N Bases / Nucleosides / Nucleotides / Nucleic Acid Structures (Review)

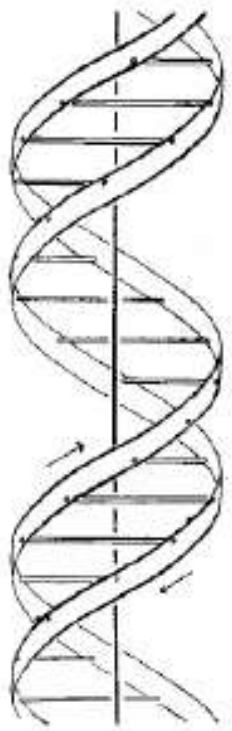
Goals for this review unit:

1. **Recognize the common building blocks of nucleic acids: names / 1-letter abbrev.**
2. **Nomenclature for nucleosides and nucleotides (structure ofATP)**
3. **Primary structures of RNA and DNA**
4. **Conformations in DNAs**
5. **Characteristics of B-DNA, A-DNA and Z-DNA**
6. **Denaturation of DNA**
7. **Features of RNA / Functions of RNA**
8. **DNA Sequencing (Maxam – Gilbert vs. Sanger Dideoxy)**

The Birth of Molecular Biology: DNA Structure

inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

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 di-ester 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



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

Nature – 1953

Nature – 2001

The Nobel Prize in Physiology or Medicine 1962

"for their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material"

Francis Harry Compton Crick

1/3 of the prize

United Kingdom

MRC Laboratory of Molecular Biology
Cambridge, United Kingdom

b. 1916
d. 2004

James Dewey Watson

1/3 of the prize

USA

Harvard University
Cambridge, MA, USA

b. 1928

Maurice Hugh Frederick Wilkins

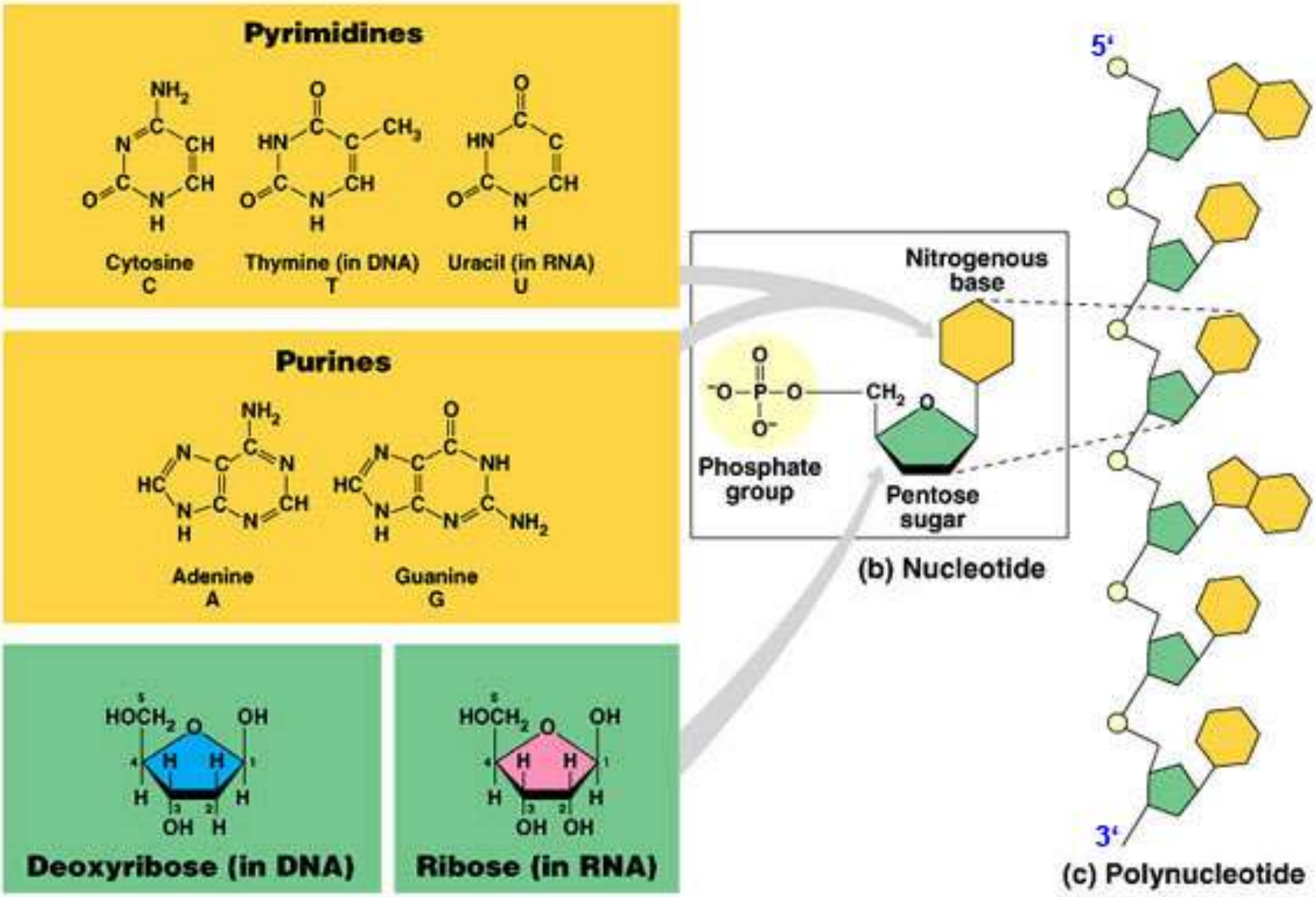
1/3 of the prize

United Kingdom and New Zealand

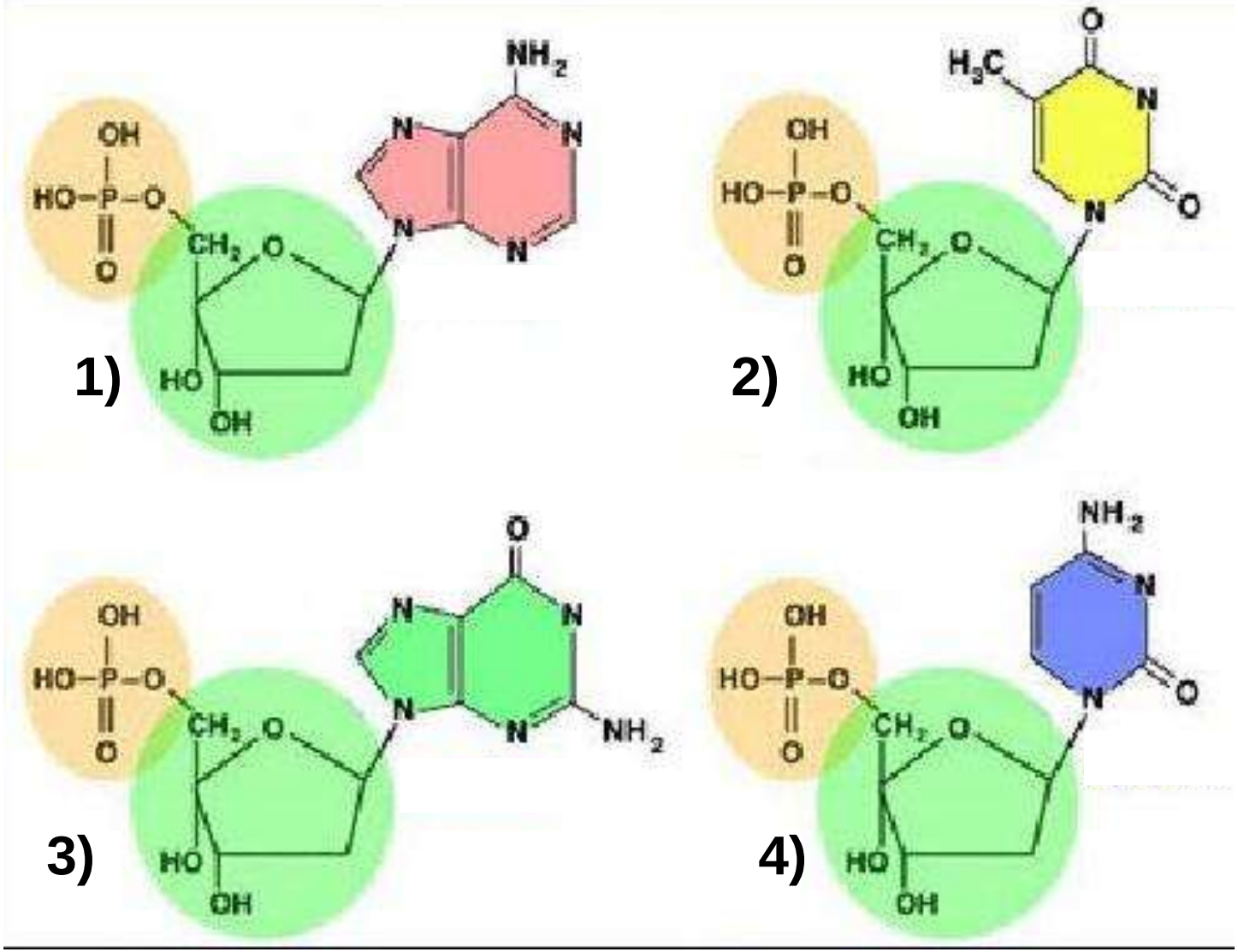
London University
London, United Kingdom

b. 1916
(in Pongaroa, New Zealand)
d. 2004

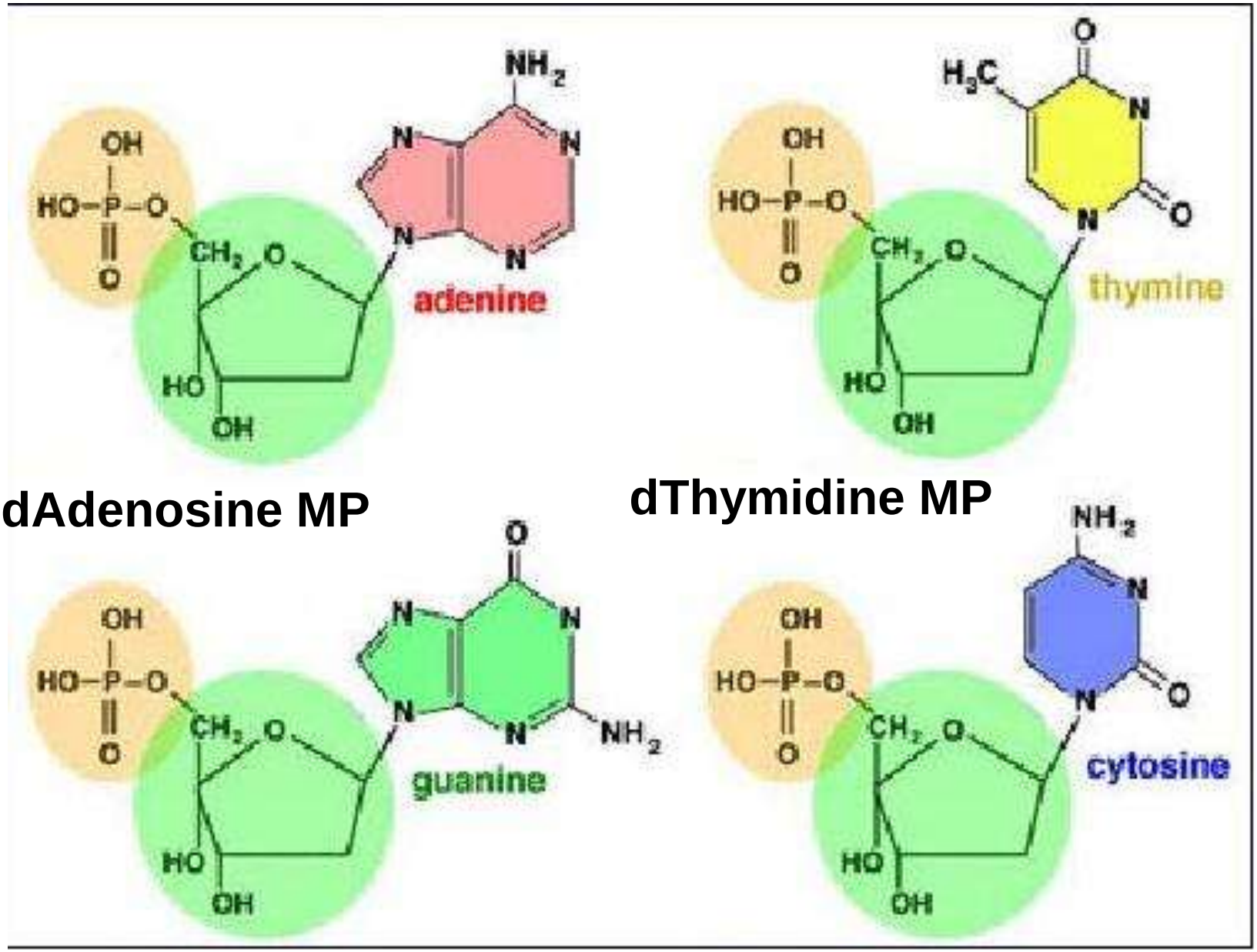
Left to right: Maurice Wilkins, John Steinbeck, John Kendrew, Max Perutz, Francis Crick and Jim Watson after the Nobel Ceremony in Stockholm in December 1962.



What is the correct name for #2?



A) Guanosine
B) deoxyThymine MP
C) Thymine
D) dexoyThymidine MP
E) Cytidine MP



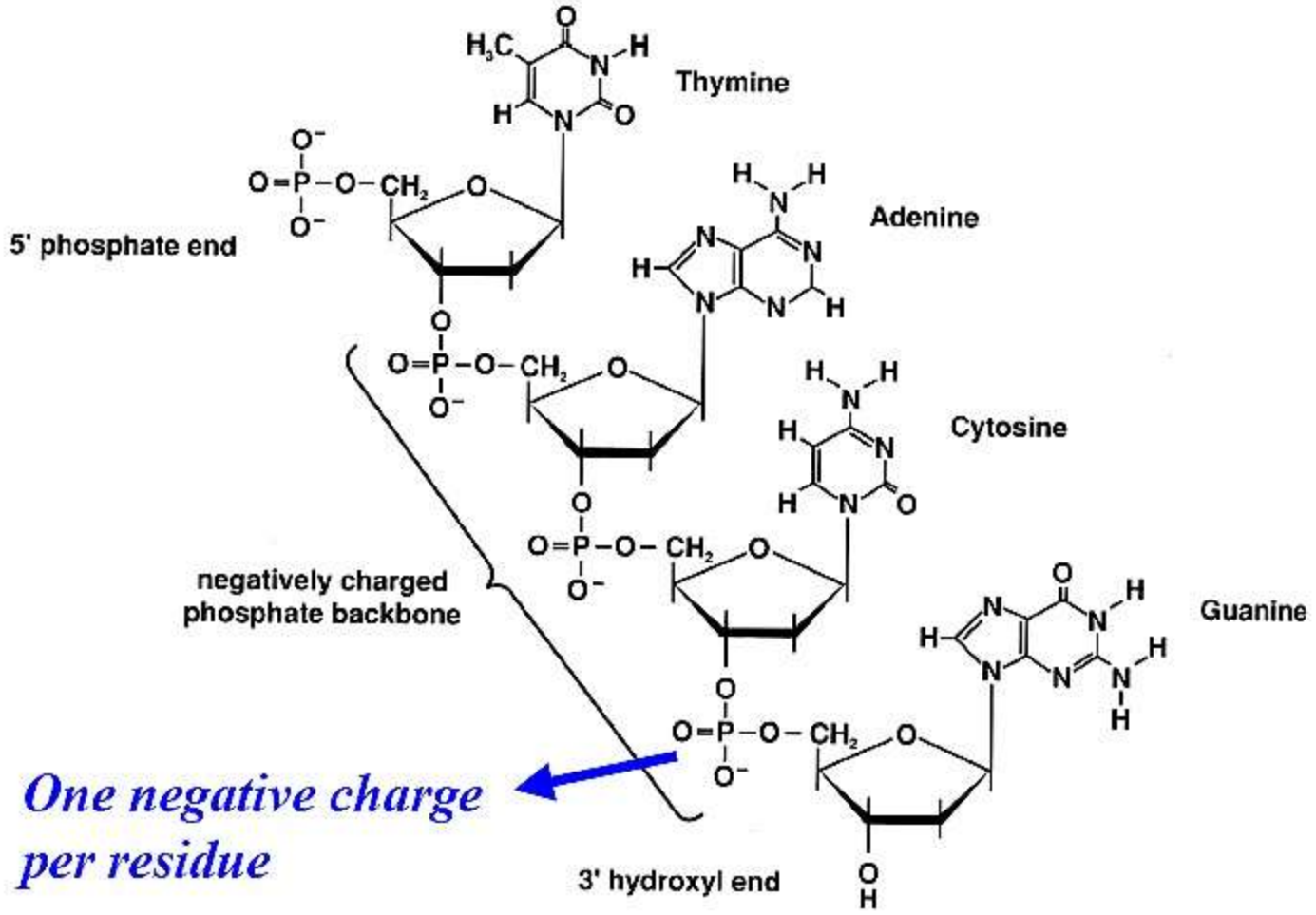
dAdenosine MP

dThymidine MP

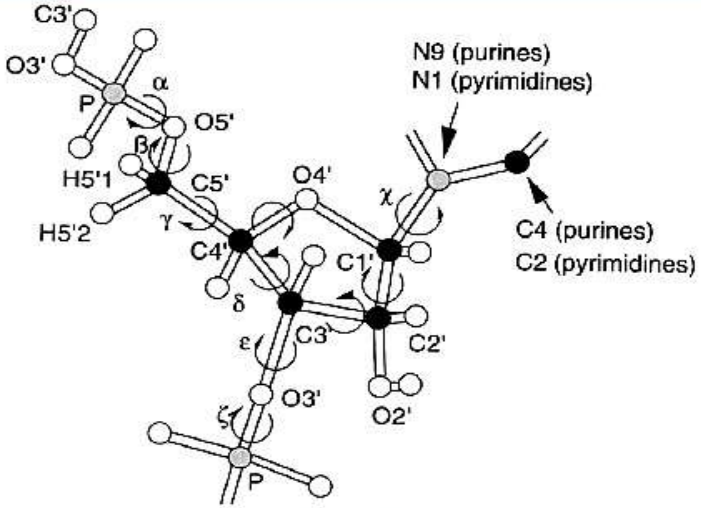
dGuanosine MP

dCytidine MP

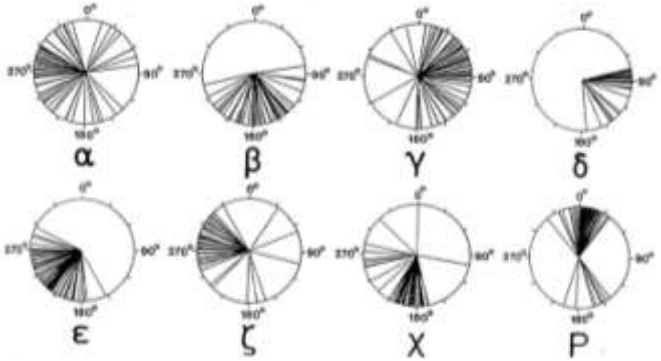
DNA primary structure

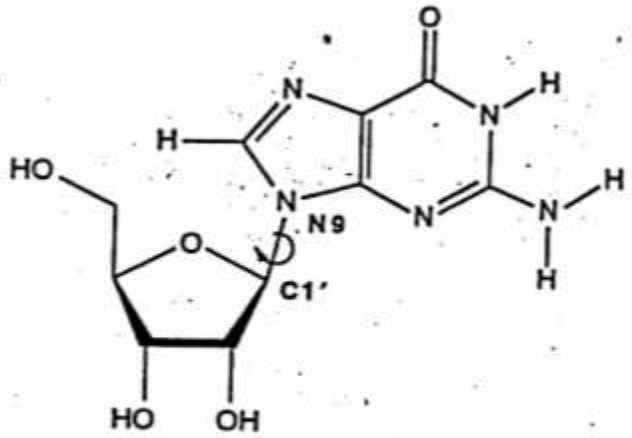


Rotational angles of phosphodiester chain

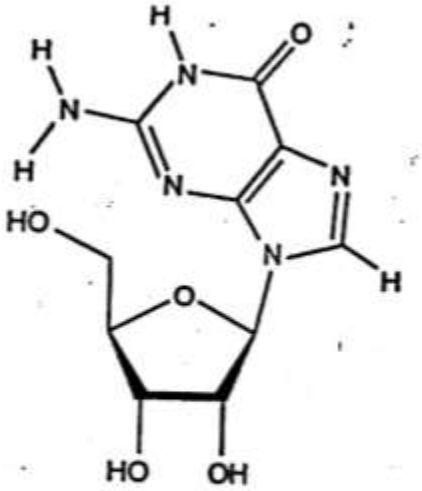


	B-DNA	A-DNA	Z-DNA C	Z-DNA G
$\alpha(\omega)$	−41	−90	138	100
$\beta(\phi)$	136	211	−94	−108
$\gamma(\psi)$	38	47	80	−70
$\delta(\psi')$	139	83	48	−130
$\varepsilon(\phi')$	−133	−185	180	−140
$\zeta(\omega')$	−57	−45	−170	56
χ	78	27	20	−100

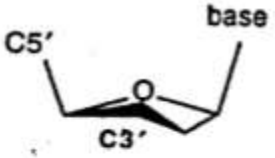




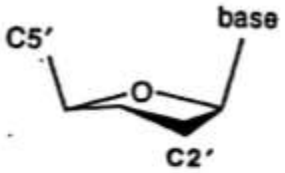
guanosine-anti



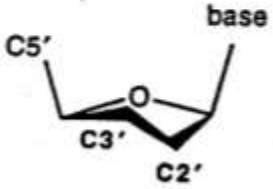
guanosine-syn



C3′ endo

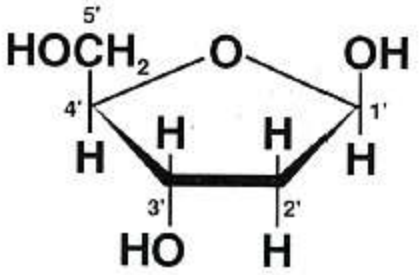


C2′ exo

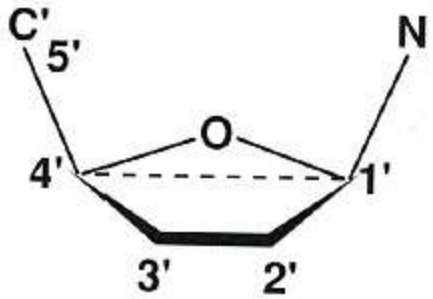


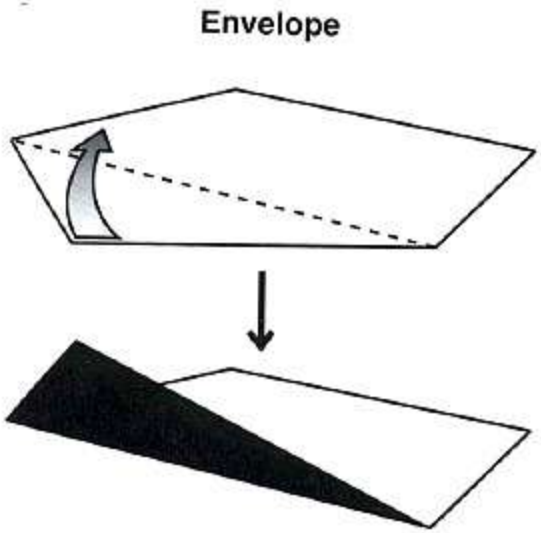
C3′ endo–C2′ exo

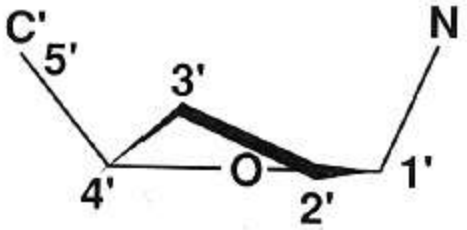
Sugar pucker in DNA



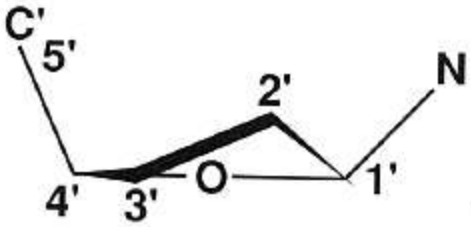
β-D-2-Deoxyribose





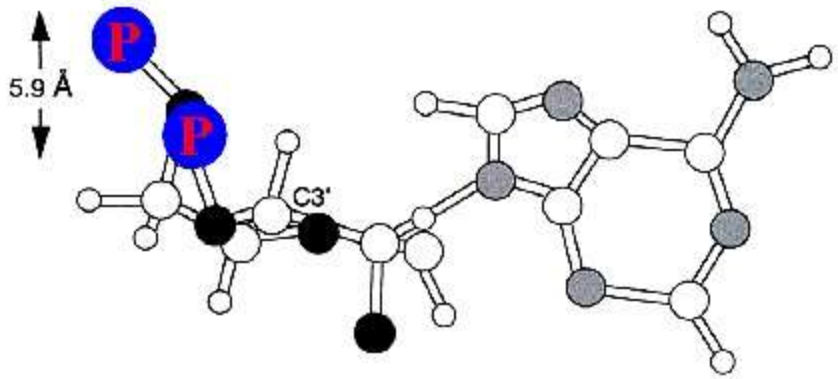


C3' endo

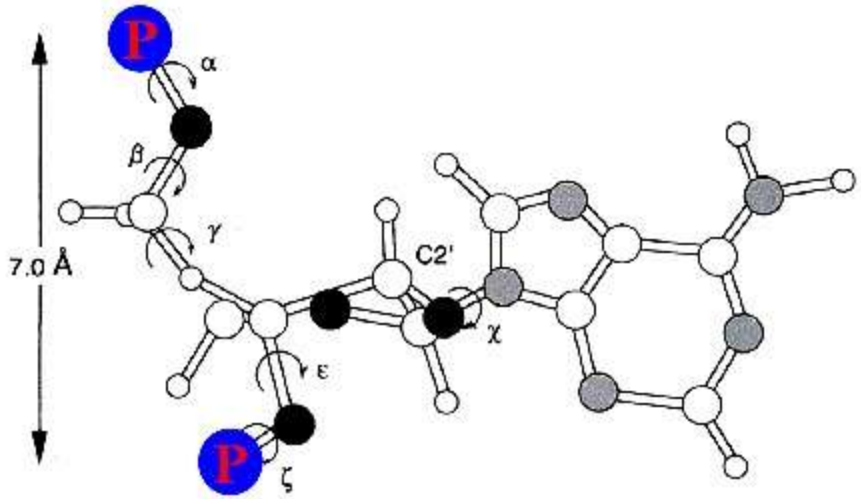


C2' endo

Sugar pucker in DNA



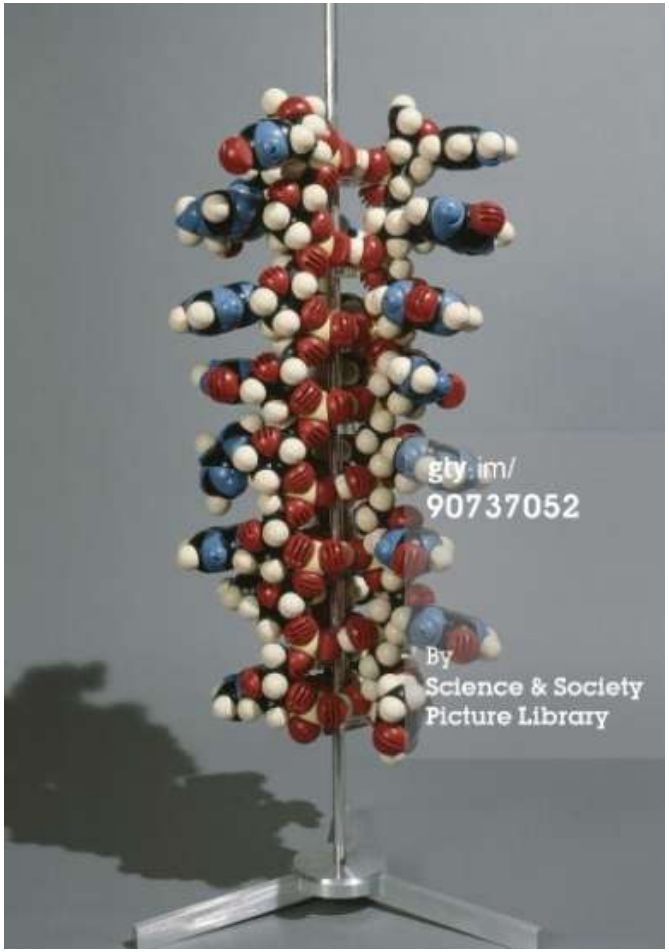
C3' endo

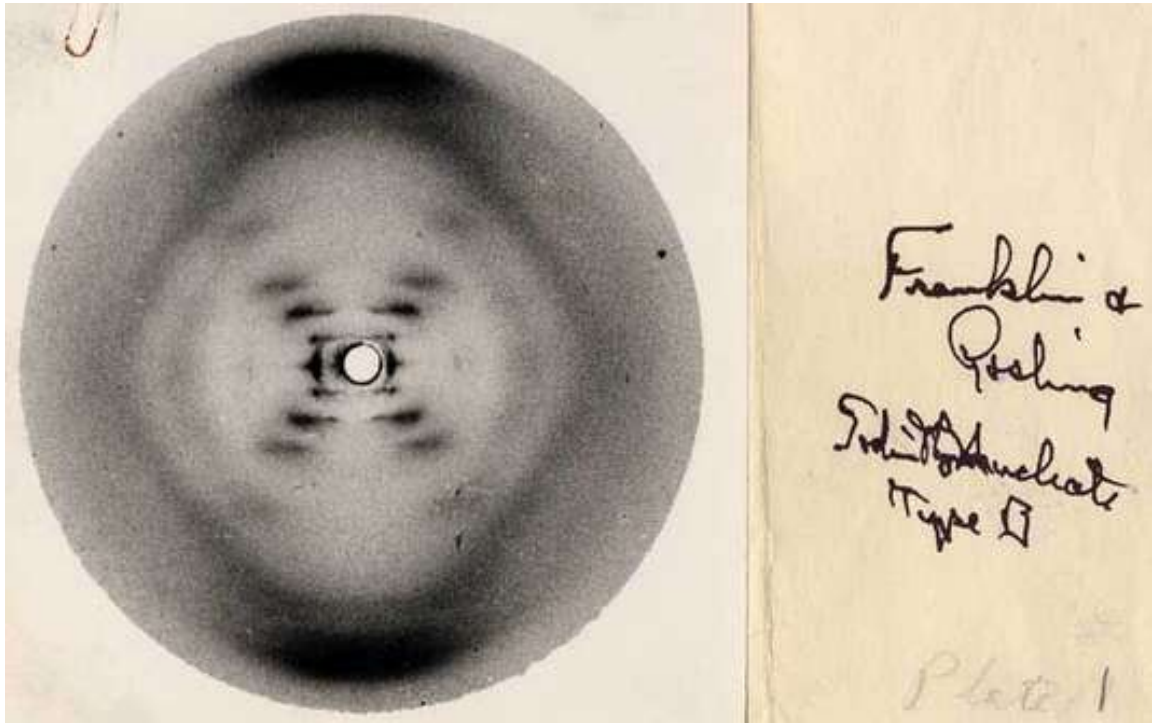


C2' endo

Pauling triple helix model

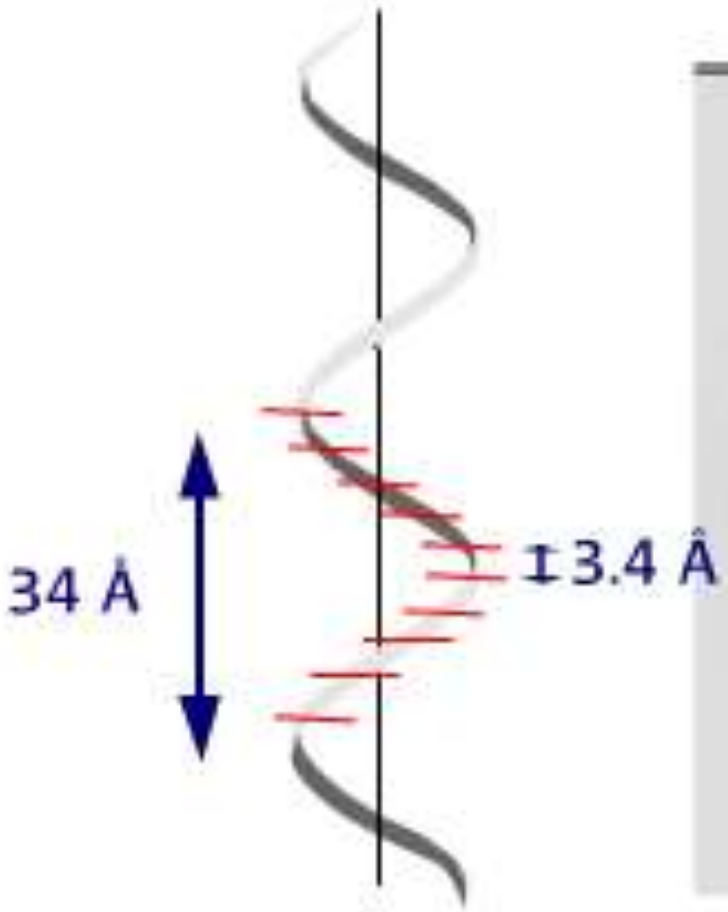
One of the failed hypothetical models of DNA is Linus Pauling's triple helix model. This structure would be unstable under normal cellular conditions.

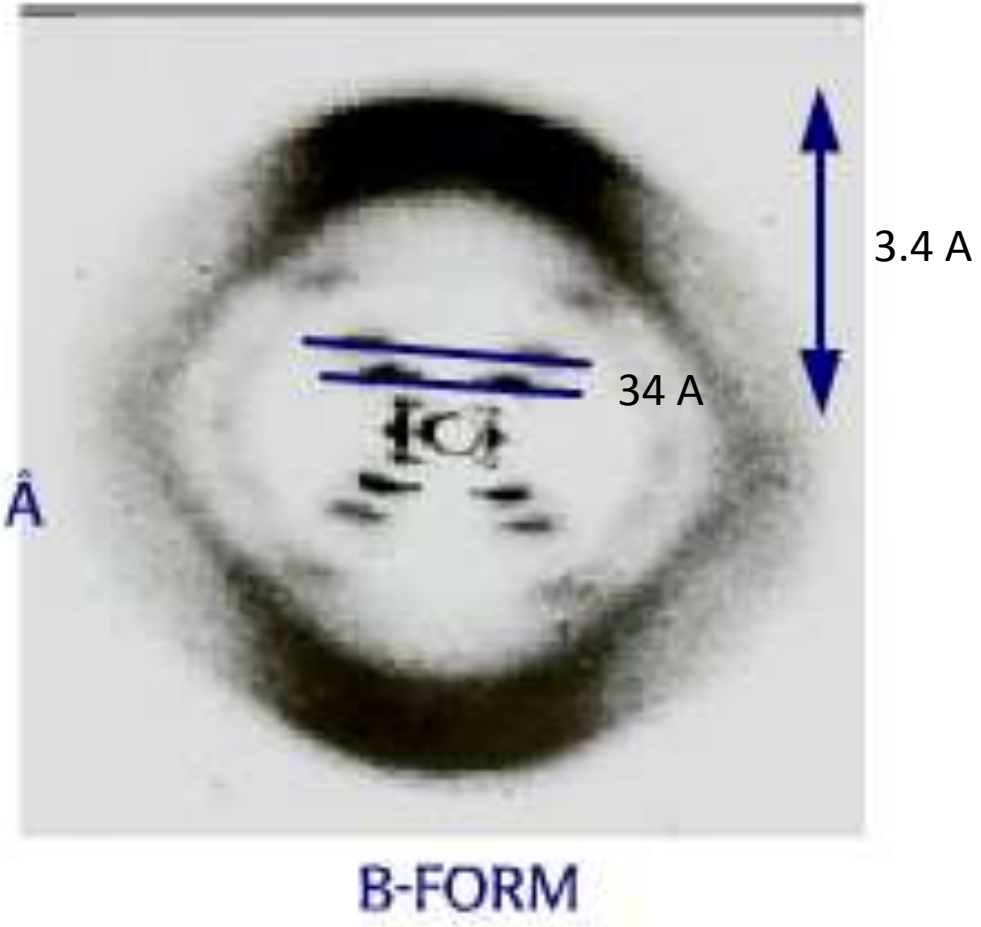




http://www.hhmi.org/biointeractive/dna/DNAi_pauling_triple_helix.html

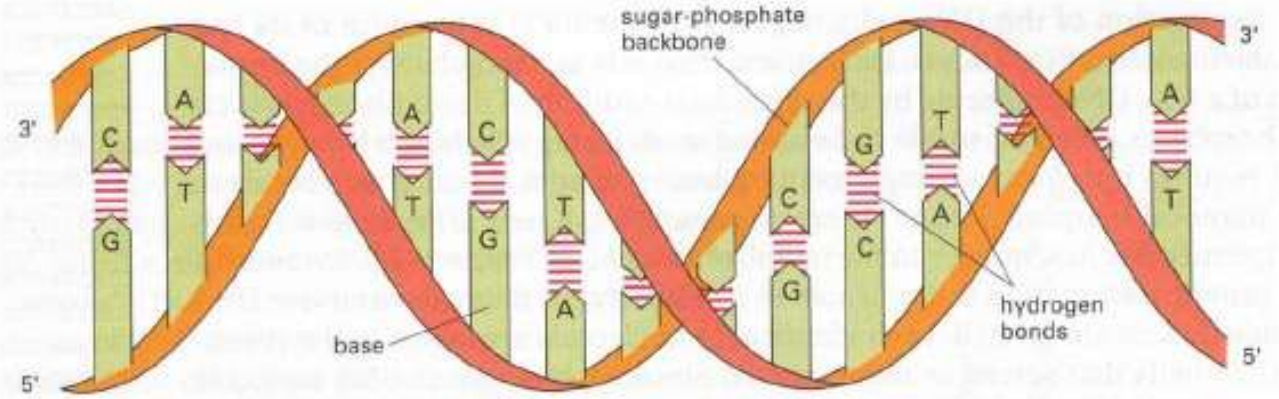
Rosalind Franklin and Raymond Gosling's B-DNA X-ray Image (interpretation)



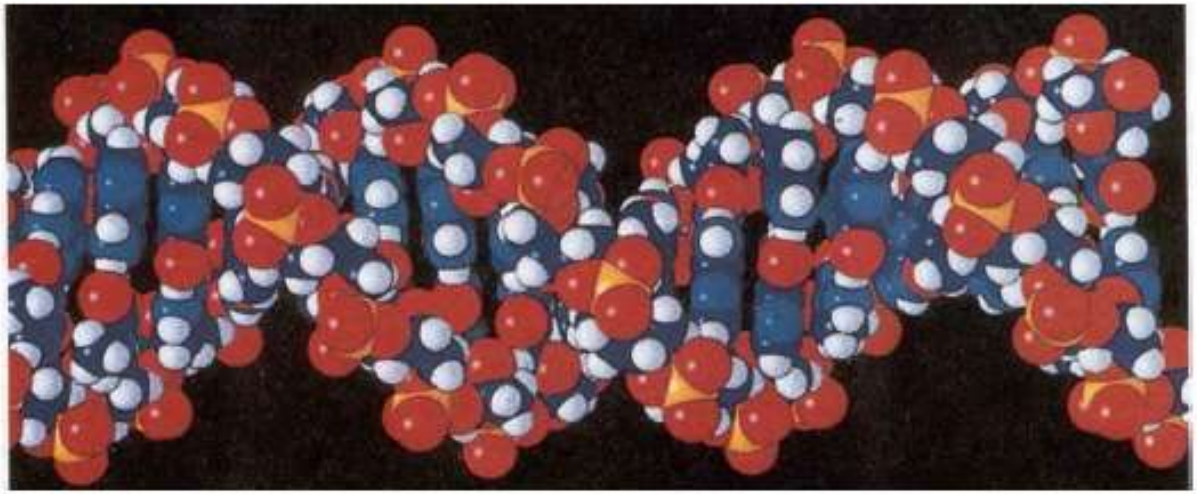


http://www.dnalc.org/view/15014-Franklin-s-X-ray-diffraction-explanation-of-X-ray-pattern-.html

Double stranded DNA

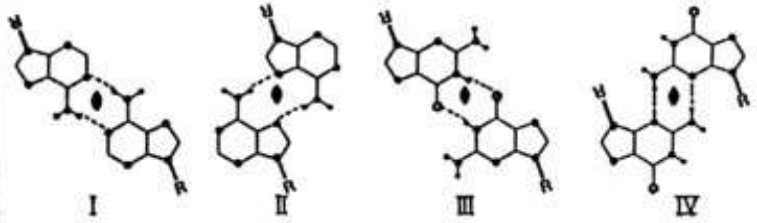


- *Two single stranded DNA paired by Hydrogen bonds.*
- *Helical structure*

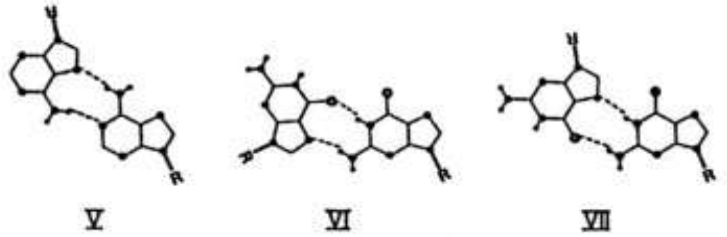


homo purine

symmetric

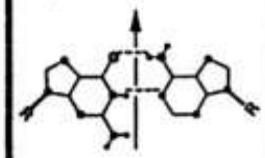


asymmetric

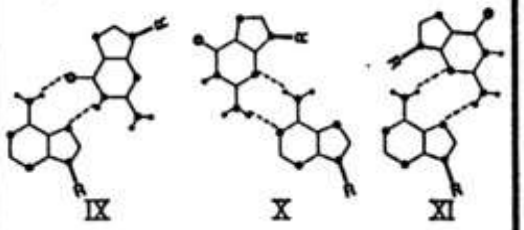


hetero purine

symmetric

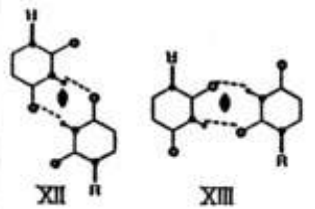


asymmetric

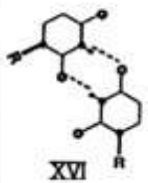


homo pyrimidine

symmetric

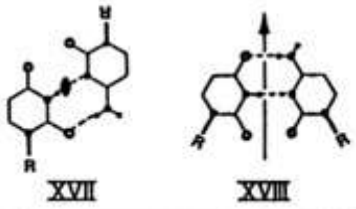


asymmetric



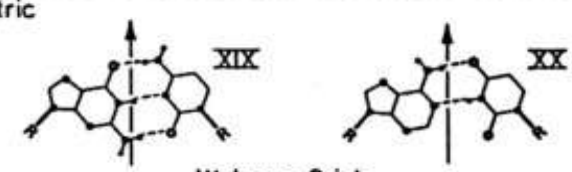
hetero pyrimidine

symmetric only



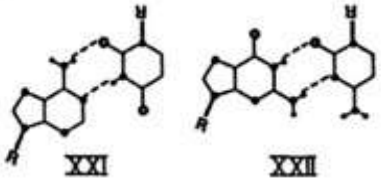
purine - pyrimidine

symmetric

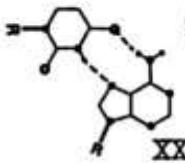


Watson - Crick

asymmetric

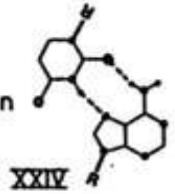


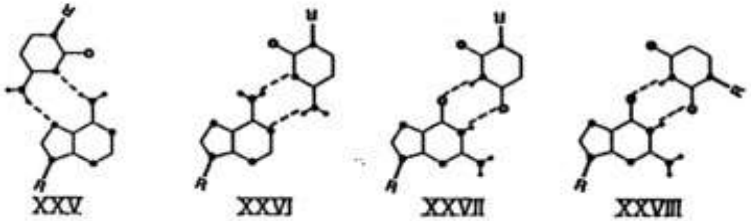
reversed
Watson - Crick



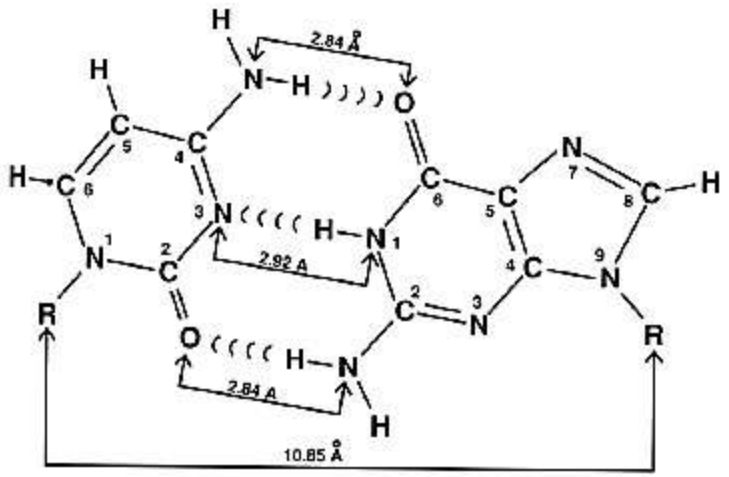
Hoogsteen

reversed
Hoogsteen

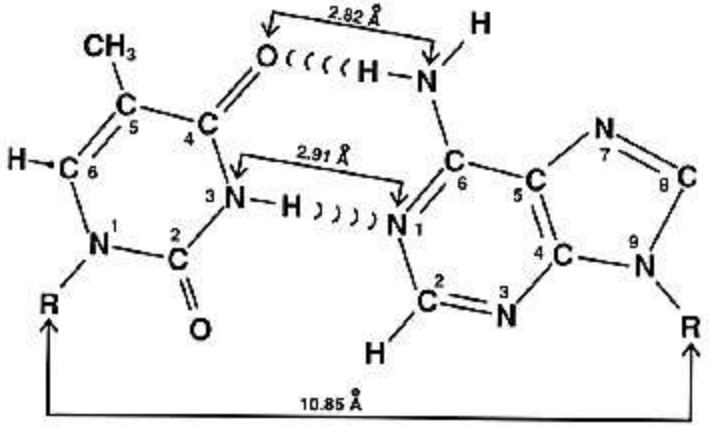




Watson-Crick base pairs



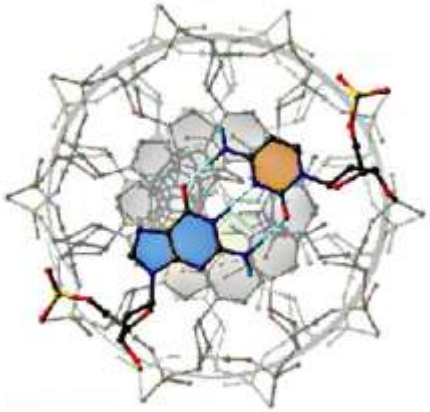
G-C base pair



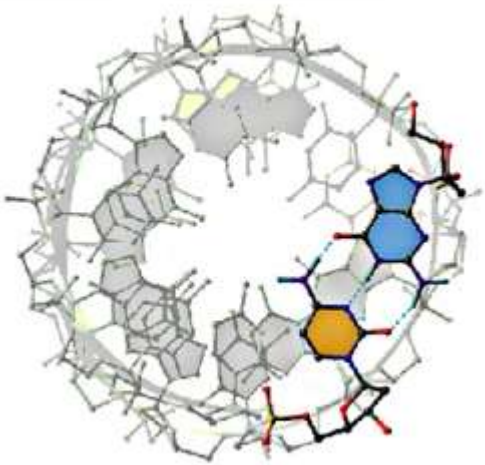
A-T base pair

HHMI: http://www.hhmi.org/biointeractive/dna/DNAi_watson_basepairing_video.html

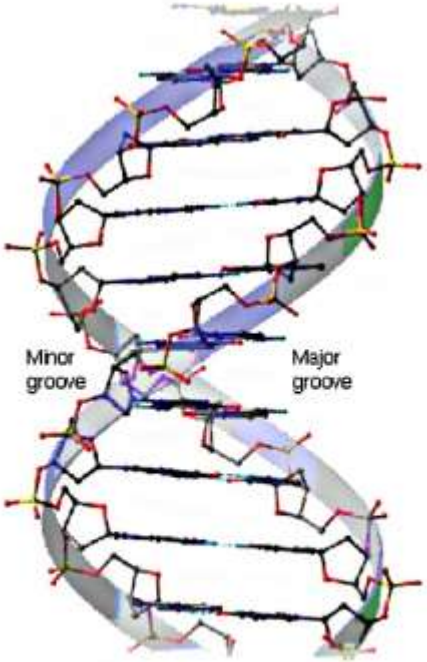
A and B Double Helices



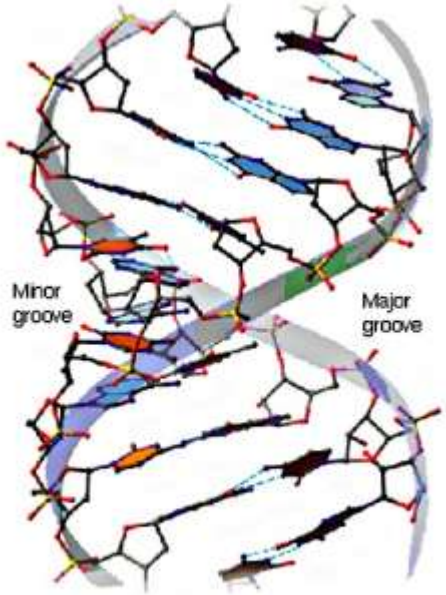
(a) B-DNA, end-on view



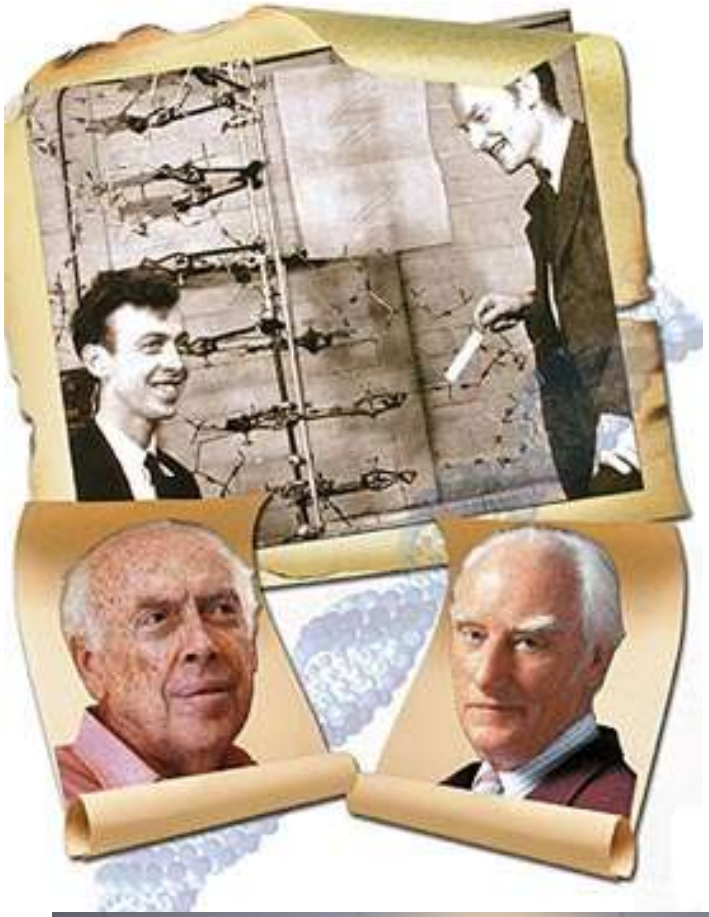
(c) A-DNA, end-on view



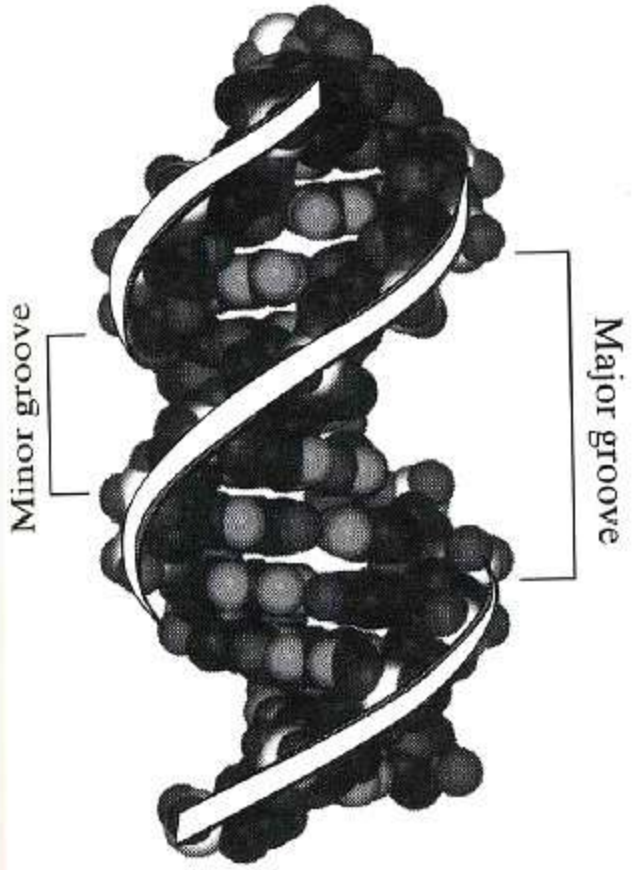
(b) B-DNA, side view



(d) A-DNA, side view



Structure of double stranded DNA (B-DNA)



B-DNA

Right handed helix

10.5 residue per turn

Helix pitch = 34Å

Base pair tilt-helical axis = -6°

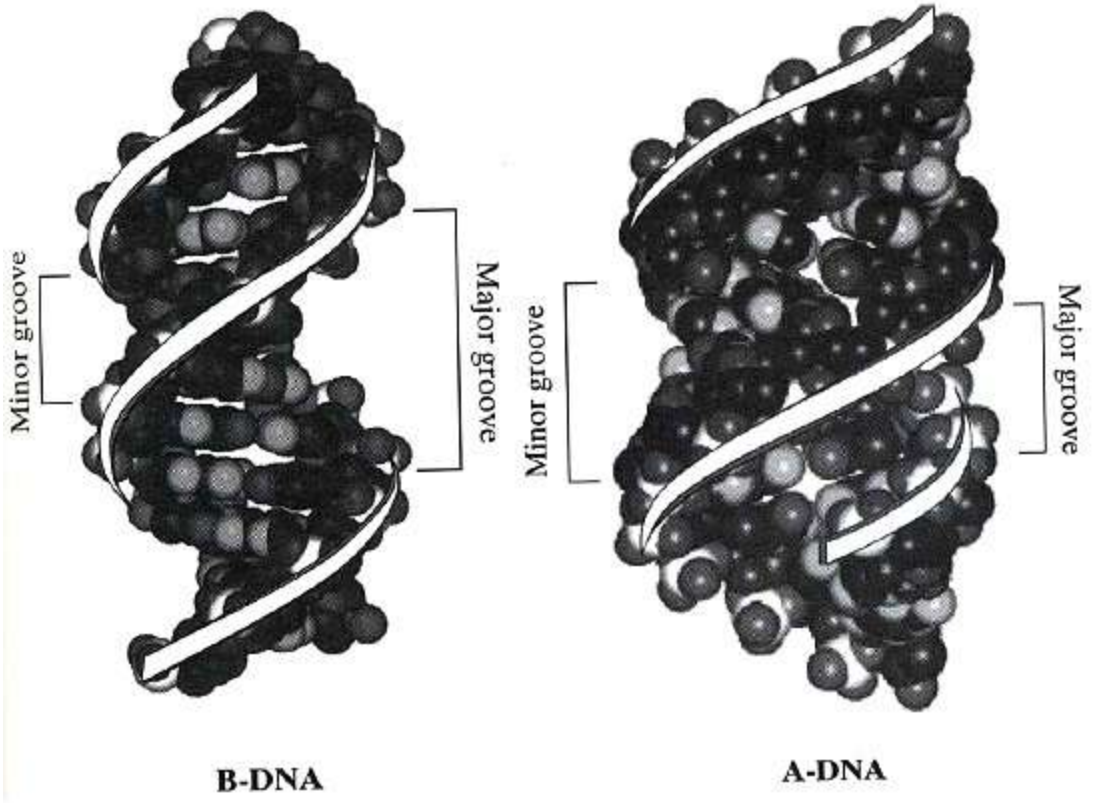
Diameter = 20Å

Sugar pucker
dA, dT, dC, dA: C2' endo

Glycosidic bond
dA, dT, dC, dA: anti

Minor grove show base diversity

A-DNA vs. B. DNA



A - DNA

11 residue per turn

Helix pitch = 28Å

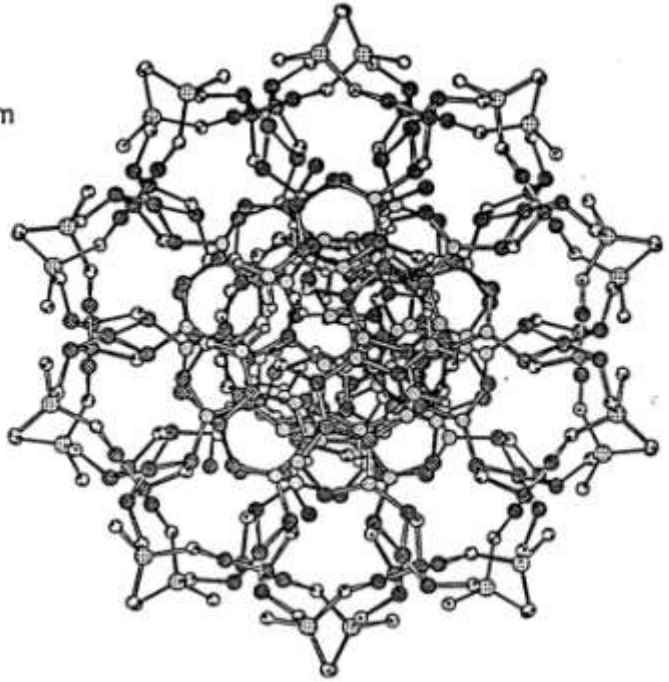
Base pair tilt = 20°

Diameter = 23Å

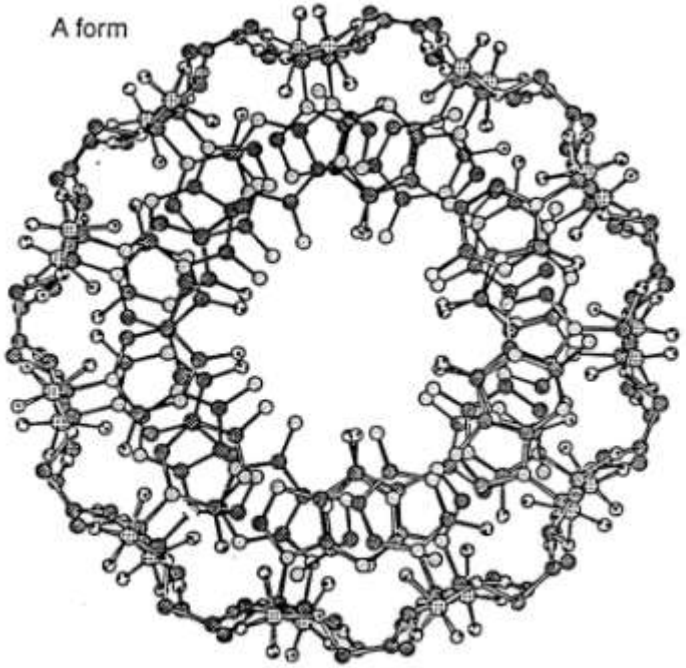
Sugar pucker
C3' endo

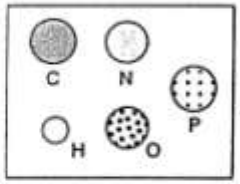
Glycosidic bond
anti

B form

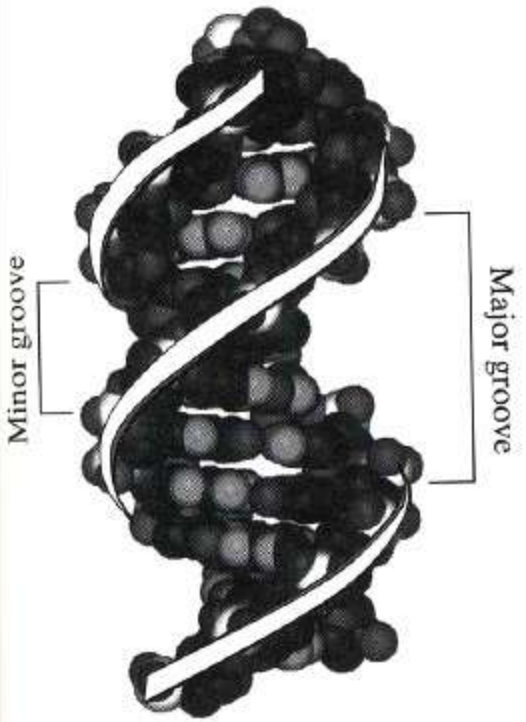


A form

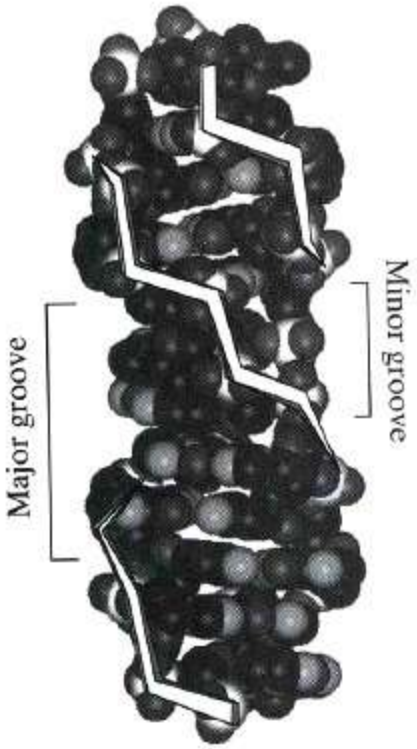




Z-DNA vs. B-DNA



B-DNA



Z-DNA

Z - DNA

left handed helix

12 residue per turn

Helix pitch = 45Å

Base pair tilt = 7°

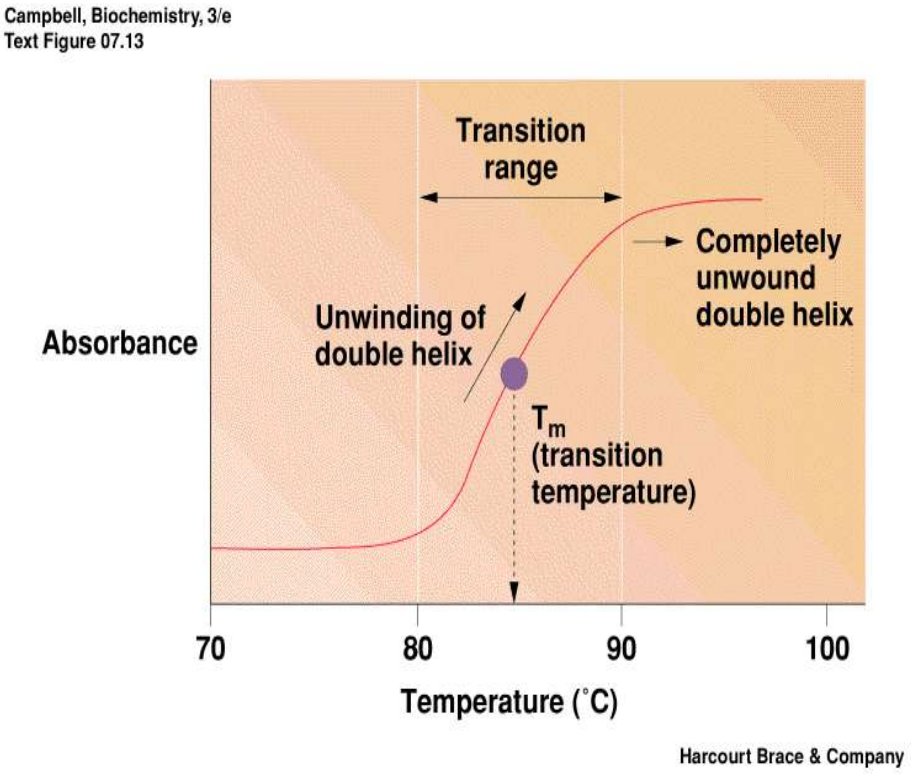
Diameter = 18Å

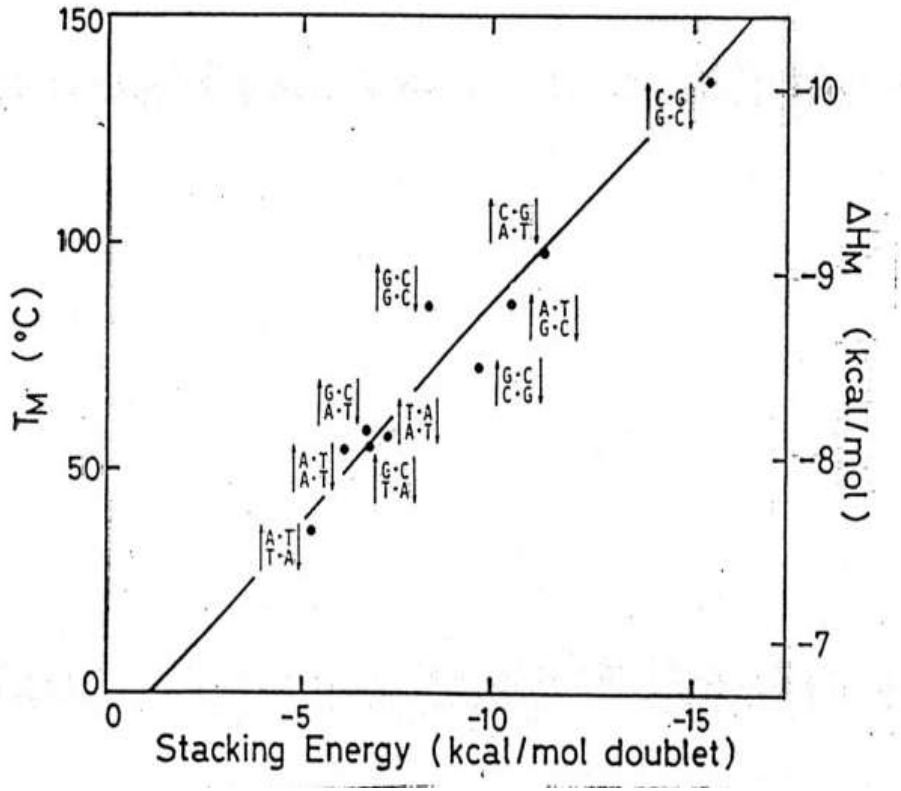
Sugar pucker
dA, dT, dC: C2' endo
dG: C3' endo

Glycosidic bond
dA, dT, dC: anti
dG: syn

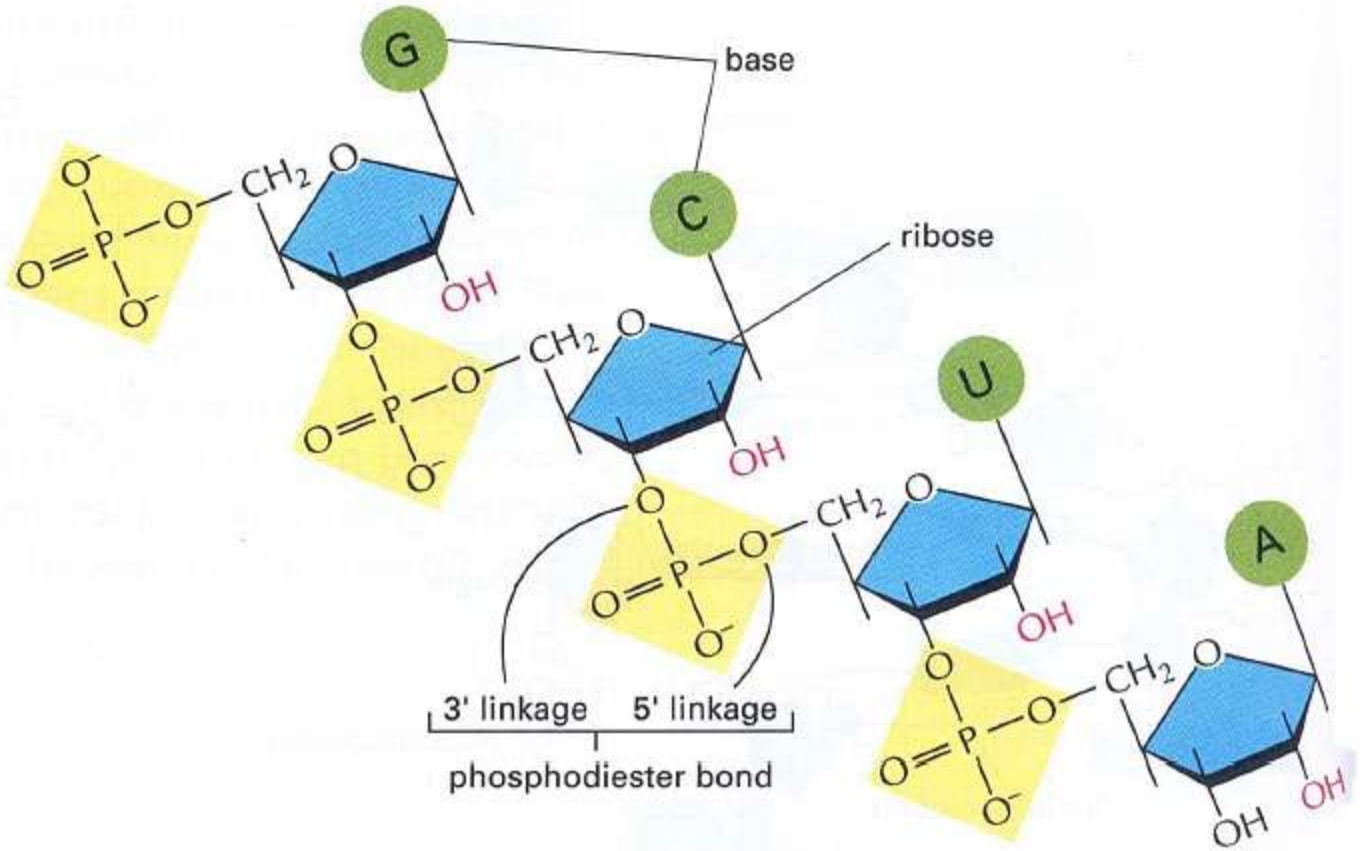
Denatuted DNA:

Heat denaturation of DNA is called "melting," The purine and pyrimidine bases exhibit very strong p-p* transitions around 260 nm. *E. coli* DNA absorption is only about 60% of that predicted from the weighted average spectrum based on its composition, this loss of intensity is called ***hypochromism***. Since the absorpance goes up as DNA "unwinds", it can be used to monitor the unstacking of DNA.

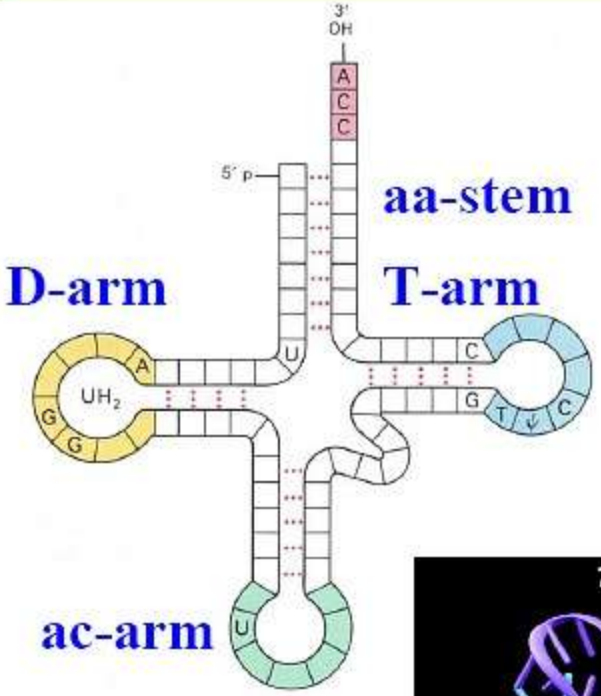


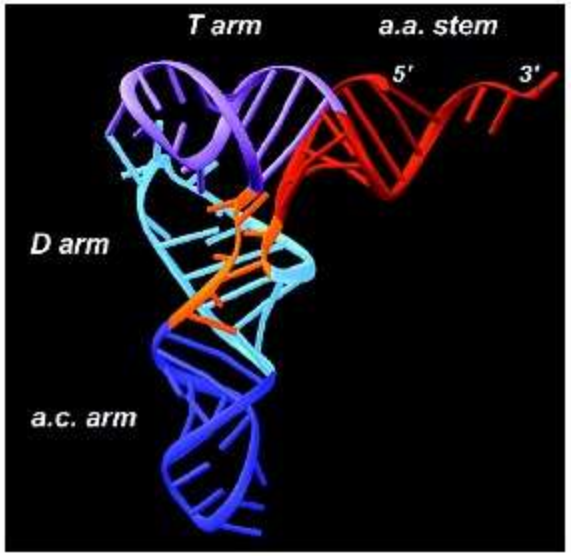


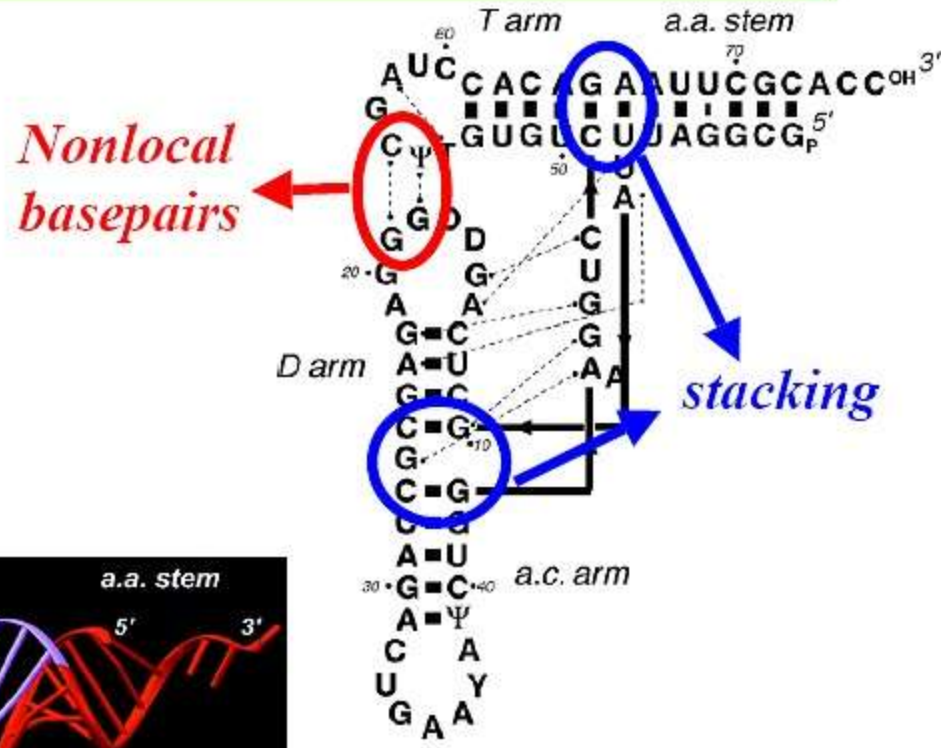
RNA primary structure



Transfer-RNA







THE SPECTACULAR ARCHITECTURE OF THE RIBOSOME AND CLUES ABOUT ITS ORIGIN

The following slides on ribosome structure are taken from material presented at the IUCr Congress in Madrid 2011.

ADA YONATH

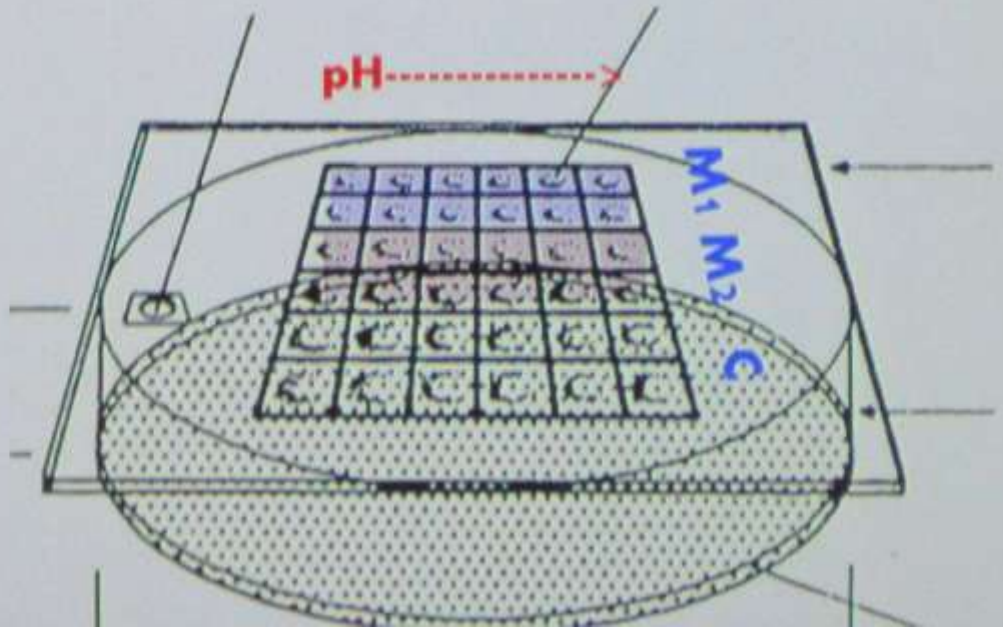
WEIZMANN INSTITUTE OF SCIENCE

Ribosomes have been considered non-crystallizable

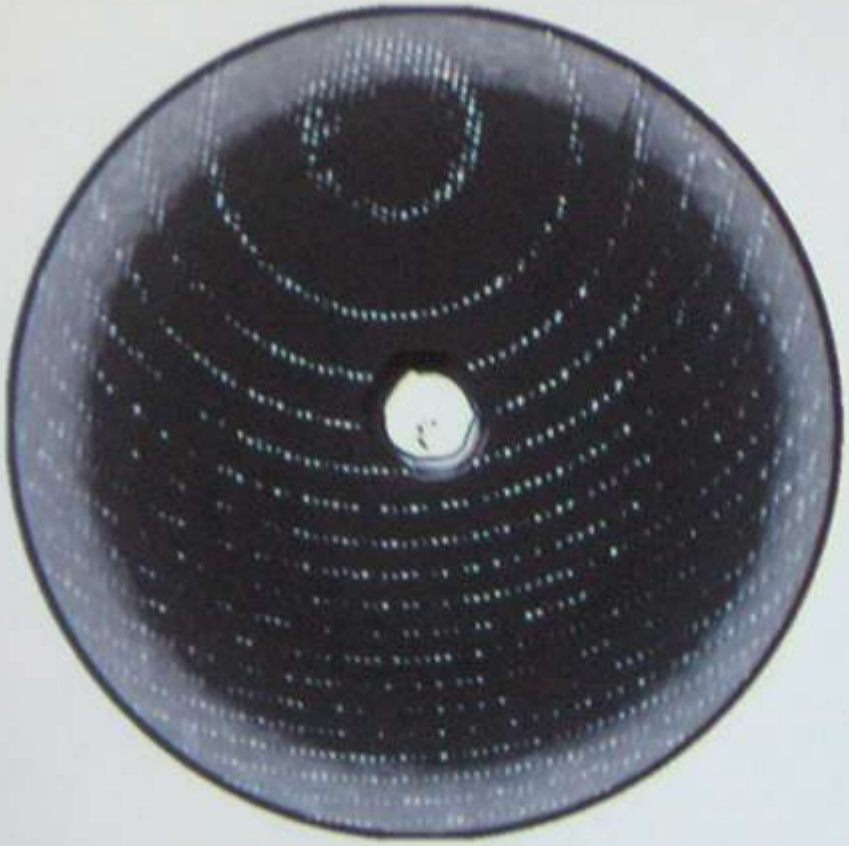
owing to:

- their high degree of internal mobility
- considerable flexibility
- functional heterogeneity
- marked tendency to deteriorate
- chemical complexity
- large size and asymmetric nature

Micro crystals, not useful for crystallographic analysis, but indicating high potential, were obtained after screening of 25,000 conditions (within 6 months)

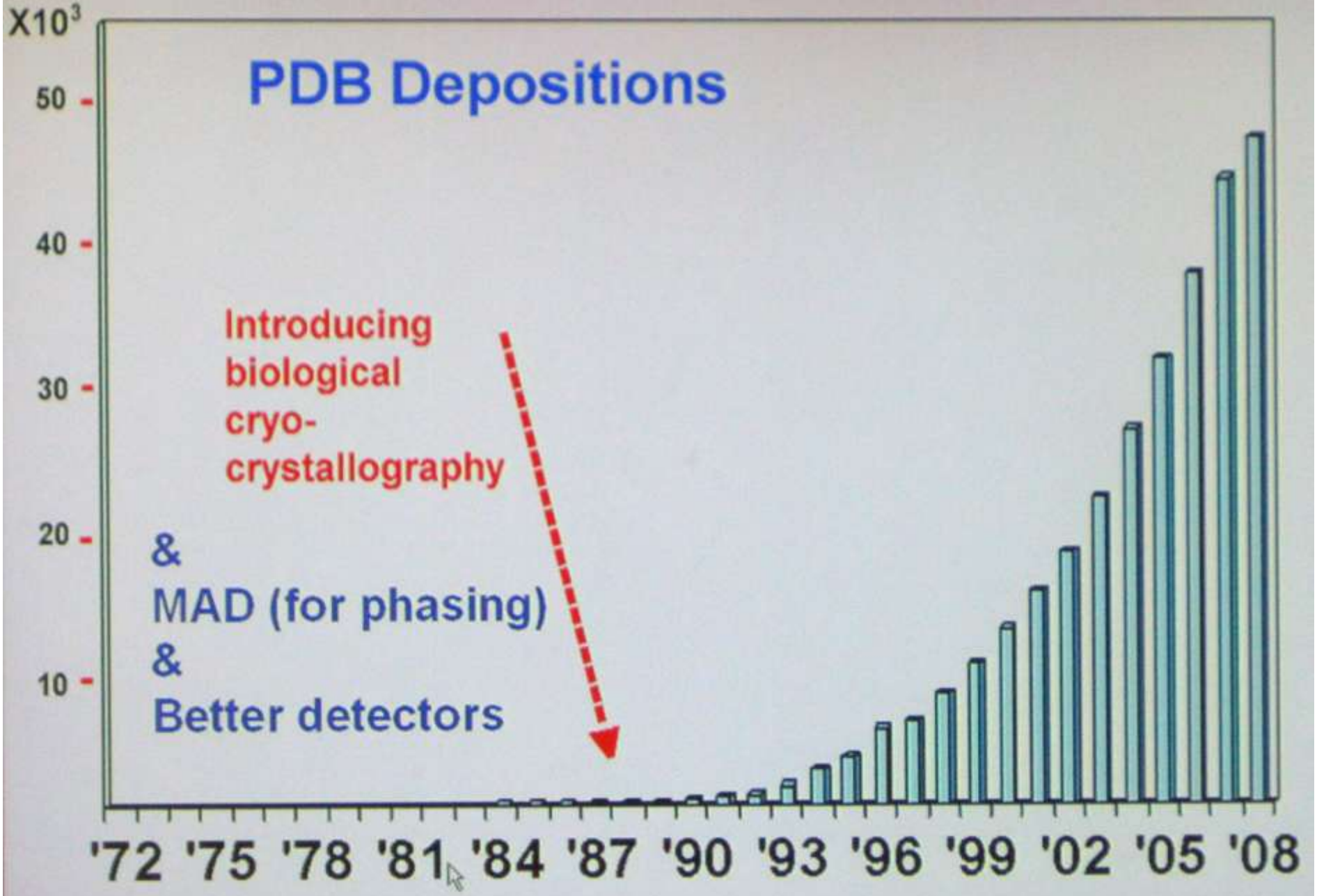


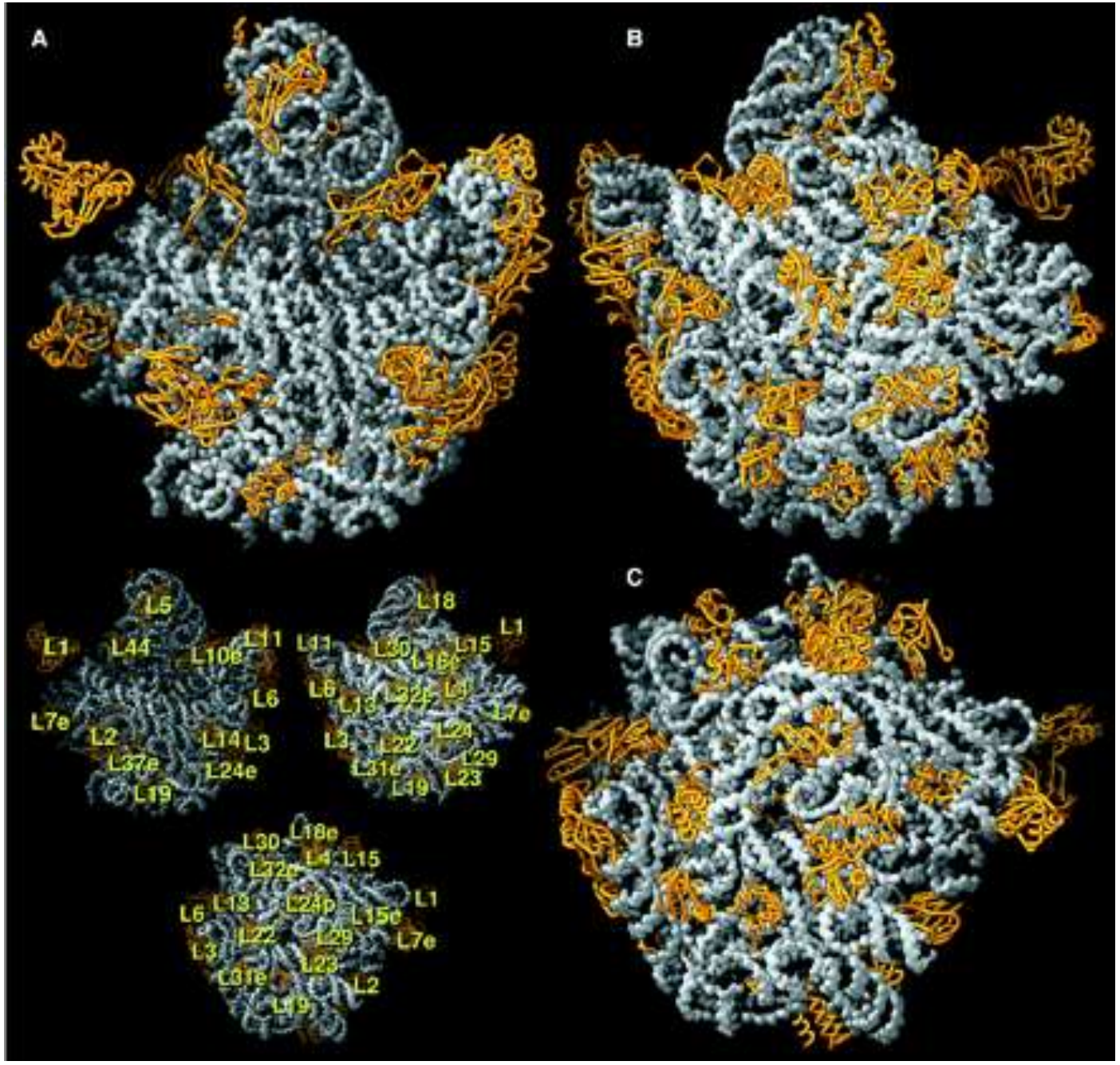
H50S diffraction, 1986



Time = 0

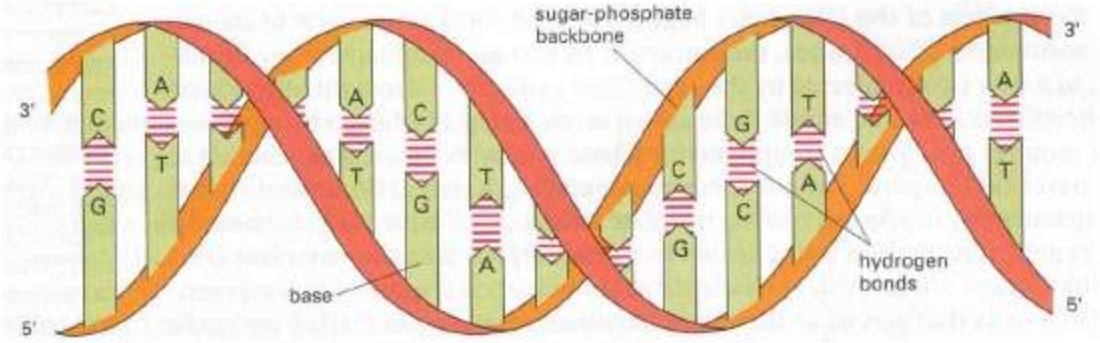
After Exposure of 0.1 second


X10^3
PDB Depositions
50
40
30
20
10
Introducing
biological
cryo-
crystallography
&
MAD (for phasing)
&
Better detectors
'72 '75 '78 '81 '84 '87 '90 '93 '96 '99 '02 '05 '08


A
B
C
L5
L1
L44
L10e
L11
L41
L18
L1
L30
L15
L16e
L6
L13
L32e
L9
L7e
L7e
L2
L14
L3
L3
L22
L24
L37e
L24e
L31e
L29
L19
L49
L23
L18e
L30
L4
L15
L32e
L1
L6
L13
L24p
L15e
L22
L29
L7e
L3
L23
L31e
L2
L19

Head Full of Ribosomes

Genetic information



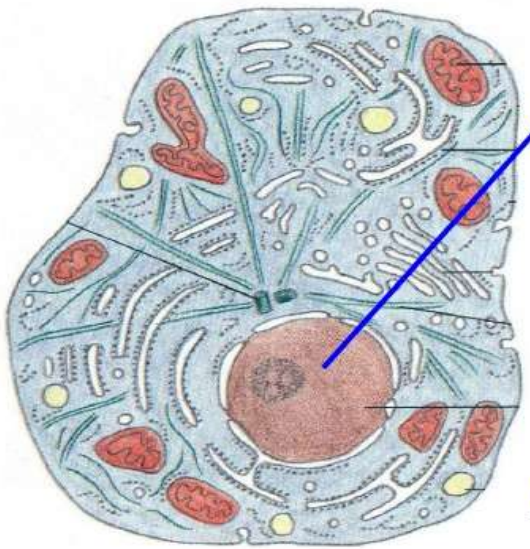
... G T A C T G A A C G C A G G T...

Genetic code

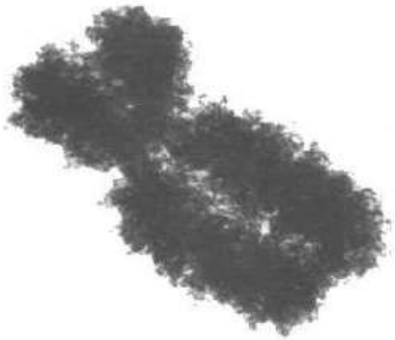
Human being: ~ 3,000,000,000 base-pairs
~ 30,000 – 40,000 Genes

(Public Human Genome Project and Celera Genomics)

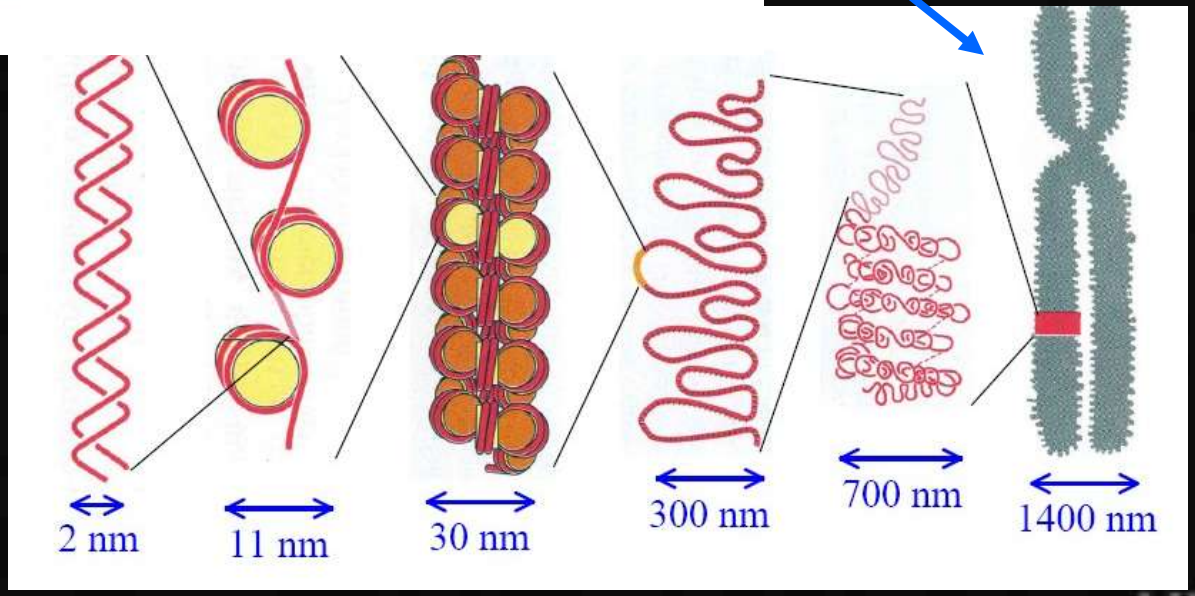
Chromosome



Chromosomes



Human being has 24 chromosomes.



HHMI: http://www.hhmi.org/biointeractive/dna/DNAi_packaging_vo2.html

Sequencing DNA

Prior to the mid-1970's no method existed by which DNA could be directly sequenced. Knowledge about gene and genome organization was based upon studies of prokaryotic organisms and the primary means of obtaining DNA sequence was so-called reverse genetics in which the amino acid sequence of the gene product of interest is back-translated into a nucleotide sequence based upon the appropriate codons.

- **Maxam-Gilbert DNA Sequencing**
- **Sanger (didexoy) DNA Sequencing**

The Nobel Prize in Chemistry 1958

"for his work on the structure of proteins, especially that of insulin"

Frederick Sanger

United Kingdom

University of Cambridge
Cambridge, United Kingdom

b. 1918

The Nobel Prize in Chemistry 1980

"for his fundamental studies of the biochemistry of nucleic acids, with particular regard to recombinant-DNA"

"for their contributions concerning the determination of base sequences in nucleic acids"

Paul Berg

1/2 of the prize

USA

Stanford University
Stanford, CA, USA

b. 1926

Walter Gilbert

1/4 of the prize

USA

Harvard University,
Biological Laboratories
Cambridge, MA, USA

b. 1932

Frederick Sanger

1/4 of the prize

United Kingdom

MRC Laboratory of
Molecular Biology
Cambridge, United
Kingdom

b. 1918

Maxam-Gilbert DNA Sequencing

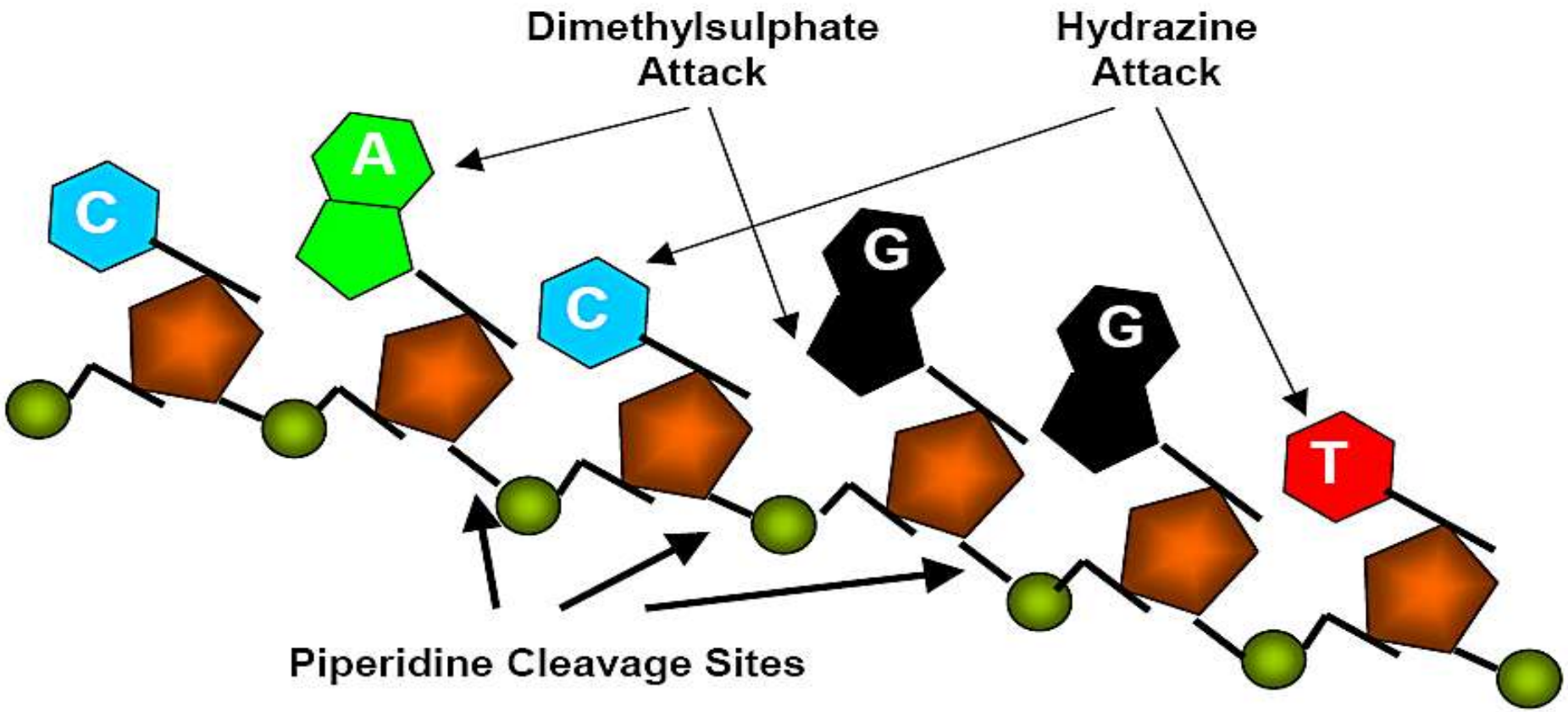


Figure 1. Chemical targets in the Maxam-Gilbert DNA sequencing strategy. Dimethylsulphate or hydrazine will attack the purine or pyrimidine rings respectively and piperidine will cleave the phosphate bond at the 3' carbon.

http://www.idtdna.com/support/technical/TechnicalBulletinPDF/DNA_Sequencing.pdf

Allan Maxam / Walter Gilbert DNA Sequencing

Sequencing single-stranded DNA

Two-step catalytic process:

1) Break glycoside bond between the ribose sugar and the base / displace base

Purines react with dimethyl sulfate

Pyrimidines react with hydrazine

2) Piperidine catalyzes phosphodiester bond cleavage where base displaced

---

“G” - dimethyl sulfate and piperidine

“A + G” - dimethyl sulfate and piperidine in formic acid

“C” - hydrazine and piperidine in 1.5M NaCl

“C + T” - hydrazine and piperidine

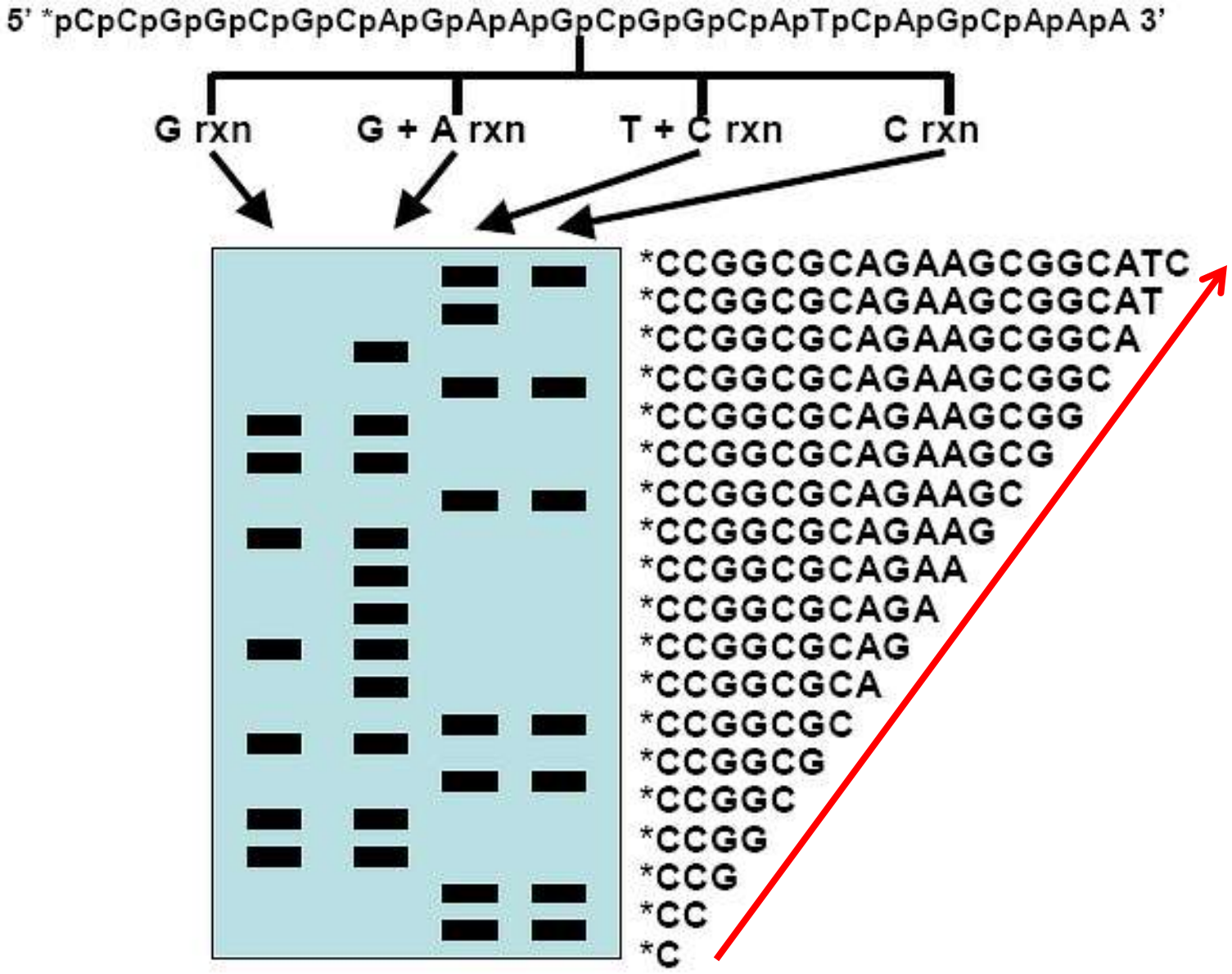


Figure 2. The Maxam-Gilbert manual sequencing scheme. The target DNA is radiolabeled and then split into the four chemical cleavage reactions. Each reaction is loaded onto a polyacrylamide gel and run. Finally, the gel is autoradiographed and base calling proceeds from bottom to top.

Maxam-Gilbert DNA Sequencing

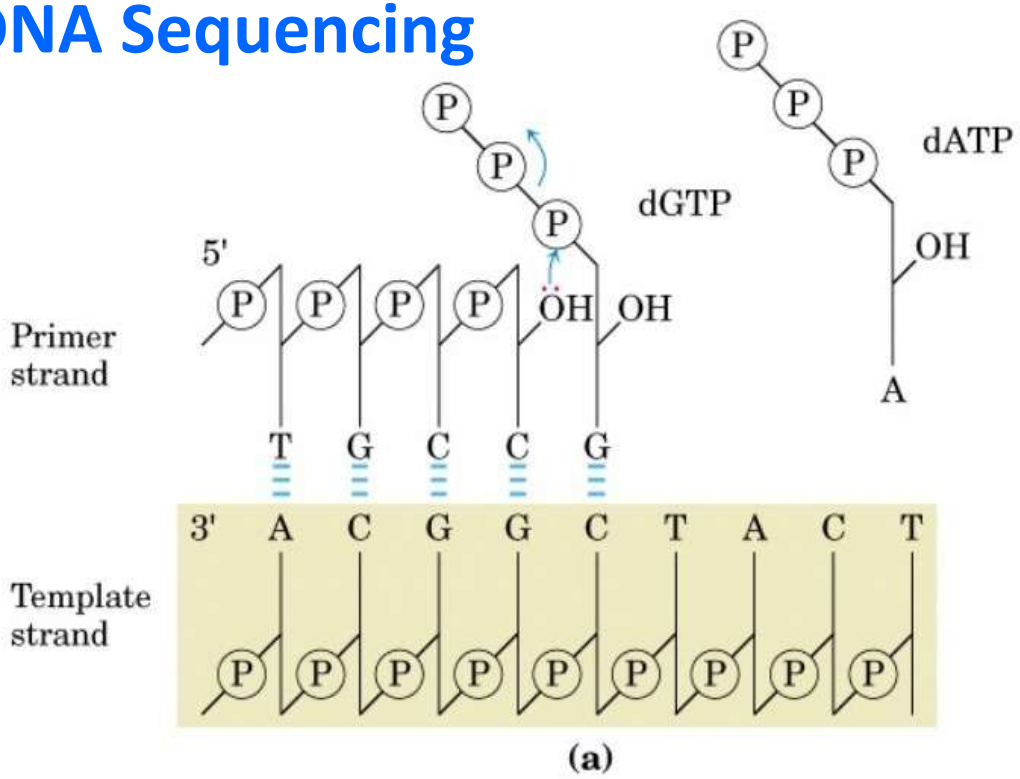
- **200-300 bases of DNA sequence every few days**
- **Use large amounts of radioactive material, ^{35}S or ^{32}P**
- **Constantly pouring large, paper thin acrylamide gels**
- **Hydrazine is a neurotoxin**

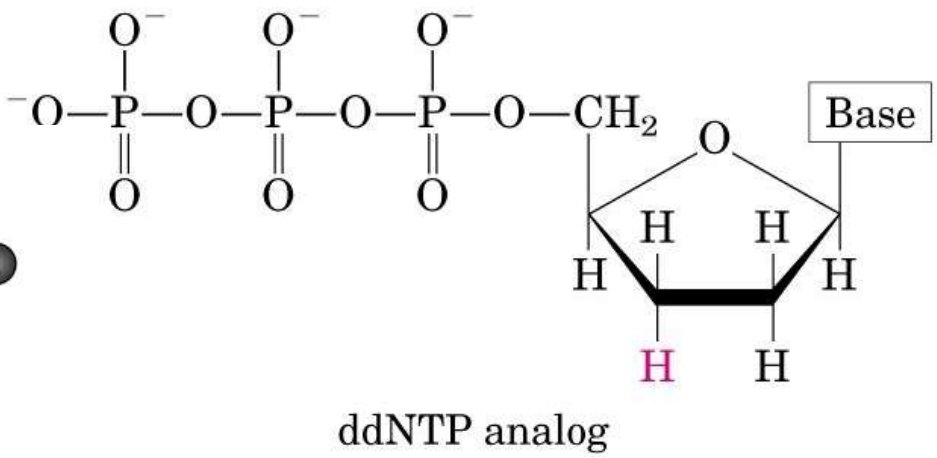
Early Benefits -

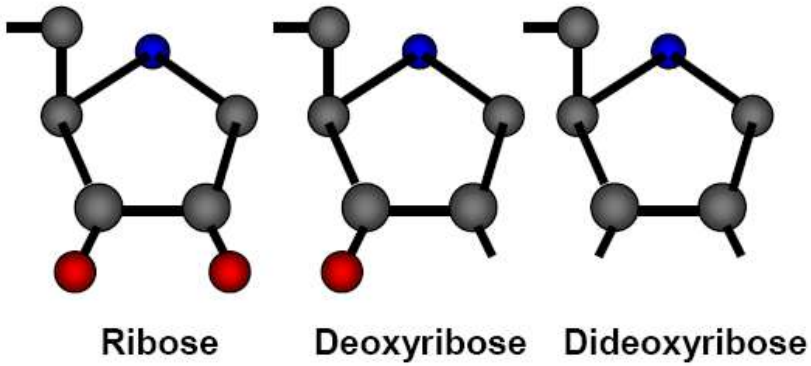
Discovery that the gene for ovalbumin in chicken and the gene encoding β-globin in rabbit contained non-coding gaps in the coding regions. These gaps were flanked by the same dinucleotides in the two genes; GT on the 5' end of the gaps and AG on the 3' end of the gaps. Soon, the terms intron and exon were added to the genetic lexicon to describe the coding and non-coding regions of eukaryotic genes (1977).

Fred Sanger (dideoxy) DNA Sequencing

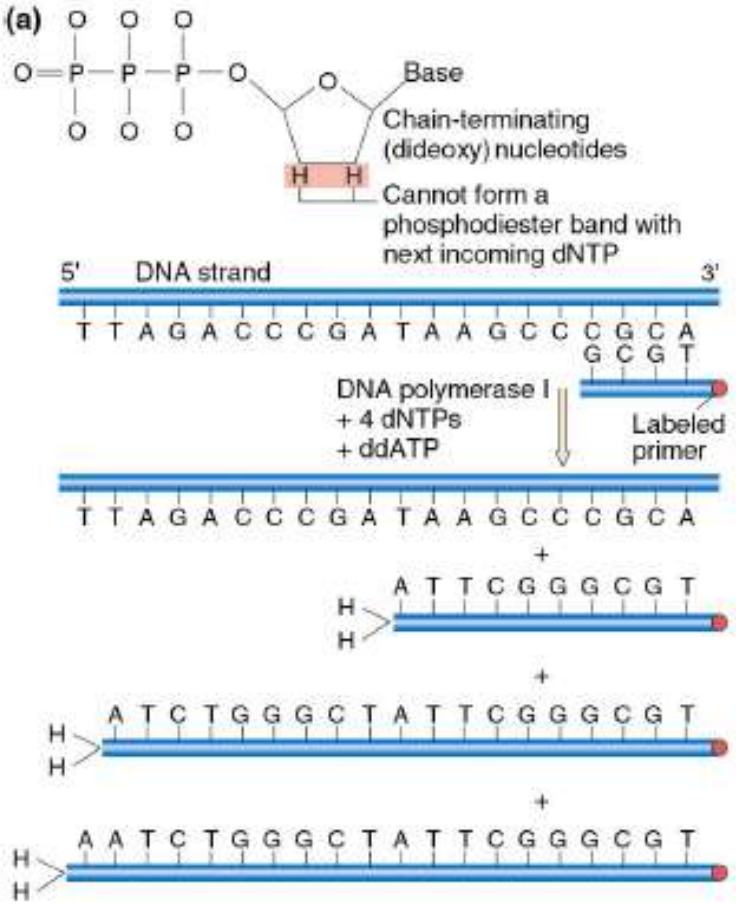
Sanger knew that, whenever a dideoxynucleotide was incorporated into a polynucleotide, the chain would irreversibly stop, or terminate. Thus, the incorporation of specific dideoxynucleotides in vitro would result in selective chain termination.

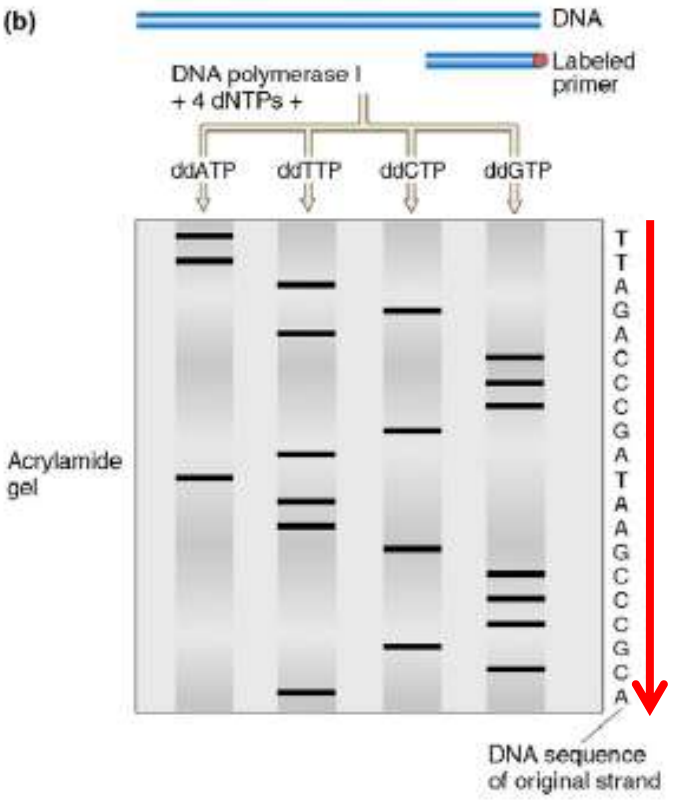


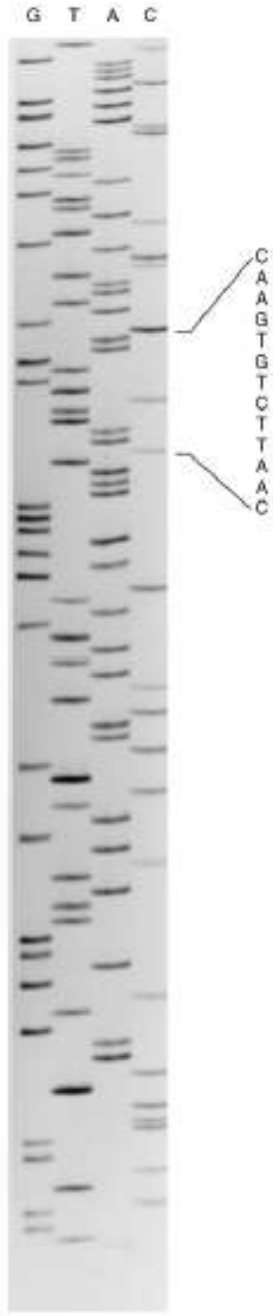




Sanger (dideoxy) DNA Sequencing



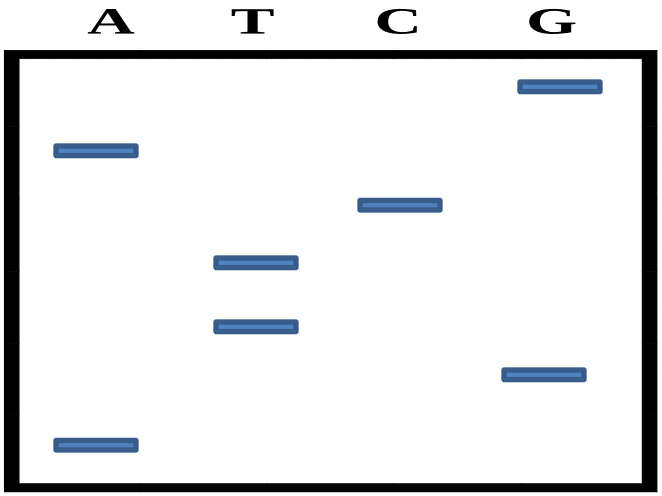




Consider the following nucleic acid sequencing gel experiment using the Sanger dideoxy sequencing method:

What is the expected sequence (5’ -> 3’) of the original DNA sample assuming the primer was labeled with a 5’-prime fluorescent label?

[DNA polymerase I + 4 dNTPs + ddATP ddTTP ddCTP ddGTP]



iClicker Question:

A) GACTTGA
B) CTGAACT
C) AGTTCGA
D) TCAAGTC
E) None of the above

Advantages of dideoxy DNA Sequencing

- **Elimination of dangerous chemicals (hydrazine)**
- **Greater efficiency**

Taq polymerase makes DNA strands off of a template at rate of about 500 bases per minute

Chemical synthesis of a 25-mer oligonucleotide takes more than two hours.

→ **High Throughput Methods (Human Genome Project)**

Automated Fluorescence Sequencing

In 1986, Leroy Hood and colleagues reported on a DNA sequencing method in which the radioactive labels, autoradiography, and manual base calling were all replaced by fluorescent labels, laser induced fluorescence detection, and computerized base calling.

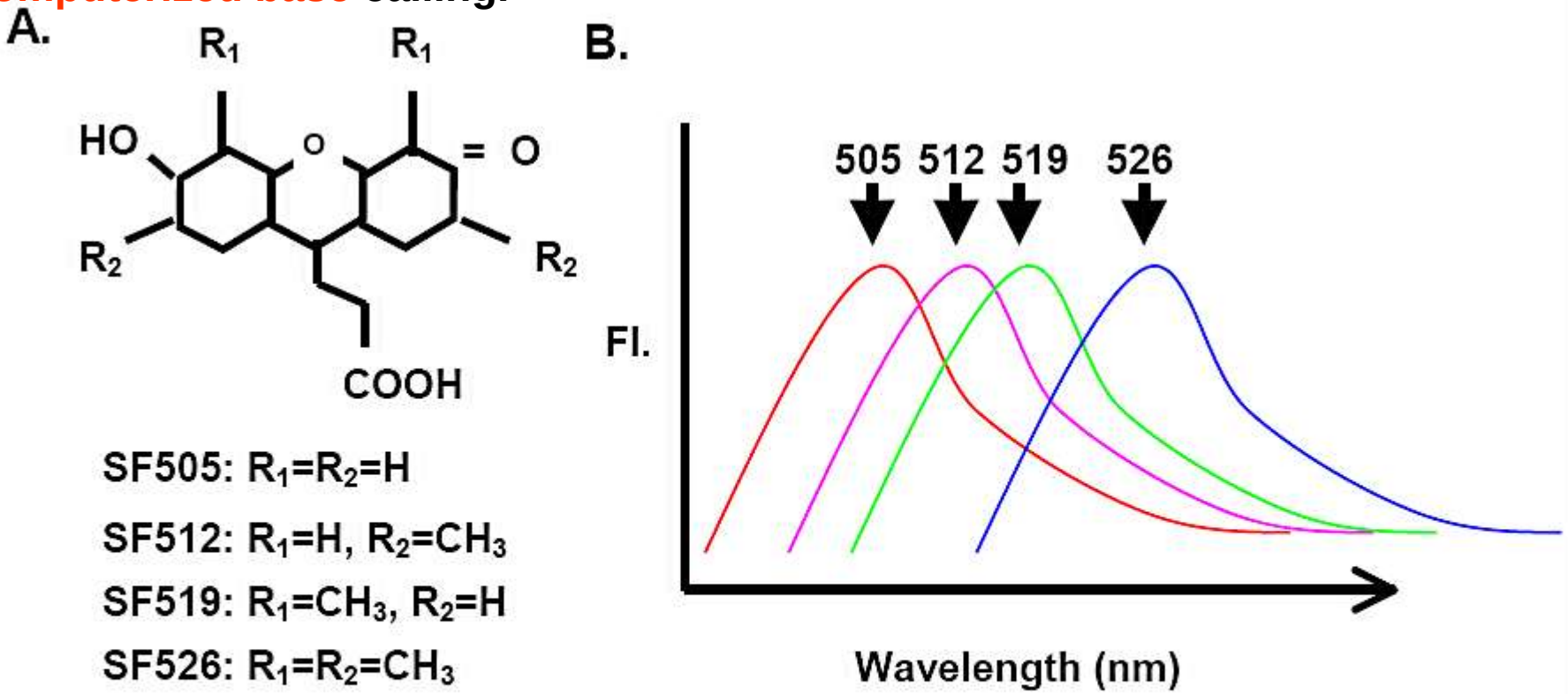
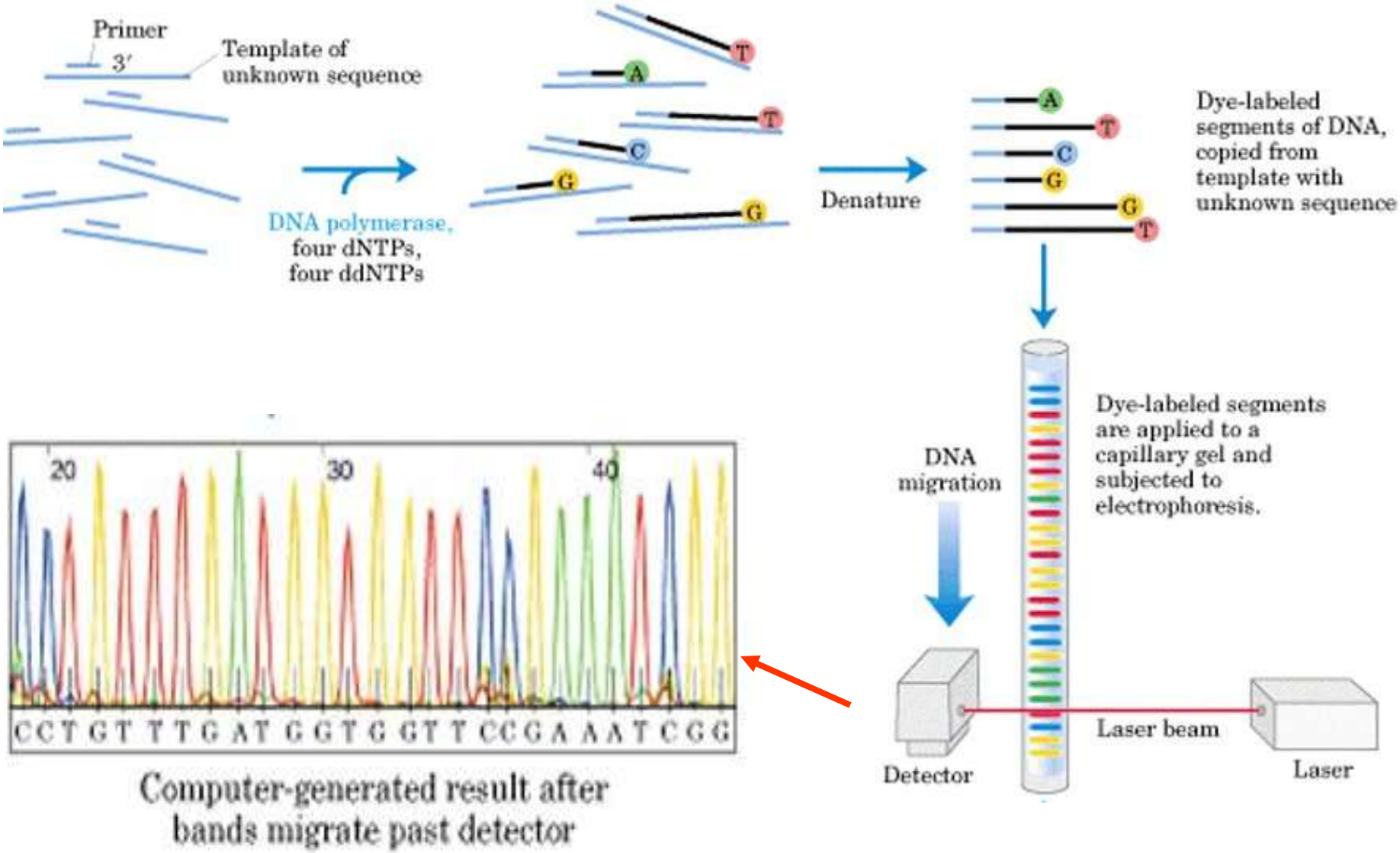


Figure 5. A. Chemical structure of the four succinylfluorescein dyes developed at DuPont. B. Normalized fluorescence emission spectra for each of the four dyes following excitation at 488nm. Shifts in the spectra were achieved by changing the side groups R_1 and R_2.

Automated DNA sequencing



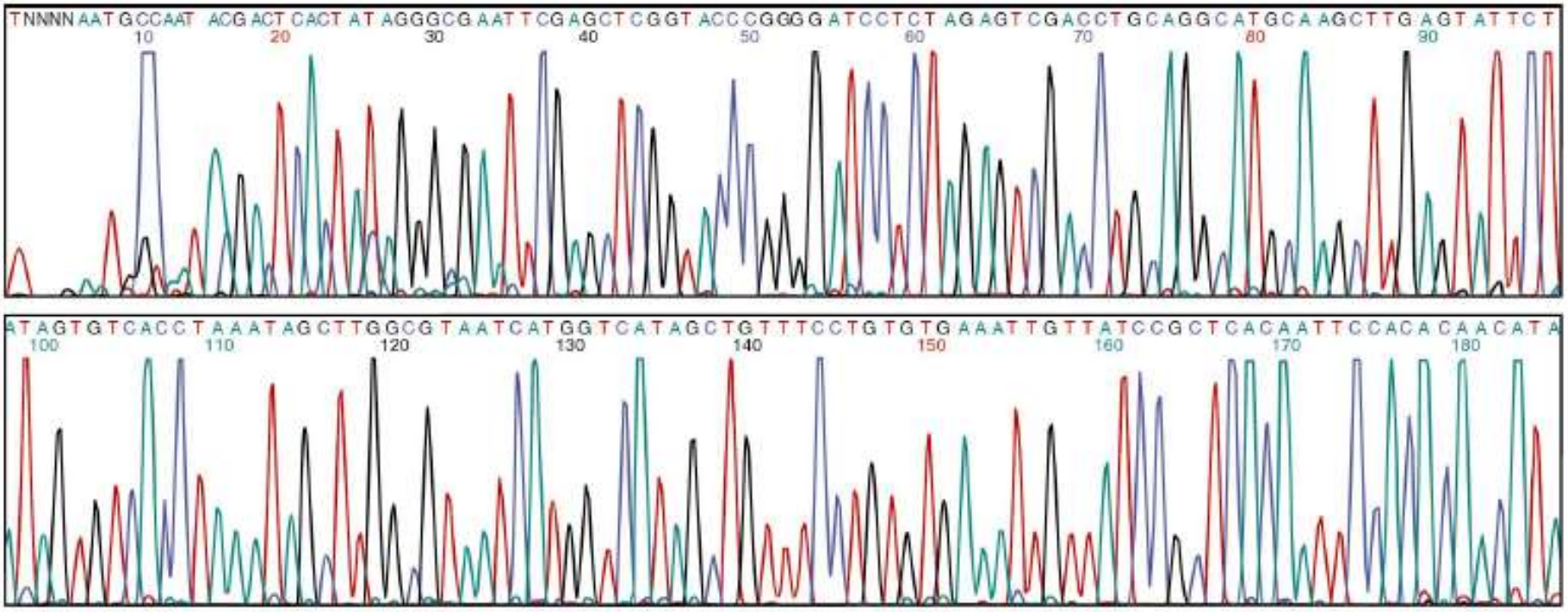
Automated dye-terminator sequencing

4-fluorescently labelled dideoxy dye terminators

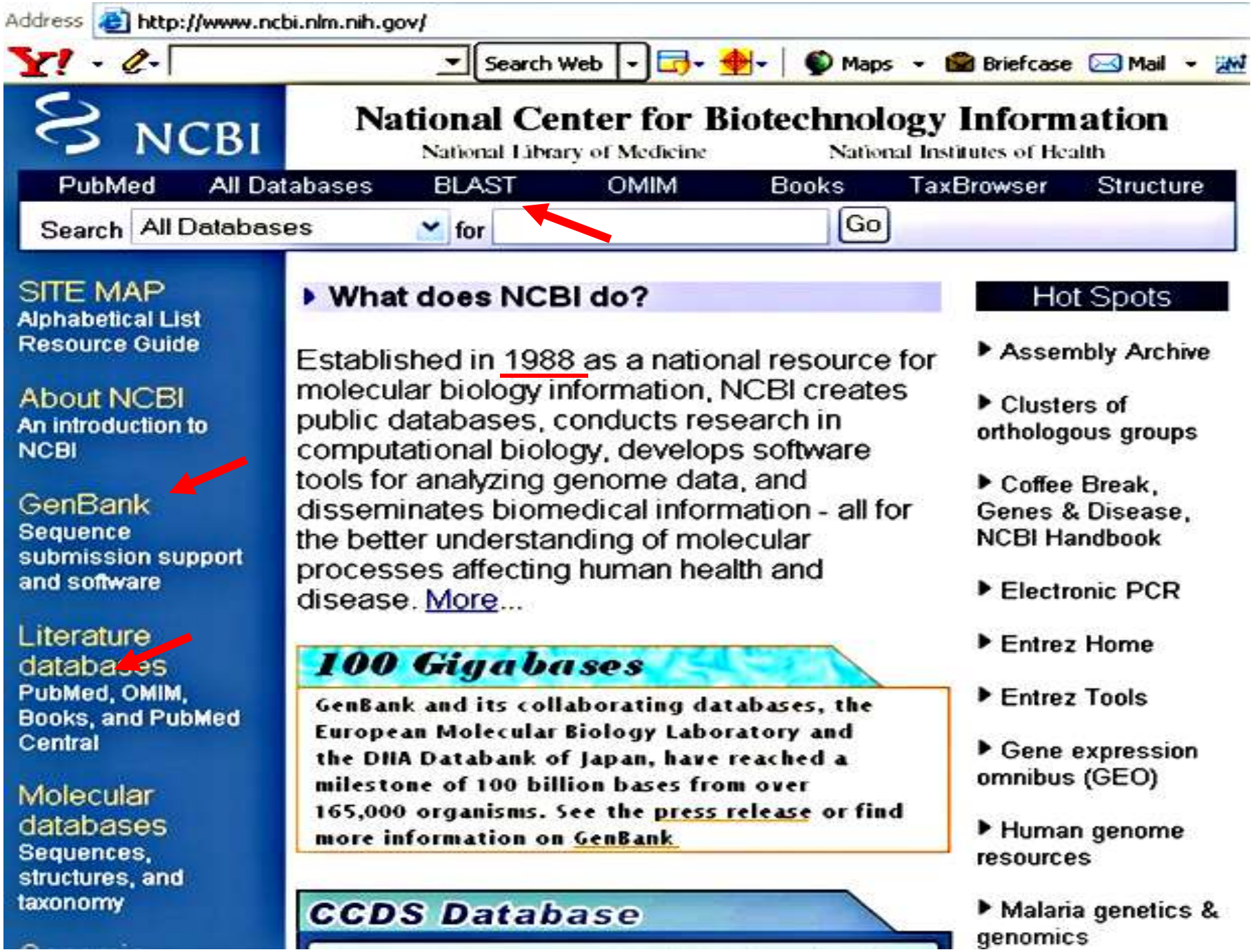
ddATP
ddGTP
ddCTP
ddTTP

pool and load in a single well or capillary

- scan with laser + detector specific for each dye
- automated base calling
- very long reads (~ 1000 bases)/run



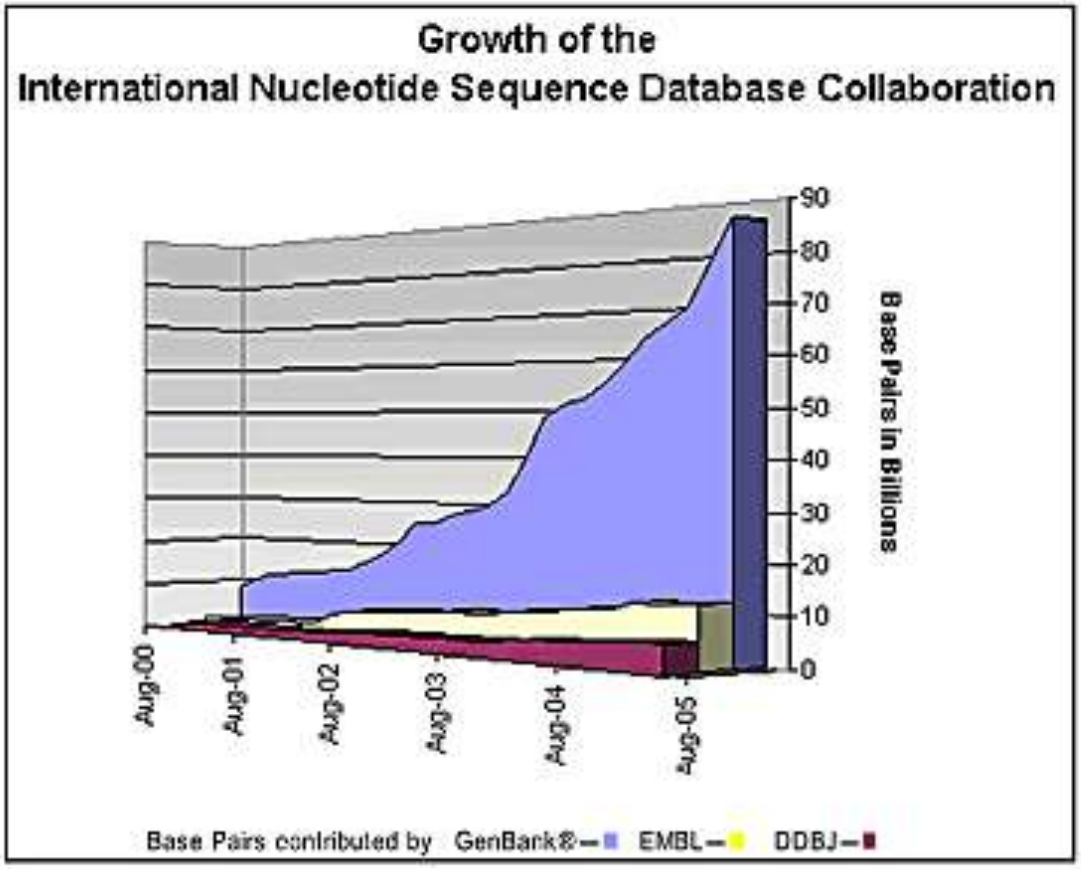
[DNA dideoxysequencing animation]


Address http://www.ncbi.nlm.nih.gov/
Search Web
Maps
Briefcase
Mail
NCBI
National Center for Biotechnology Information
National Library of Medicine
National Institutes of Health
PubMed
All Databases
BLAST
OMIM
Books
TaxBrowser
Structure
Search All Databases for Go
SITE MAP
Alphabetical List
Resource Guide
About NCBI
An introduction to NCBI
GenBank
Sequence submission support and software
Literature databases
PubMed, OMIM, Books, and PubMed Central
Molecular databases
Sequences, structures, and taxonomy
What does NCBI do?
Established in 1988 as a national resource for molecular biology information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. More...
100 Gigabases
GenBank and its collaborating databases, the European Molecular Biology Laboratory and the DNA Databank of Japan, have reached a milestone of 100 billion bases from over 165,000 organisms. See the press release or find more information on GenBank
CCDS Database
Hot Spots
Assembly Archive
Clusters of orthologous groups
Coffee Break, Genes & Disease, NCBI Handbook
Electronic PCR
Entrez Home
Entrez Tools
Gene expression omnibus (GEO)
Human genome resources
Malaria genetics & genomics

International sequence databases exceed 100 gigabases

In August 2005, the INSDC announced the DNA sequence database exceeded 100 gigabases. GenBank is proud of its contributions toward this milestone. We thank all the scientists who have worked through the submission process at GenBank and made their sequence data available to the world. See the related press release.

>100,000,000,000 bases



> 200,000 organisms!!

Amino Acids and Proteins

The one letter code for lysine is _____.

Answers: A) A B) L C) S D) K E) H

The approximate charge at pH 1 of the oligopeptide "C - H – A – N – D – R – A " is ___.

Answers: A) +3 B) +2 C) +1 D) -1 E) -2

The estimated isoelectric point of the oligopeptide "C - H – A – N – D – R – A " is ____.

Answers: A) 4 B) 5 C) 6 D) 7 E) 9

pI estimator on-line

The world at 1,000,000 X (1 nm → 1 mm)

At this magnification, double stranded DNA would be approximately _____ in width.

Answers: A) 1 mm B) 1 m C) 2 mm D) 20 m E) 5 mm