of genetic duplication can be described in detail. What are the polynucleotide precursors ? What makes the pair of chains unwind and separate ? What is the precise role of the protein ? Is the chromosome one long pair of deoxyribonucleic acid chains, or does it consist of patches of the acid joined together by protein ?

# Lessons from Watson and Crick

- Only tackle Important Problems
- Be in the "Right Place at the Right Time"
- Collaboration is Critical for Success
  - Different Scientific/Cultural Perspectives
  - Complementary Technical Skills
- Competition is Central to Rapid Progress in Research
- Have the Courage of Your Convictions

# More About the DNA Double Helix




- Right handed
- Minor groove: deep and narrow
- Major groove: wide and shallow

- Right handed
- Minor groove: deep and narrow
- Major groove: deep and wide

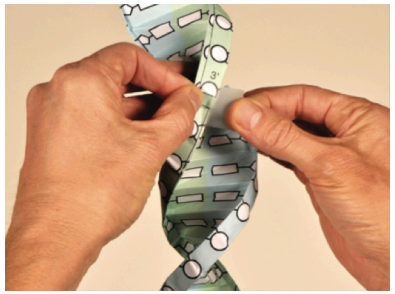
- Left handed
- V. deep minor groove
- Major groove on outside

https://cdn.rcsb.org/pdb101/geis/images/600/geis-0394-dna.png

# Exercise 1

- Search the PDB for structures of B-DNA:
  1. When was the first B-form duplex DNA structure determined? What is its PDB Identifier?
  2. Open the structure summary page for this PDB entry and answer the following:
     - Describe the overall structure. Where are the bases, phosphates, sugars?
     - Comment on the H-bonding pattern between bases
     - Where are the major and minor groves in this structure? How did you figure this out?

# Further Exploration

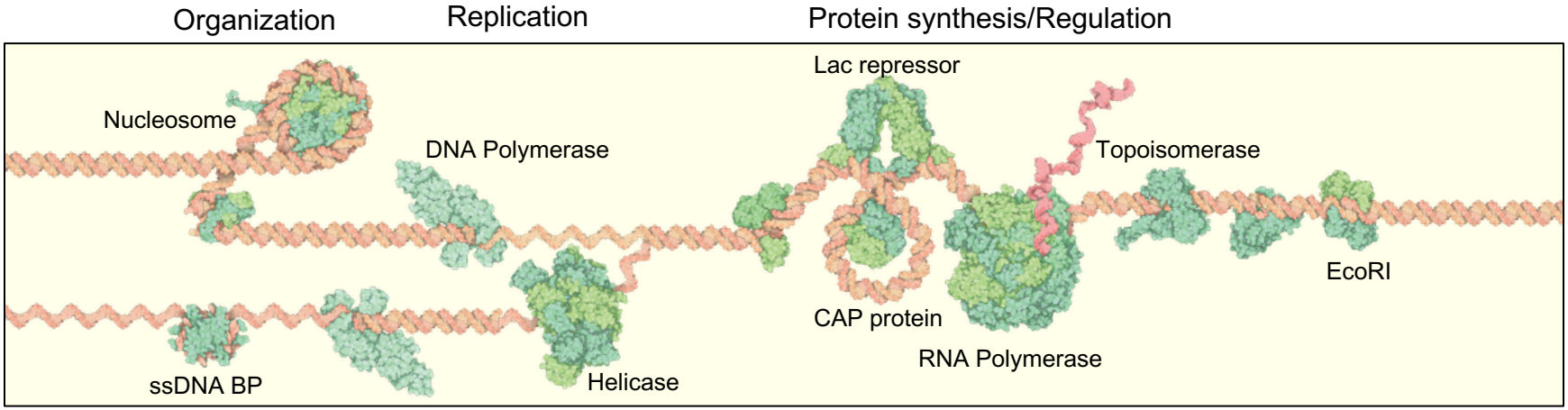


https://cdn.rcsb.org/pdb101/learn/resources/dna-model-2013_2.pdf

- Make a paper model of the DNA
- Search the PDB for structures of A-DNA and Z-DNA
- Visualize these structures and answer the following questions:
  - What is the handedness of each of these DNA helices?
  - In these structures, is the H-bonding pattern within base pairs the same or different?
  - Where are the major and minor grooves located? How did you figure this out?

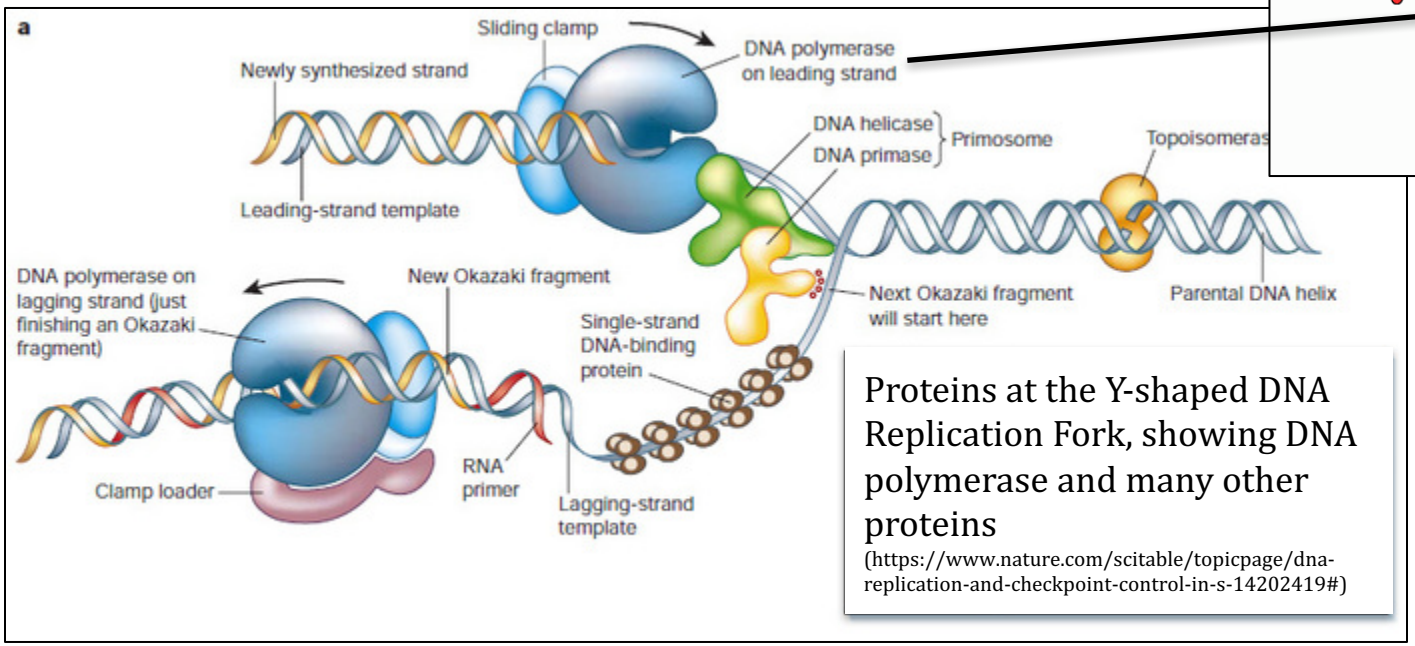
# Learning Objectives

- Modeling in 3D
- Modeling the Double Helix
- Functions of the Double Helix
  - Genetic Blueprint: Replication
  - Genetic Code: Protein Synthesis
  - Organization of DNA in higher organisms
- Designing with the Double Helix

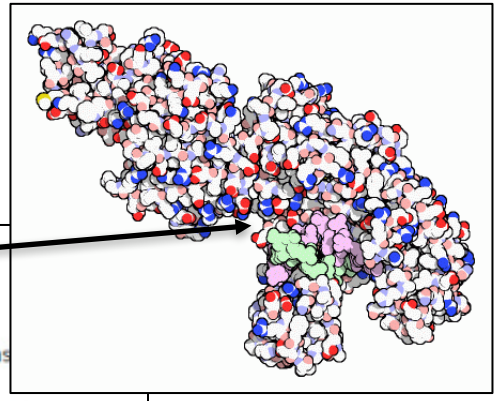


# Double Helix and Duplicating DNA

- Semiconservative replication – with leading and lagging strands
- Involves many proteins



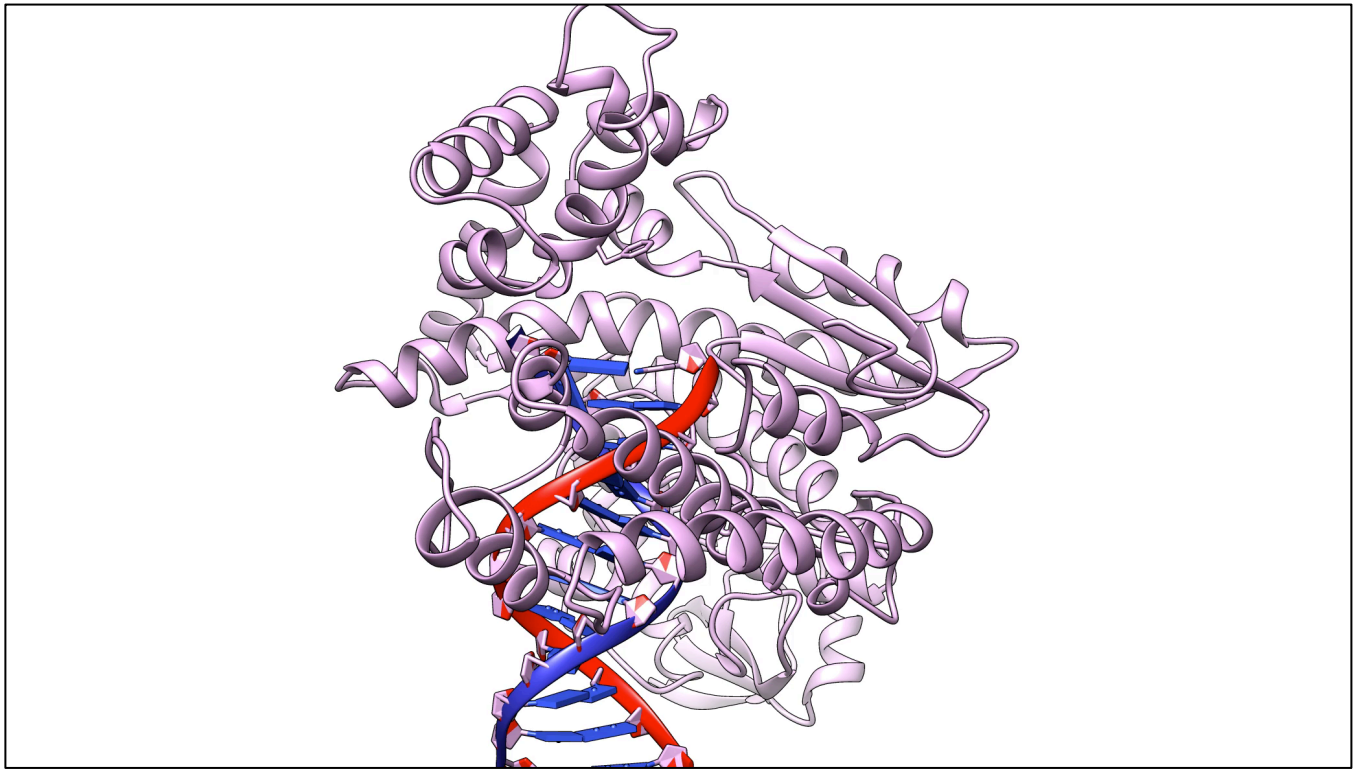

Proteins at the Y-shaped DNA Replication Fork, showing DNA polymerase and many other proteins
(https://www.nature.com/scitable/topicpage/dna-replication-and-checkpoint-control-in-s-14202419#)



Taq DNA polymerase

http://pdb101.rcsb.org/motm/3

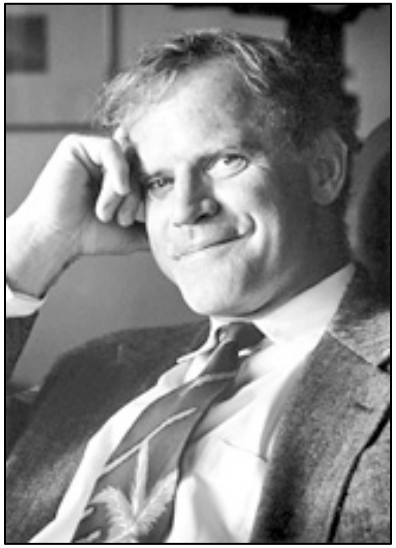
# Polymerase Chain Reaction



Taq Polymerase in action, showing the template and primer strands.
PDB entries 4ktq, 2ktq, and 3ktq, (Li, Korolev, Waksman 1998)

http://www.hhmi.org/biointeractive/polymerase-chain-reaction

Mullis



http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1993/mullis.html

# Exercise 2

- Read the MotM feature on DNA polymerase (pdb101-alpha.rcsb.org/motm/3). Compare the *E.coli* and *T. aquaticus* DNA polymerase enzymes:
  - List one similarity in the functions of the enzymes.
  - List one difference in the functions of the enzymes.
- Open the structure summary (SS) page for the structure of Taq polymerase (PDB entry 1tau).
  - Name the polymers present in this structure. Hint: examine the Macromolecules section of the SS page
  - How many domains does the protein have? Name them.
  - How many structures of the Taq polymerase can you find in the PDB? Explain your approach to this answer.
  - Why are there so many Taq polymerase structures?

# Further Exploration

- What does the Taq polymerase do when the template strand is missing a base or has a modified base?
- Can the Taq polymerase incorporate unnatural bases?
- What do other proteins involved in DNA replication look like? What are their functions? Visualize their structures and explain their functions.

# The Genetic Code

- 1958, Crick → Central Dogma of Biology:
  - DNA → RNA → Protein
- mRNA transmits genetic information from DNA (nucleus) to ribosome (cytoplasm)
- 1961, the genetic code is:
  - a group of 3 bases code for an aa
  - non-overlapping
  - read from a starting point
  - probably degenerate

Crick and Brenner


wellcomelibrary.org/content/2843/23971/brenner-crick.jpg

No. 4809 December 30, 1961 NATURE 1227

GENERAL NATURE OF THE GENETIC CODE FOR PROTEINS

By Dr. F. H. C. CRICK, F.R.S., LESLIE BARNETT, Dr. S. BRENNER and Dr. R. J. WATTS-TOBIN

Medical Research Council Unit for Molecular Biology, Cavendish Laboratory, Cambridge

| | | MUTATION | PHENOTYPE |
|---|---|---|---|
| Wild-type sequence | A B C A B C A B C A B C A B C A B C A B C A B C | NONE | rII+ |
| FC0 mutant | A B C A A B C A B C A B C A B C A B C A B C A B | + | rII− |
| Supression of FC0 | A B C A A B A B C A B C A B C A B C A B C A B C | + − | rII+ |
| Two base additions | A B C A A B C A B C B A B C A B C A B C A B C A | + + | rII− |
| Three base additions | A B C A A B C A B C B A B C C A B C A B C A B C | + + + | rII+ |

+ Base addition
− Base deletion

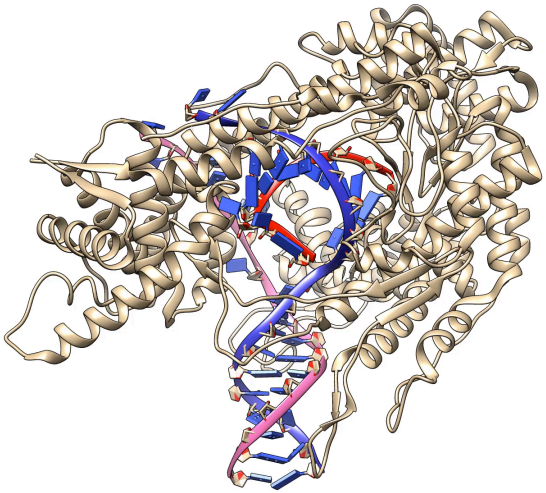
Cell, Volume 128, Issue 5, 9 March 2007, Pages 815-818

# The Genetic Code – Cont.

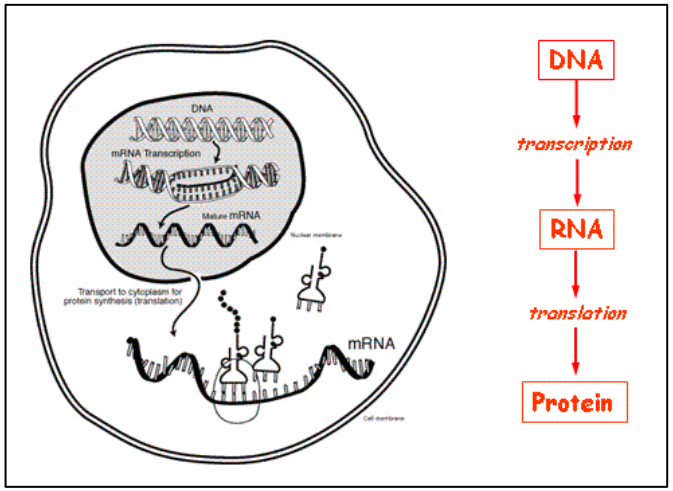
- Nirenberg:
  - 1961: Cell free protein synthesis studies
    poly U (RNA) → codes for poly Phe
  - By 1966 all 64 codons for 20 aa deciphered
- Khorana:
  - 1963-66: synthesized deoxypolynucleotides
    - Template for RNA polymerase → RNA polymers
  - 1979: Total synthesis of a gene
- Holley:
  - 1965: isolated, sequenced and determined the structure of Ala tRNA

https://www.nobelprize.org/nobel_prizes/medicine/laureates/1968/

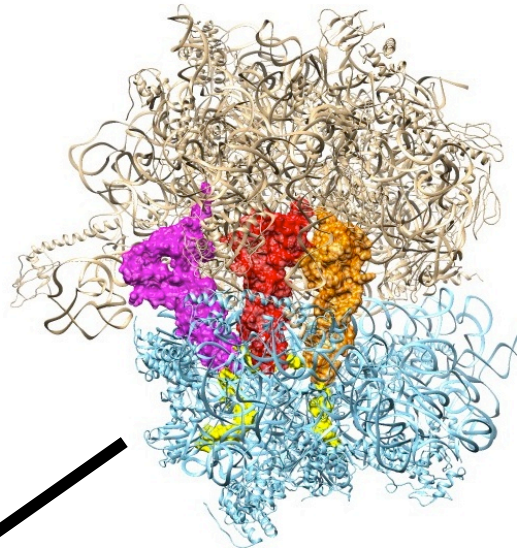
# Double Helix and The Central Dogma



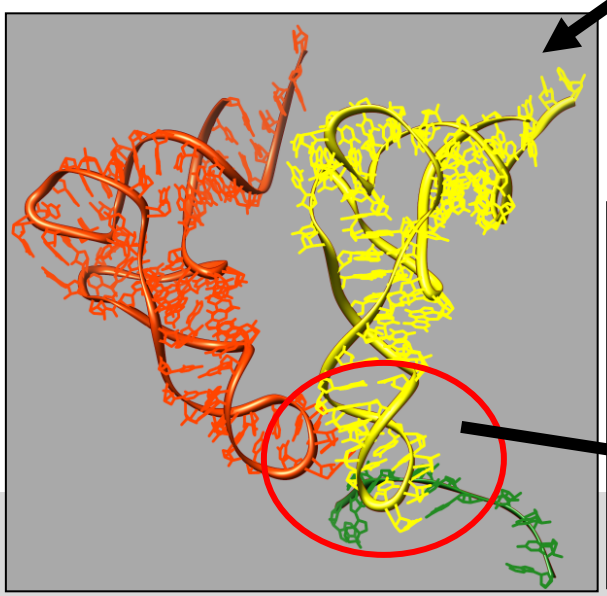
T7 RNA Polymerase in action showing the coding, non-coding DNA strands, and the transcribed RNA. PDB entry 1msw (Yin and Stietz, 2002)

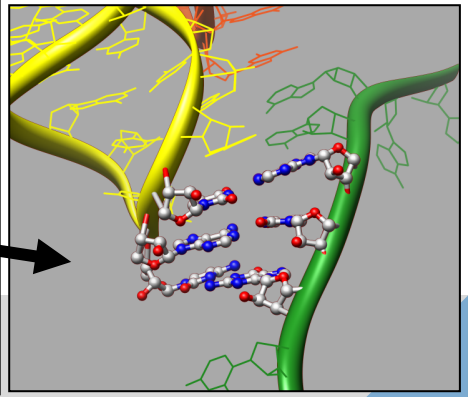



https://www.ncbi.nlm.nih.gov/Class/MLACourse/Modules/MolBioReview/images/central_dogma.gif



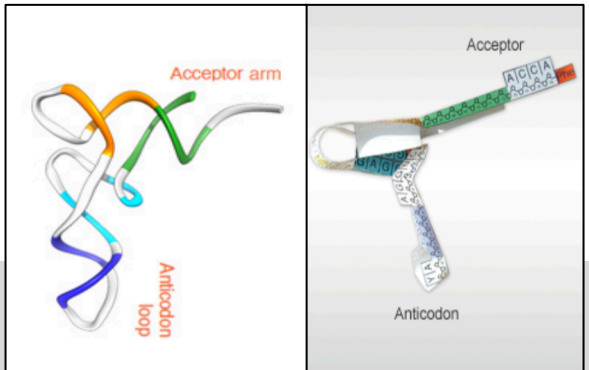
Ribosome in action – showing mRNA and tRNAs bound to ribosome. PDB entry 4v6f (Jenner et al., 2009)





# Further Explorations

- Read the MotM on RNA polymerase (pdb101-alpha.rcsb.org/motm/40).
  - Are all DNA and RNA polymerases accurate?
  - What is the importance of proofreading?
- Examine the structure of an RNA polymerase in action (PDB entry 1msw). What is the helical form of the DNA:DNA and DNA:RNA duplexes?
- Make a paper model of tRNA and try the tRNA activity (pdb101.rcsb.org/learn/paper-models/trna)



# Chromosomes: History and Function

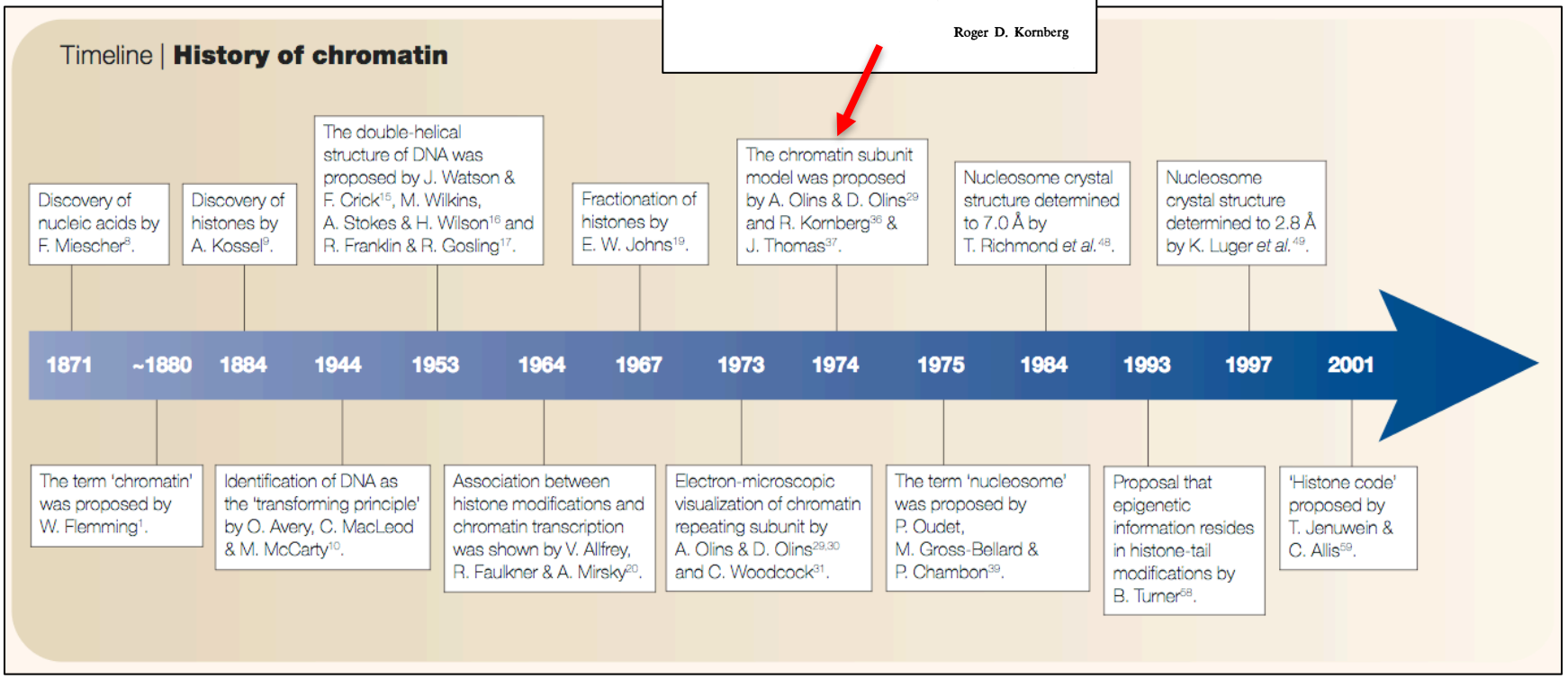
- "Chromatin" coined, histones identified in 1880s!

**Chromatin Structure: A Repeating Unit of Histones and DNA**

Chromatin structure is based on a repeating unit of eight histone molecules and about 200 DNA base pairs.

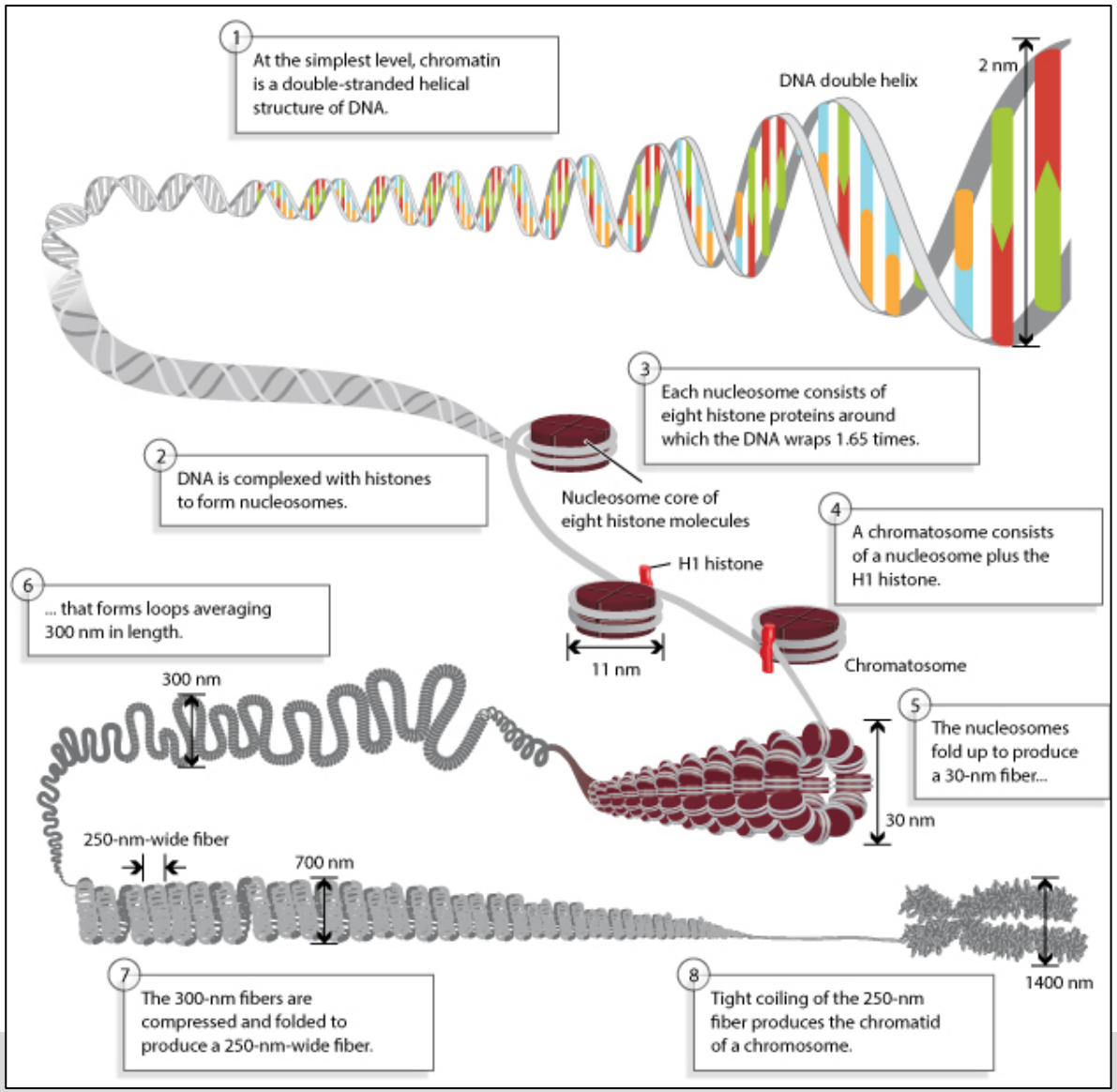
Roger D. Kornberg

*Science* 24 May 1974:
Vol. 184, Issue 4139, pp. 868-871
DOI: 10.1126/science.184.4139.868



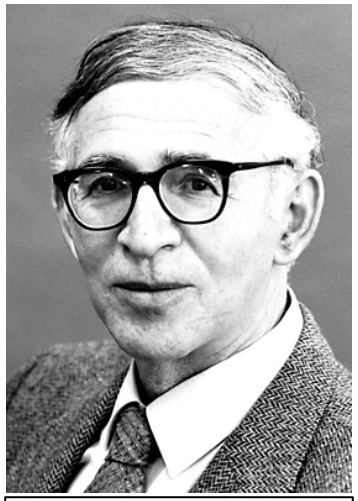

https://www.nature.com/scitable/content/Chromatin-history-our-view-from-the-bridge-30352

# Double Helix: The Packing Problem

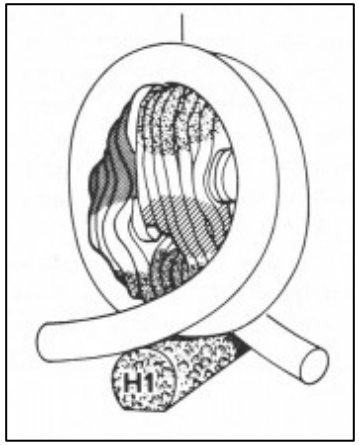



https://www.nature.com/scitable/content/ne0000/ne0000/ne0000/ne0000/113158606/18847_6.jpg

Klug



https://www.nobelprize.org/nobel_prizes/chemistry/laureates/1982/klug_postcard.jpg

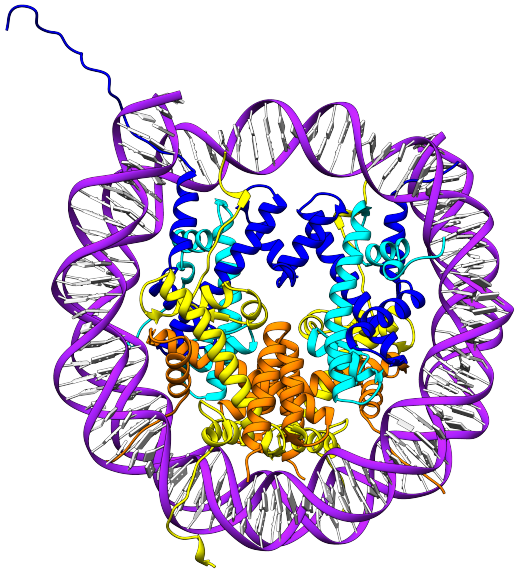



www2.mrc-lmb.cam.ac.uk/wordpress/wp-content/uploads/klug_nobel_research2-415x520.jpg

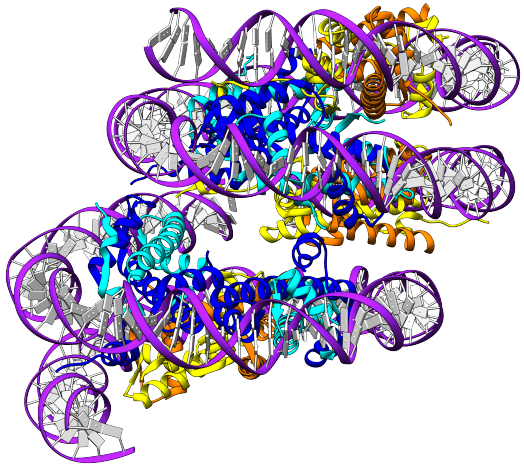
Nucleosome model (EM data)

# Structure of the Nucleosome

- Luger et al., 1997 structure revealed:
  - Nucleosome core – has 2 copies each of H2A, H2B, H3 and H4 (octamer) w 145–147 (DNA bps)
- Chromatin structure → transcription regulation
  - Histones collaborate w transcription factors → their removal and/or modification → gene derepression



Nucleosome core particle bound to DNA. PDB entry 1aoi (Luger et al., 1997)



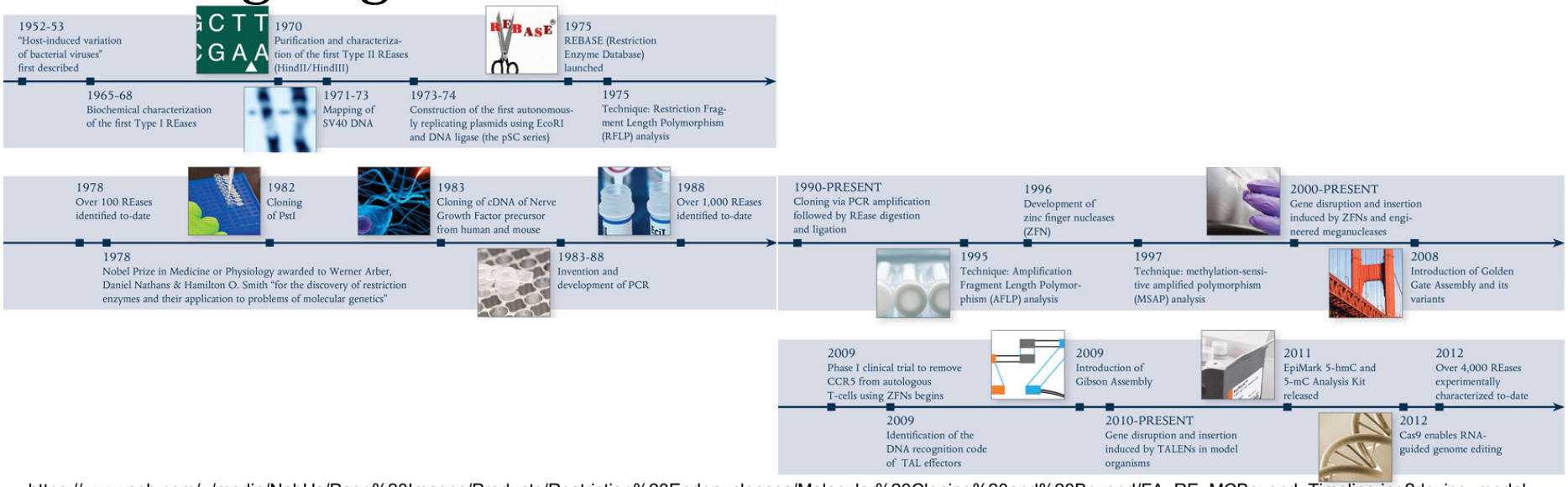
Overlapping dinucleosome provides insights into chromatin remodeling. PDB entry 5gse (Kato et al., 2017)

# Further Explorations

- Read the MotM on Nucleosomes (pdb101.rcsb.org/motm/7)
- Visualize and explore the structure of a nucleosome (PDB entry 1aoi). Answer the following questions based on your explorations:
  - What is the conformation of the DNA in this structure?
  - Describe the histone protein:DNA interactions. Are they sequence specific?

# Learning Objectives

- Modeling in 3D
- Modeling the Double Helix
- Functions of the Double Helix
- Designing with the Double Helix




https://www.neb.com/~/media/NebUs/Page%20Images/Products/Restriction%20Endonucleases/Molecular%20Cloning%20and%20Beyond/FA_RE_MCBeyond_Timeline.jpg?device=modal

# Restriction Endonucleases

- Bacterial defense against viral infection – innate immunity
- Discovered 1960s
- Cleaves viral (foreign) DNA at specific locations
- Works with DNA methylases → modify host restriction sites to protect from being cleaved

Arber

Smith

https://www.nobelprize.org/nobel_prizes/medicine/laureates/1978/

**TYPES AND ACTIVITIES OF RESTRICTION ENZYMES**

**Type I**
Cleaves DNA at random sites far from its recognition sequence

**Type II**
Cleaves DNA at defined positions close to or within its recognition sequence

**Type IIG**
Cleaves outside its recognition sequence with both REase and MTase enzymatic activities in the same protein

**Type IIP**
Cleaves symmetric targets and cleavage sites

**Type IIS**
Recognizes asymmetric sequences

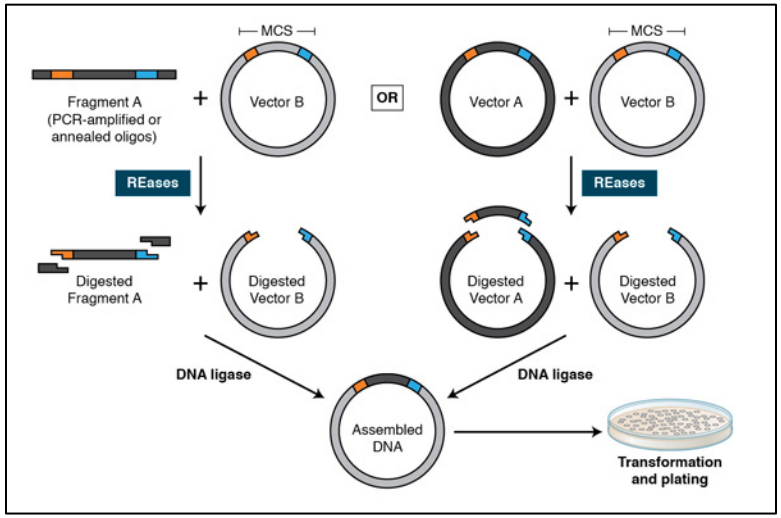
**Type III**
Cleaves outside its recognition sequence and require two sequences in opposite orientations within the same DNA

**Type IV**
Cleaves modified (e.g., methylated) DNA

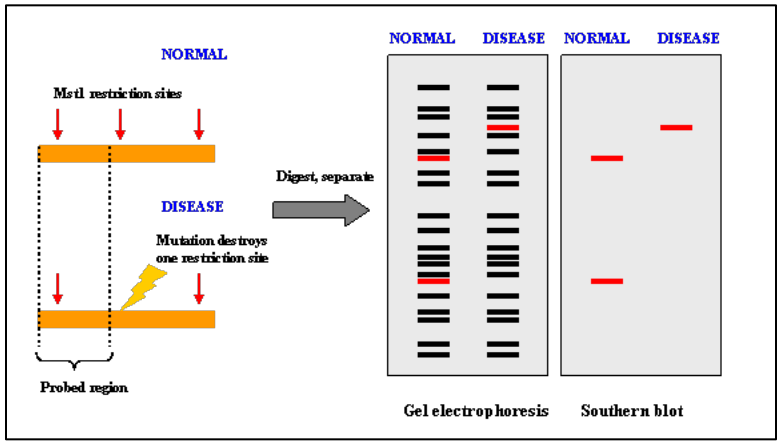
https://www.neb.com/~/media/NebUs/Page%20Images/Products/Restriction%20Endonucleases/Molecular%20Cloning%20and%20Beyond/FA_RE_MCBeyond_RETypes.jpg

# Restriction Endonucleases: Applications

- Cloning:
  - Uses restriction enzymes, gene of interest and ligases
  - Used for recombinant DNA
- DNA mapping:
  - Difference in homologous DNA sequences → DNA fragments of different lengths after digestion with specific REs
  - Used for genotyping, forensics, paternity tests, hereditary disease diagnostics

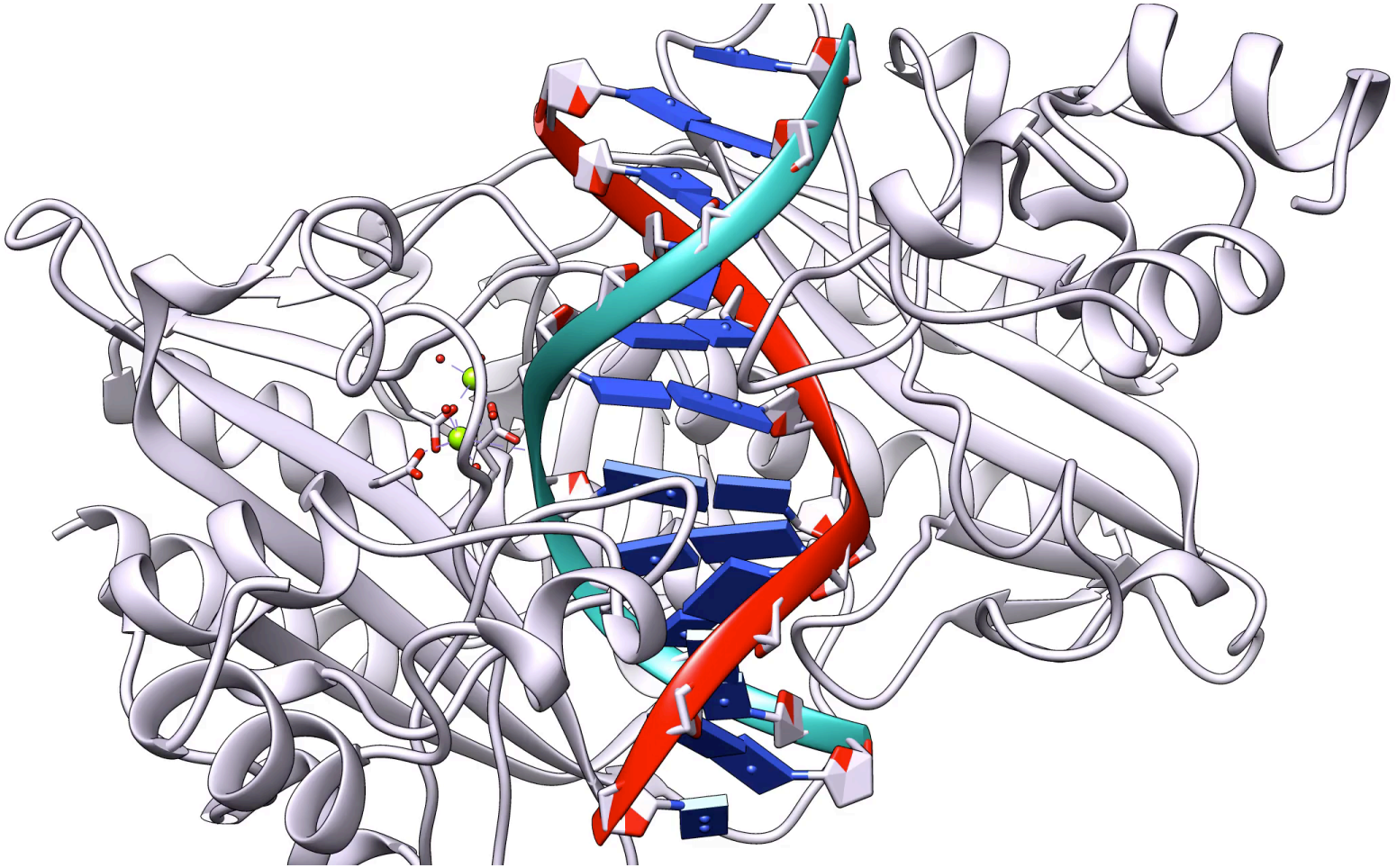



https://www.neb.com/~/media/NebUs/Page%20Images/Products/Restriction%20Endonucleases/Molecular%20Cloning%20and%20Beyond/FA_RE_MCBeyond_Fig2_ClassicCloning.jpg




https://www.ncbi.nlm.nih.gov/probe/docs/techrflp/

# Double Helix in Restriction Endonuclease

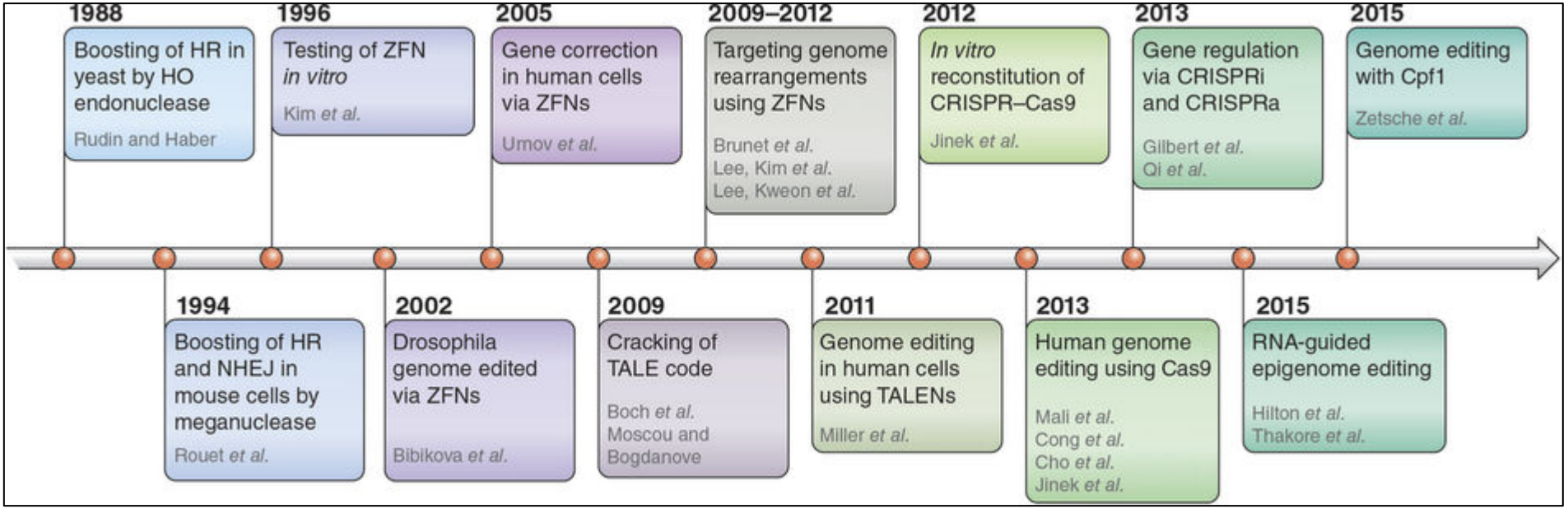


Animation showing restriction digestion by EcoRV. PDB entries 1rvb and 1rvc (Kostreva and Winkler 1995)

# Further Explorations

- Read the MotM feature on restriction endonucleases (pdb101-alpha.rcsb.org/motm/8)
- Examine the structures of the restriction endonuclease EcoRV (PDB entries 1rva, 1rvb, 1rvc)
  - How does the protein recognize the specific DNA sequence?
    - Visualize the structures and explore the protein:DNA interactions.
    - Which interactions are specific to the EcoRV target site?
- Identify the structures of other restriction endonucleases (*e.g.,* HindIII, BamHI) in the PDB and explore how the enzyme recognizes its cognate sequences.

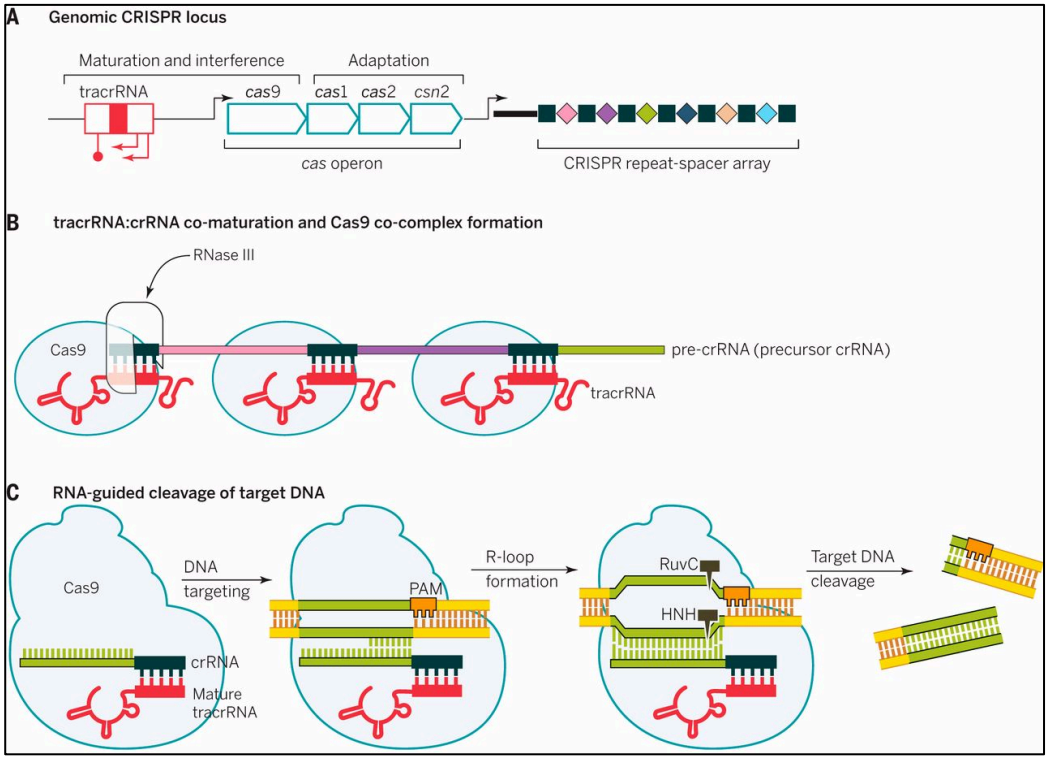
# Genome Editing




https://www.nature.com/nprot/journal/v11/n9/images/nprot.2016.104-F1.jpg

# CRISPR-Cas9

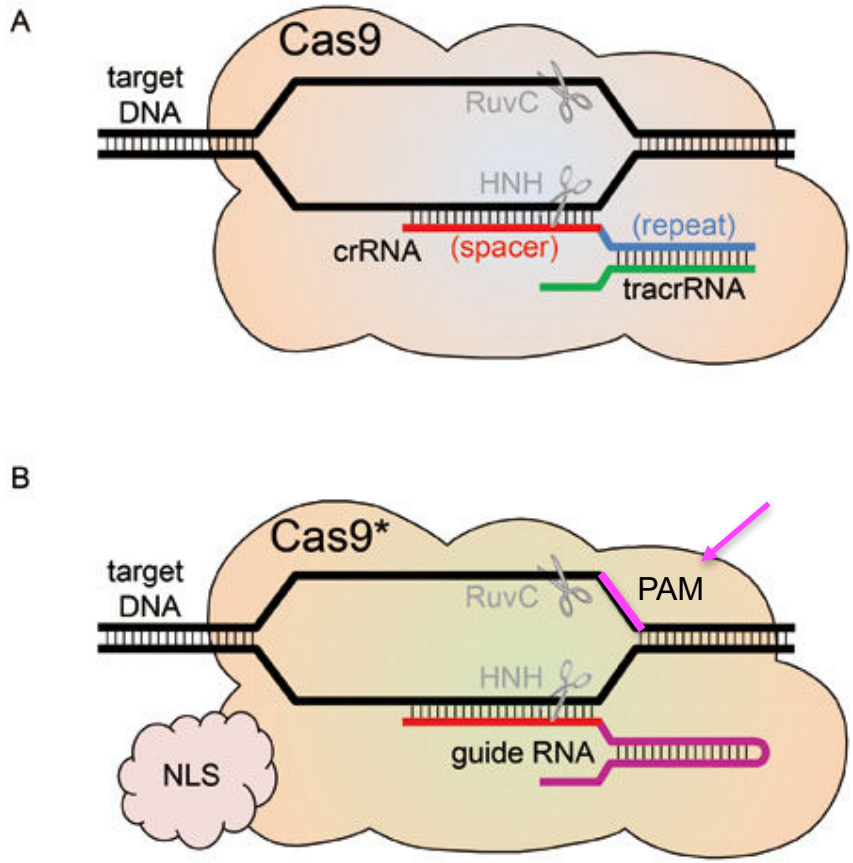
- Clustered Regularly Interspaced Palindromic Repeats (CRISPR)
- Encodes instructions for adaptive immune system → protects microbes against specific phage infections.




https://d2ufo47lrtsv5s.cloudfront.net/content/sci/346/6213/1258096/F2.large.jpg

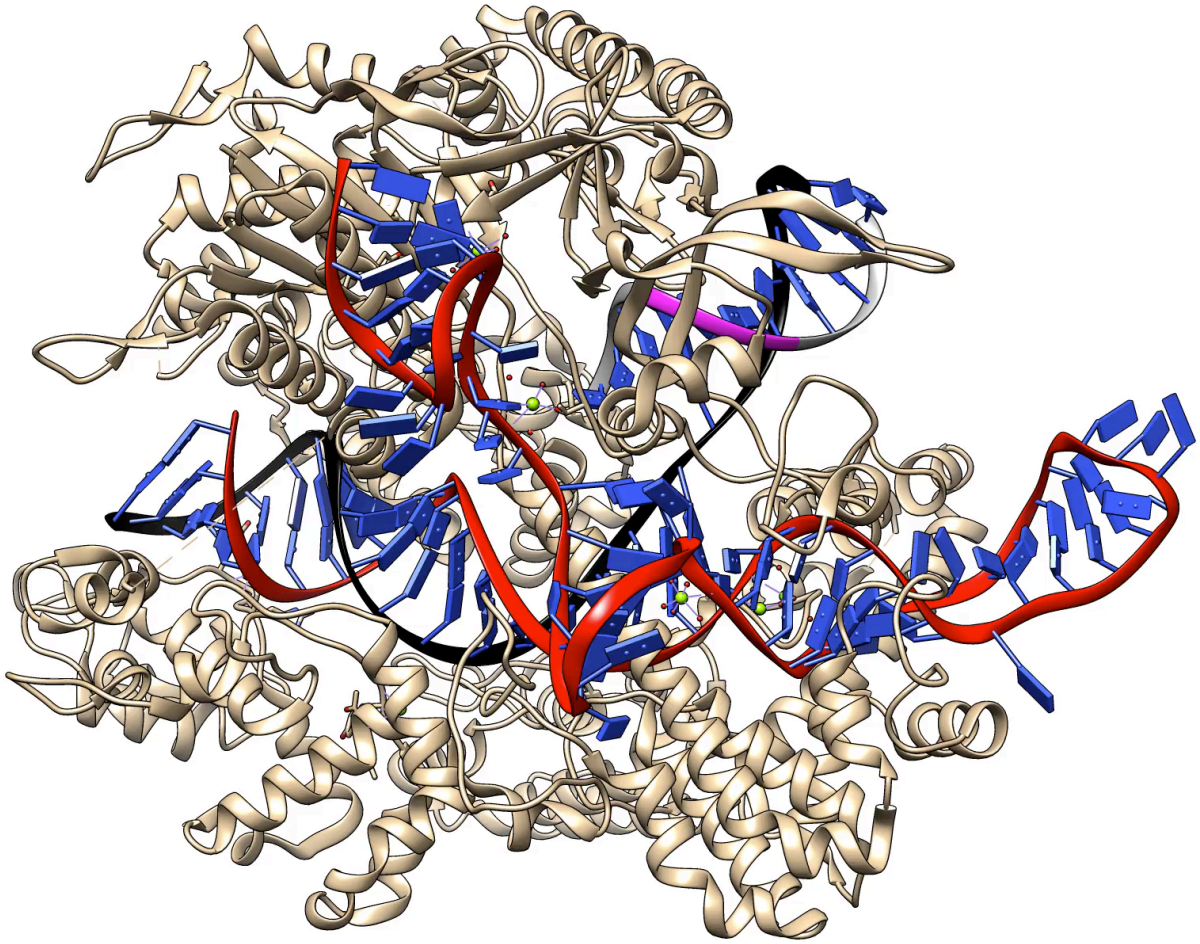
# Genome Editing Using CRISPR-Cas9

- Cas9 guided to specific locations in genomes by a short RNA search string (guide RNA)
- DNA sequences and their functional outputs editable
- Guide RNA → fuse crRNA + tracrRNA (facilitates DNA cleavage by Cas9 in vitro)




www.nature.com/cr/journal/v23/n6/images/cr201339f1.jpg

# Double Helix and CRISPR Cas-9



Structure of Cas9 bound to PAM-containing DNA target and guide RNA. PDB entry 4un3 (Anders, C., Niewoehner, O., Duerst, A., Jinek, M. 2014)

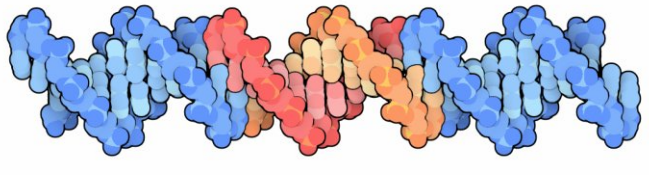
See also: http://www.cell.com/cms/attachment/2081785784/2072571835/mmc1.mp4

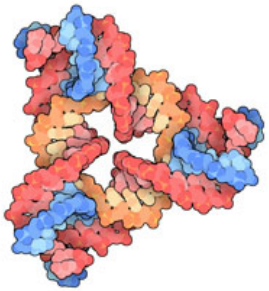
# Further Explorations

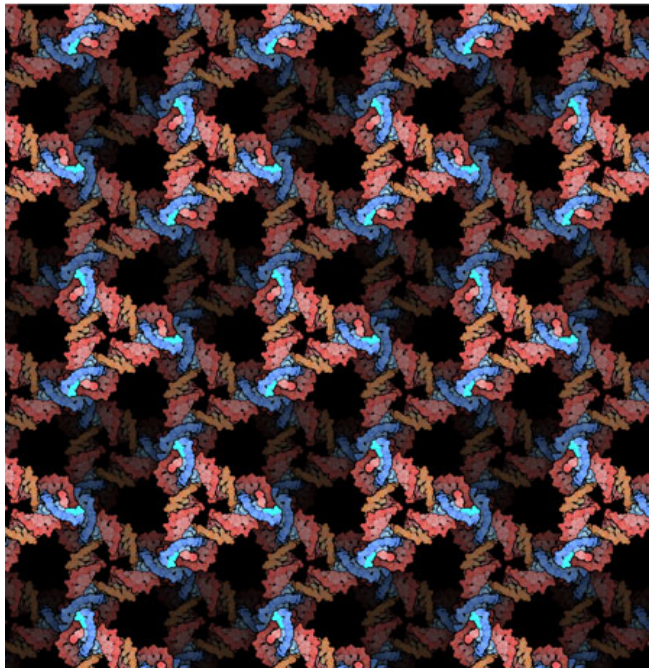
- Read the MotM feature on CRISPR proteins (pdb101-alpha.rcsb.org/motm/181)
- Examine the structure of a Cas9 bound to PAM-containing DNA target (PDB entry 4UN3):
  - Visualize and identify the target and non-target DNA strands and guide RNA
  - How is the protospacer adjacent motif (PAM) sequence recognized by Cas9? Explain at the molecular level
- Identify and explore the structures and mechanisms of function of other genome editing mechanisms, *e.g.*,:
  - Zinc finger nucleases
  - Transcription Activator Like Effector Nuclease Proteins (TALENS)

# Designer DNAs

- Designed DNA sequences with sticky ends
- Hybridize to form extended or complex structures
- Join DNAs with ligase
- Makes scaffolds that
  - Has spaces in the lattice to host other molecules, like proteins, and orient them
  - May be used for building nanoscale electrical devices



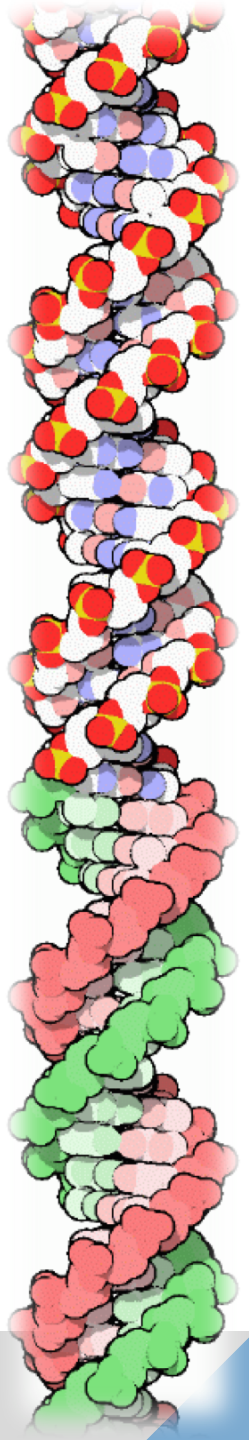




pdb101.rcsb.org/motm/119

# Summary

- Modeling in 3D
  - Protein Data Bank
  - RCSB PDB data, tools and resources
- Modeling the Double Helix
  - Interdisciplinary approaches and results
  - Modeling the Double Helix
- Functions of the Double Helix
  - Genetic Blueprint: Replication
  - Genetic Code: Protein Synthesis
  - Organization of DNA in higher organisms
- Designing with the Double Helix
  - Cloning: restriction endonucleases
  - Genome editing: CRISPR
  - Designer DNAs



# Invitation to Develop and Share Lessons

PDB-101 | Molecule of the Month | Browse | Learn | Global Health | Teach | Geis Archive | Events | More

PDB-101 — Molecular explorations through biology and medicine

Search Molecule of the Month articles and more — Go

Educational portal of RCSB PDB Protein Data Bank

**Curriculum Modules** | Overview | Discussion Forum | Contact Us | Teacher Log In

## Curriculum Modules


Biomolecular Structures and Models


Diabetes at a Molecular Level


Molecular Immunology

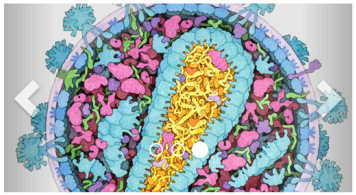

Molecular View of HIV/AIDS

## Skills

Try out and master some basic skills for learning in science - click to open the "Skills" box!

Skills

## Overview



RCSB PDB Curriculum Modules include authentic data from existing public resources, hands-on activities, teaching materials, individual and group activities and assessment suggestions. They were developed through the collaboration and participation of scientists, curriculum design experts, educators, clinicians and local teachers.

## Take a Video Tour of the Curriculum Modules