

Introduction to Microbial Genomics

Sequences as information

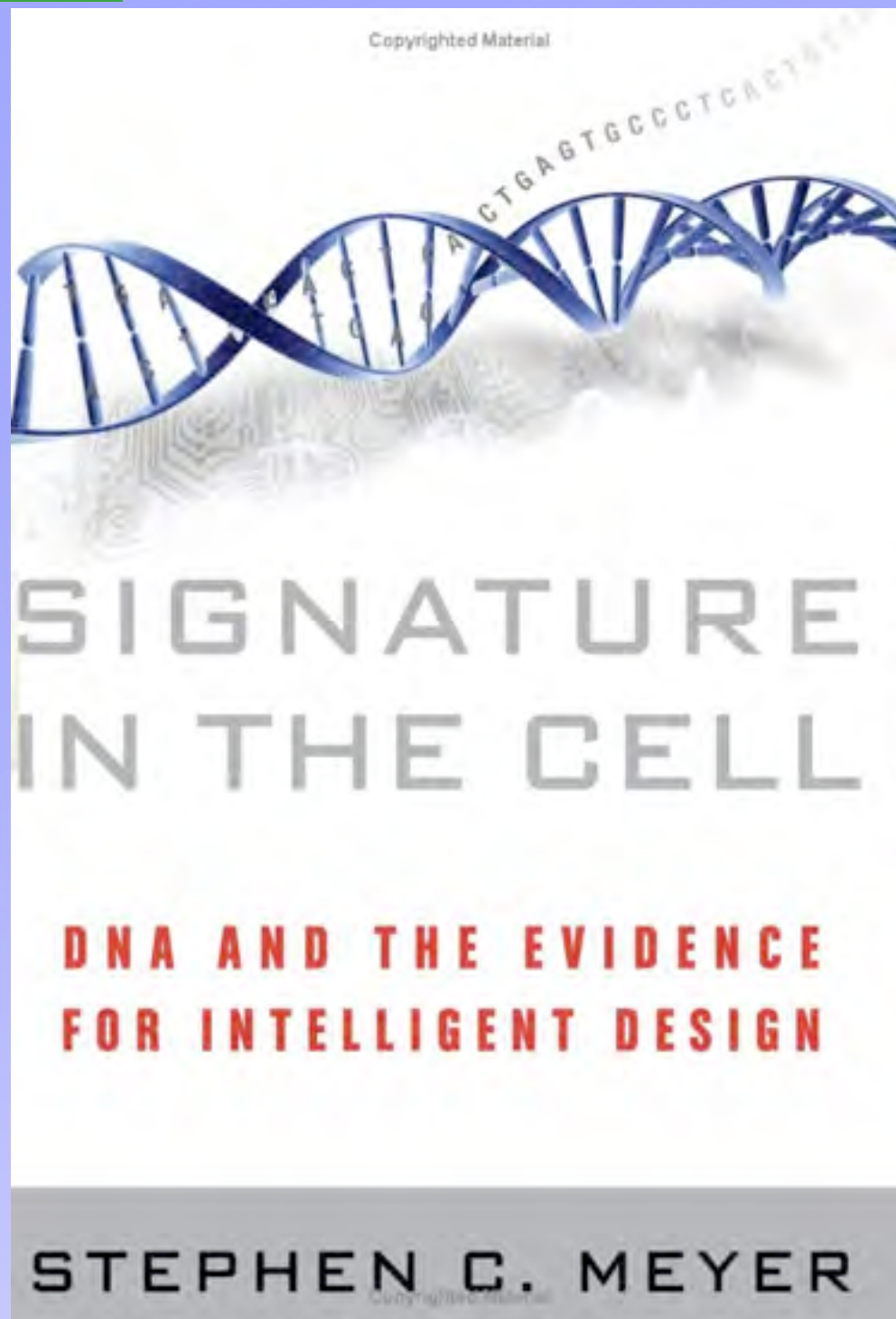
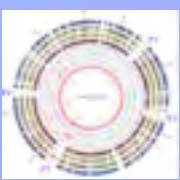


Dave Ussery

Workshop on Comparative Genomics
King Mongkut's University of Technology Thonburi
Bangkok, Thailand

Second Talk for Monday
8 March, 2010





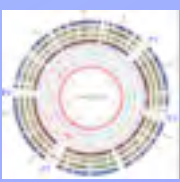
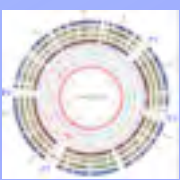


Table 1 Estimating the magnitude of microbial diversity

Number of bacteriophages on Earth	10^{31}
Number of microbes on Earth	5×10^{30}
Number of stars in the universe	7×10^{21}
Number of microbes in all humans	6×10^{23}
Number of humans	6×10^9
Number of microbial cells in one human gut	10^{14}
Number of human cells in one human	10^{13}
Number of microbial genes in one human gut	3×10^6
Number of genes in the human genome	2.5×10^4
Combined length of all bacteriophages on Earth	10^8 Ly
Diameter of the Milky Way	10^5 Ly

"An inordinate fondness of bacteriophages..."



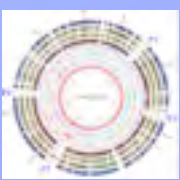
A Brief History of Biological Information

Based on three excellent books:

“THE LOGIC OF LIFE - A History of Heredity”, by Francois Jacob (Vintage Books, A Division of Random House, New York, 1973, translated by Betty E. Spillman).

“WHO WROTE THE BOOK OF LIFE? - A History of The Genetic Code”, by Lily E. Kay (Stanford University Press, Stanford, California, 2000).

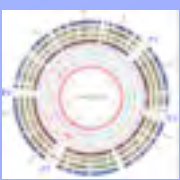
“THE INSIDE STORY - DNA to RNA to Protein”, edited by Jan Witkowski (Cold Spring Harbor Press, New York, 2005).



Aristotle ~350 B.C.



plants
animals
minerals

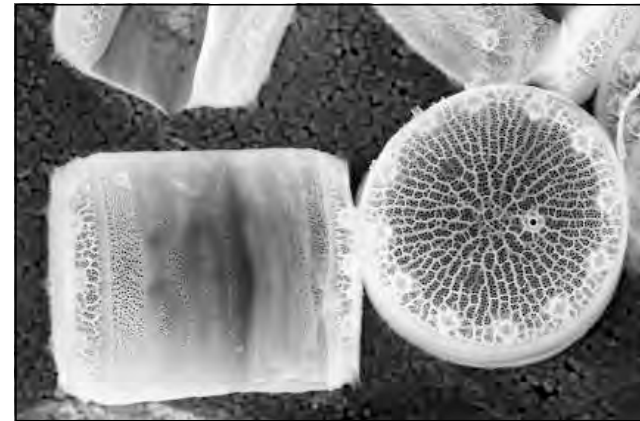


GENETICS

DNA Reveals Diatom's Complexity

Diatoms are an enigma. Neither plant nor animal, they share biochemical features of both. Though simple single-celled algae, they are covered with elegant casings sculpted from silica.

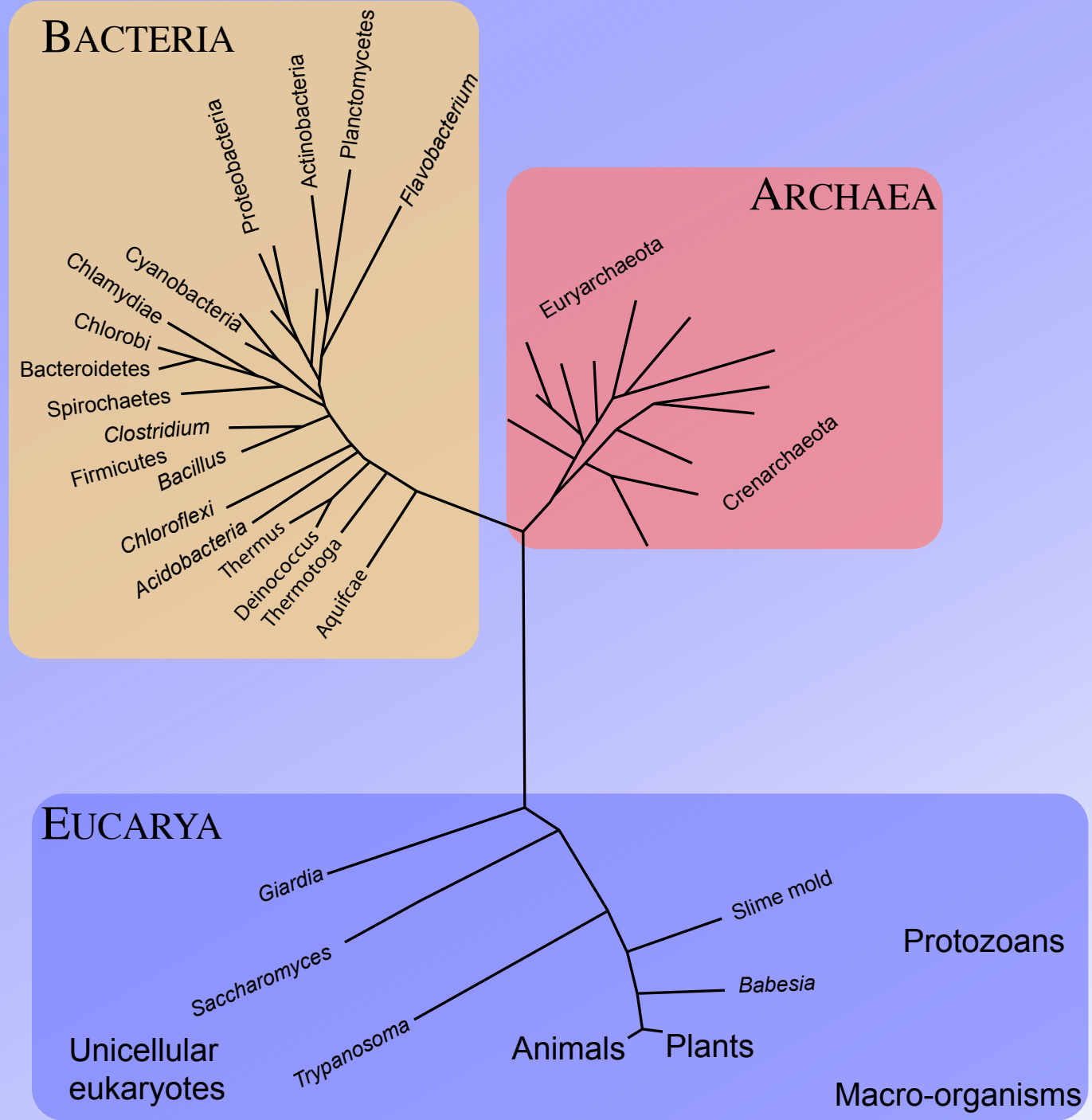
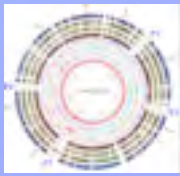
Now a team of 45 biologists has taken a big step toward resolving the paradoxical nature of these odd microbes. They have sequenced the genome of *Thalassiosira pseudonana*, which lives in salt water and is a lab favorite among diatom experts. The

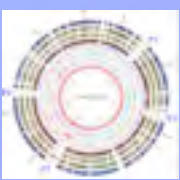


Aqueous snowflake. The sequence of a diatom should reveal the secrets of its decorative shell.

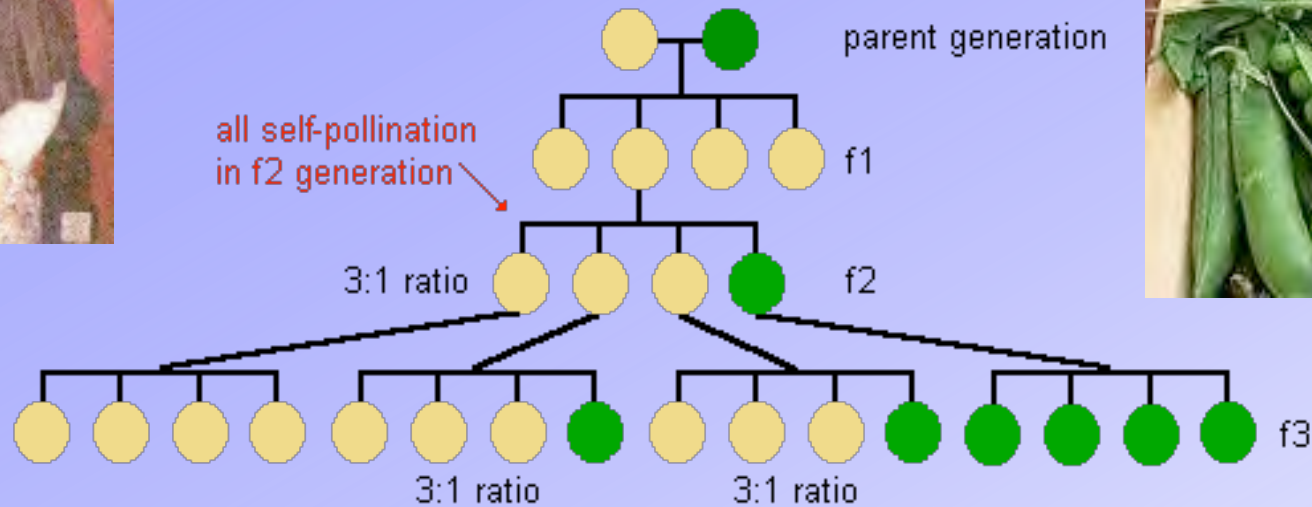
www.sciencemag.org SCIENCE VOL 306 1 OCTOBER 2004

Published by AAAS

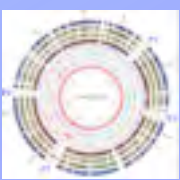




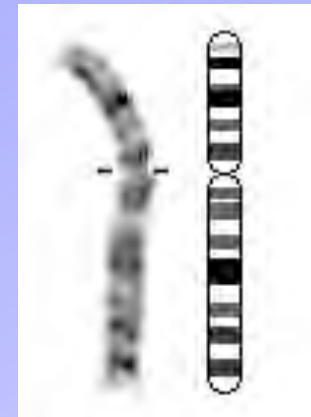
Gregor Mendel 1866



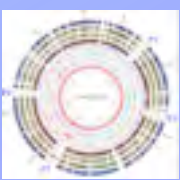
genes



T.H. Morgan 1919



Chromosomes



G. Beadle & E. Tatum



1930s



one gene



one enzyme

STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE
INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES

INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION
ISOLATED FROM PNEUMOCOCCUS TYPE III

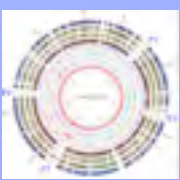
BY OSWALD T. AVERY, M.D., COLIN M. MACLEOD, M.D., AND
MACLYN McCARTY,* M.D.

(From the Hospital of The Rockefeller Institute for Medical Research)

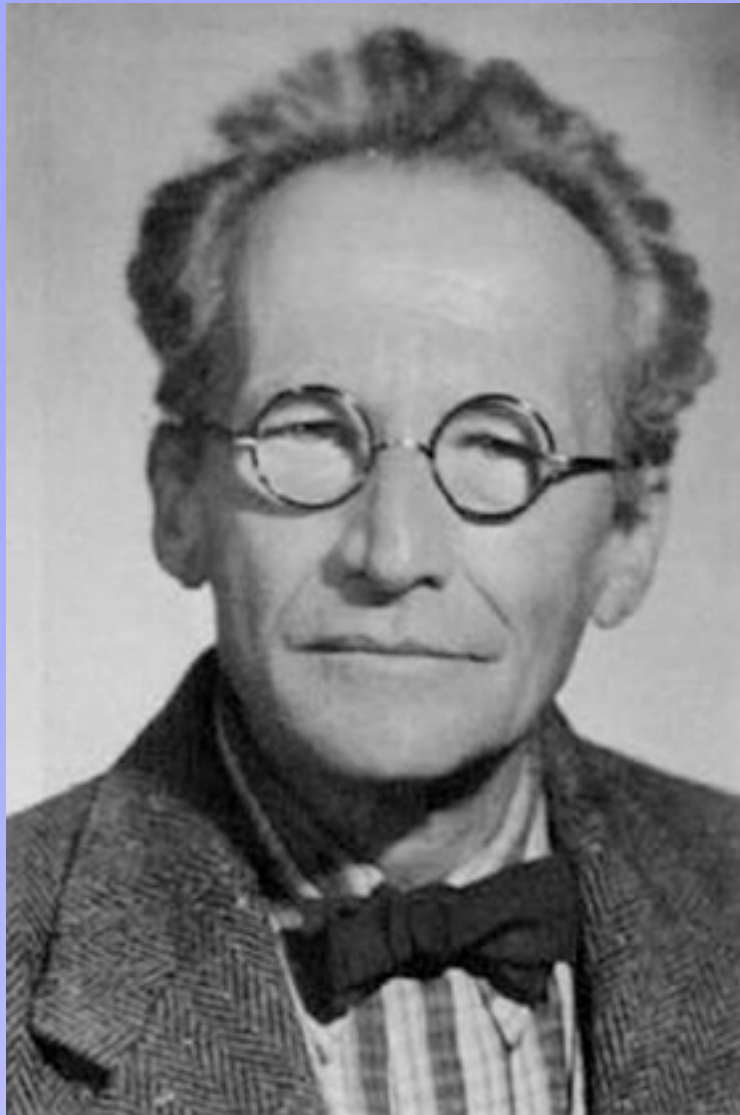
PLATE 1

(Received for publication, November 1, 1943)

Biologists have long attempted by chemical means to induce in higher organisms predictable and specific changes which thereafter could be transmitted in series as hereditary characters. Among microorganisms the most striking example of inheritable and specific alterations in cell structure and function that can be experimentally induced and are reproducible under well defined and adequately controlled conditions is the transformation of specific types of Pneumococcus. This phenomenon was first described by Griffith (1) who succeeded in transforming an attenuated and non-encapsulated (R) variant derived from one specific type into fully encapsulated and virulent (S) cells of a heterologous specific type. A typical instance will suffice to illustrate the techniques originally used and serve to indicate the wide variety of trans-



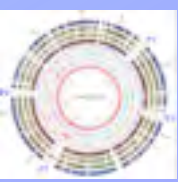
E. Schrödinger 1943



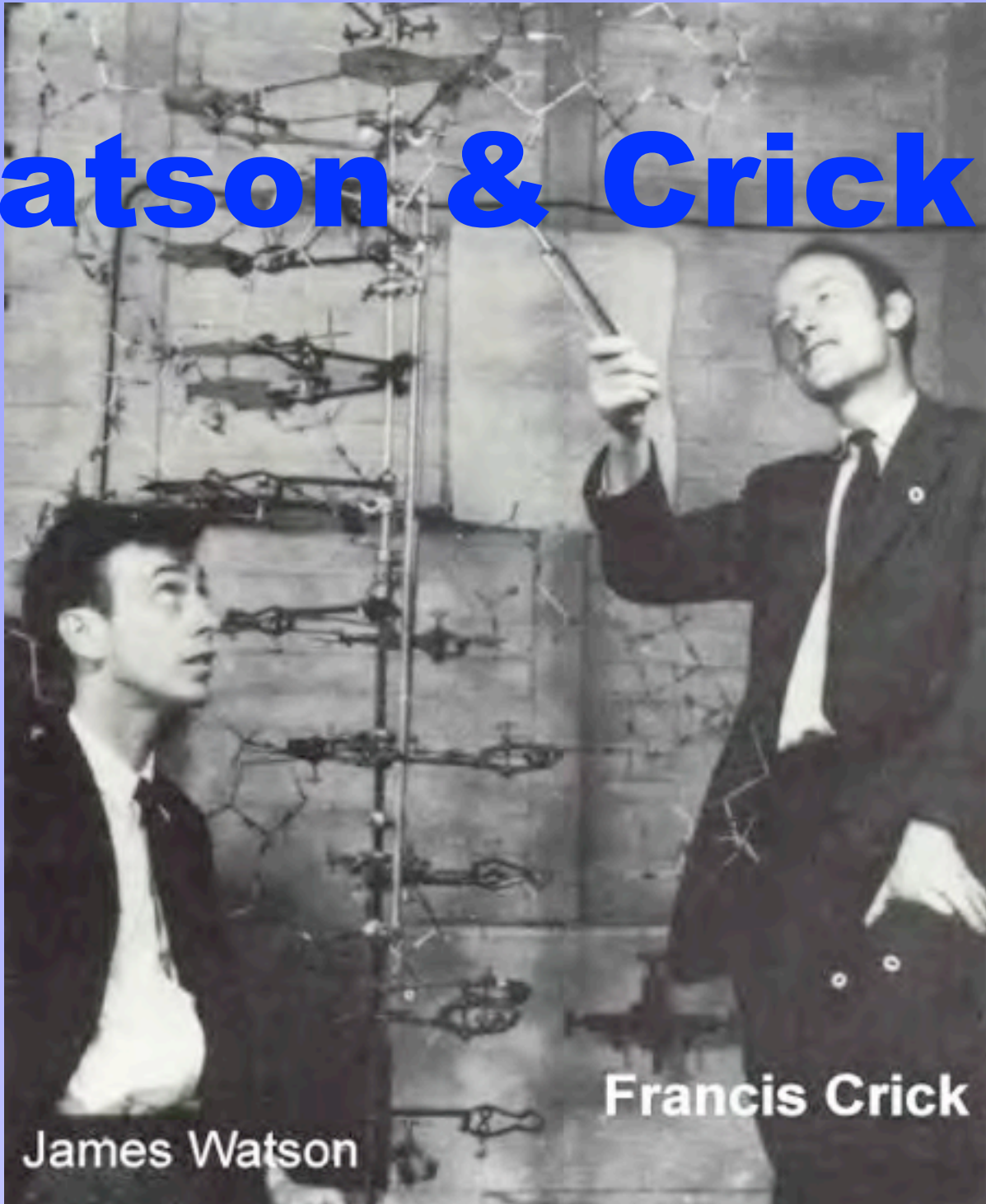
“We believe a gene - or perhaps the whole chromosome fibre - to be an periodic solid.”

“...For an illustration, think of Morse code...”

"What is Life?" by Erwin Schrödinger
(Cambridge University Press, 1944)

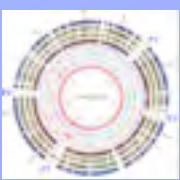


Watson & Crick 1953

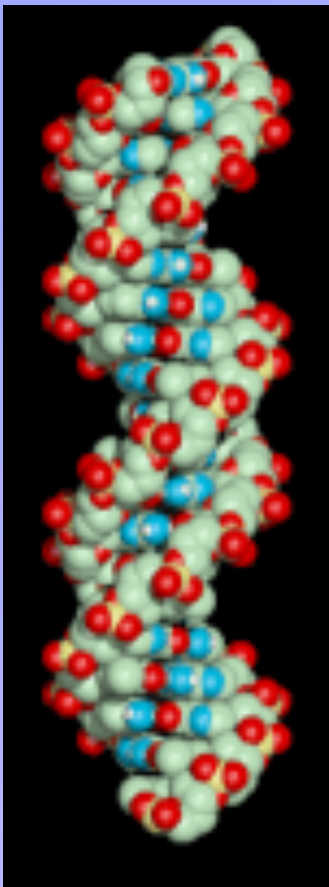


James Watson

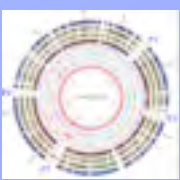
Francis Crick



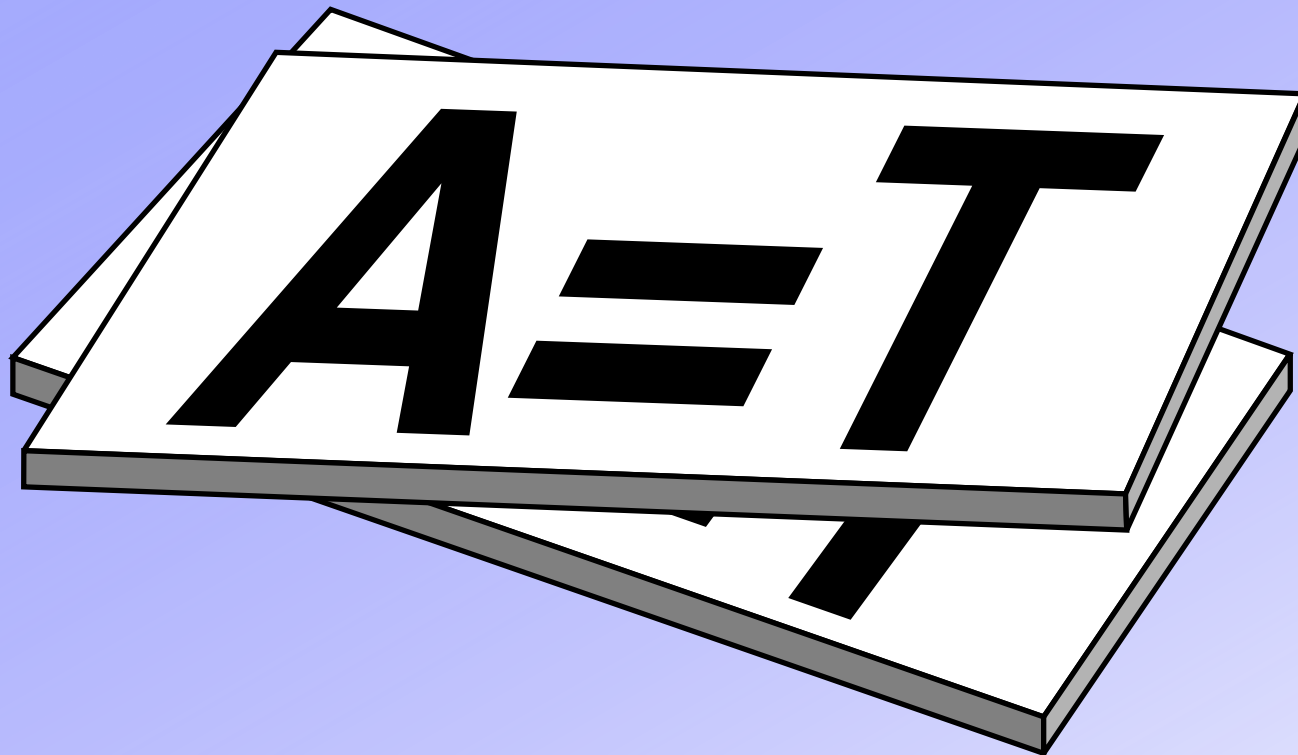
DNA is like Coca-cola!

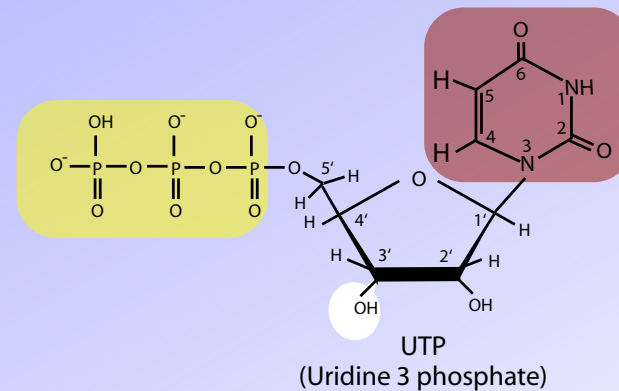
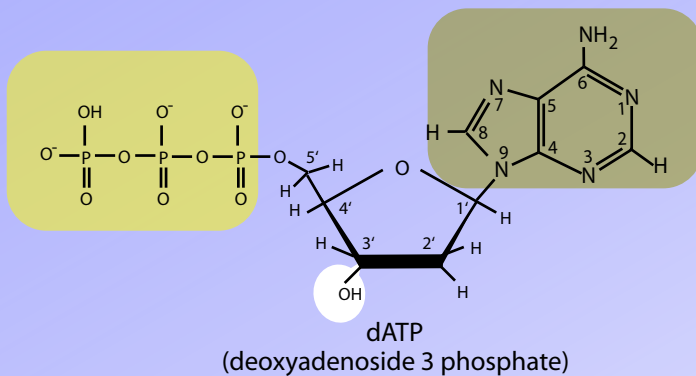
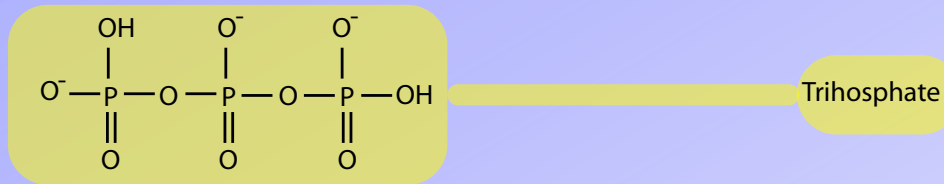
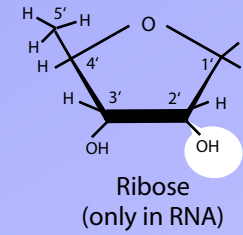
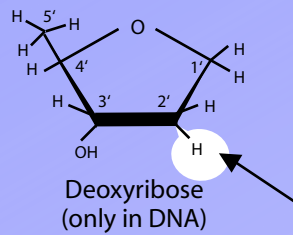
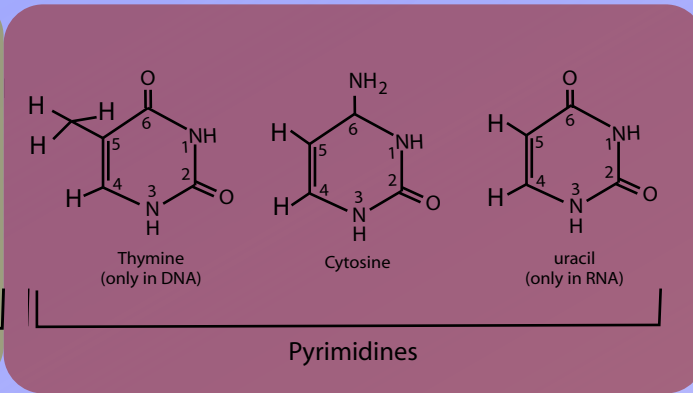
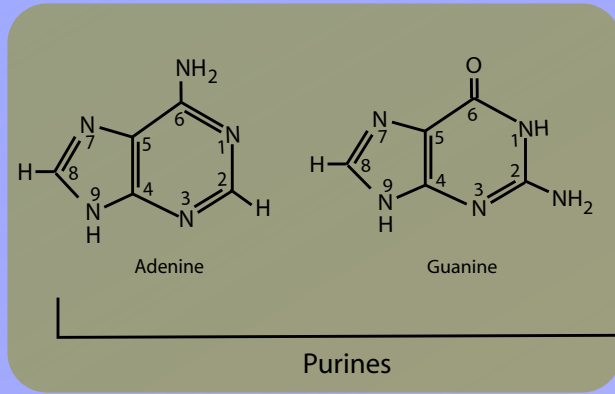
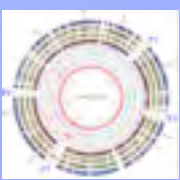


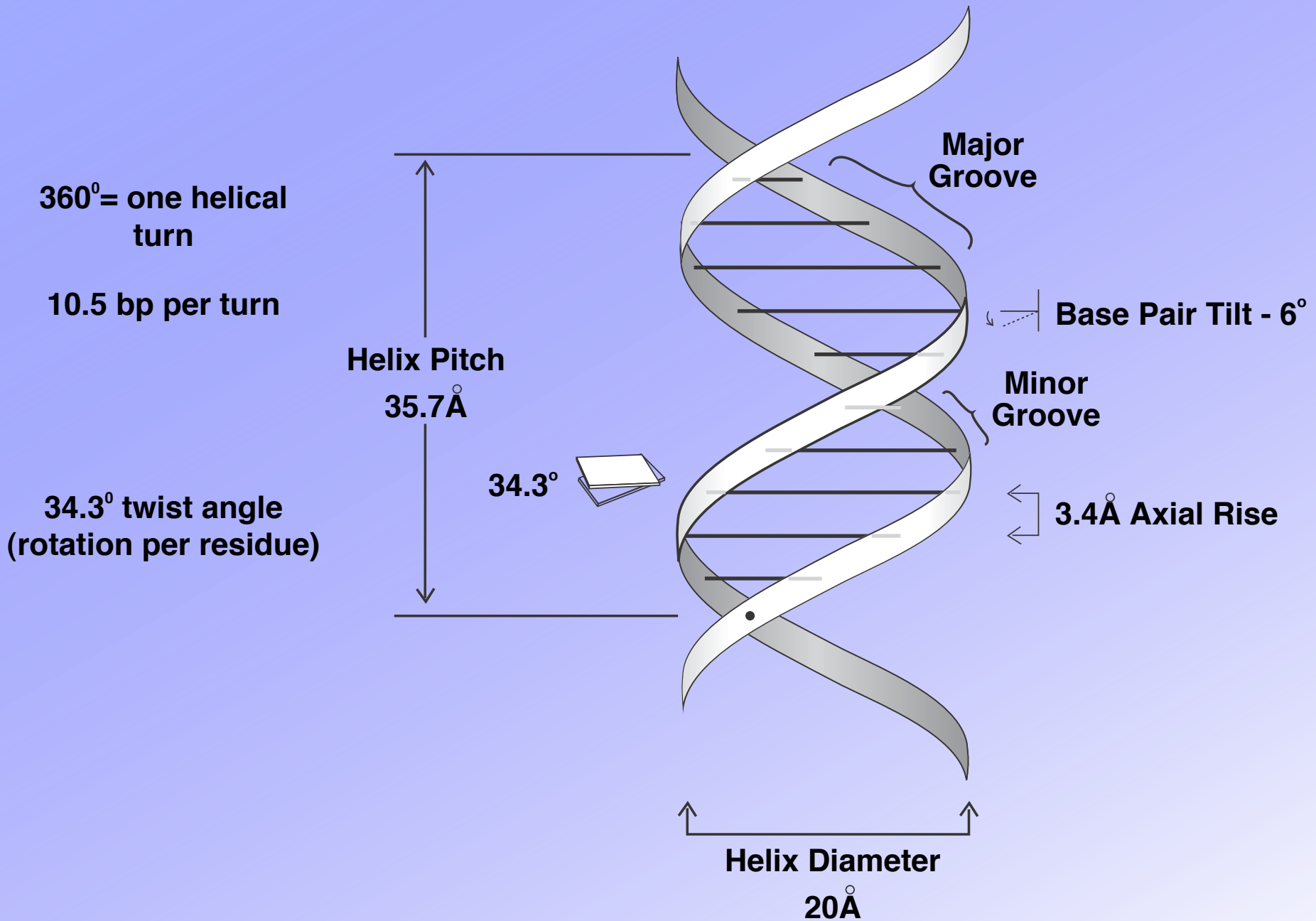
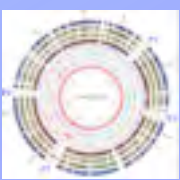
Coke	DNA
Water	Water
Sugar (sucrose)	(deoxyribose)
Phosphate acid	(PO ₄) backbone
Caffeine	Bases



DNA bases will spontaneously stack on top of each other and form a helix!







โปรดกักกลิ่นไม่สะอาด

ศูนย์อนุรักษ์แพะแดงเชียงใหม่ จังหวัดเชียงใหม่

วัตถุประสงค์ เพราะแมวกินเนื้อสัตว์เป็นอาหาร

๑. เพื่อเจ็ดพระภิกษุพระนิกายเถรวาทได้บริโภคอาหารตามหลักศาสนา

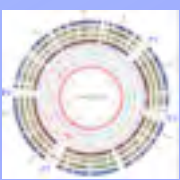
Bad smell area

๒. เพื่อ

๓. เพื่ออนุรักษ์มรดกของชาติ และ มรดกโลก

๔. เพื่อเผยแพร่ให้ชาวไทย และ ชาวต่างชาติได้รู้จักแพะไทยมากขึ้น

๕. เพื่อส่งเสริมการท่องเที่ยวเชิงวัฒนธรรม



DNA

5' ... GATCTAGCGATGCCGATGAAACATGATCATG ... 3'
3' ... CTAGATCGCTACGGCTACTTTGTACTAGTAC ... 5'

transcription

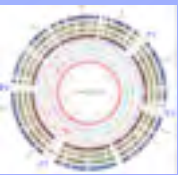
mRNA

5' ... GAUCUAGCGAUGCCGAUGAAACAUGAUGAUG ... 3'

translation

Protein

N met-pro-met-lys-his-his-his...C



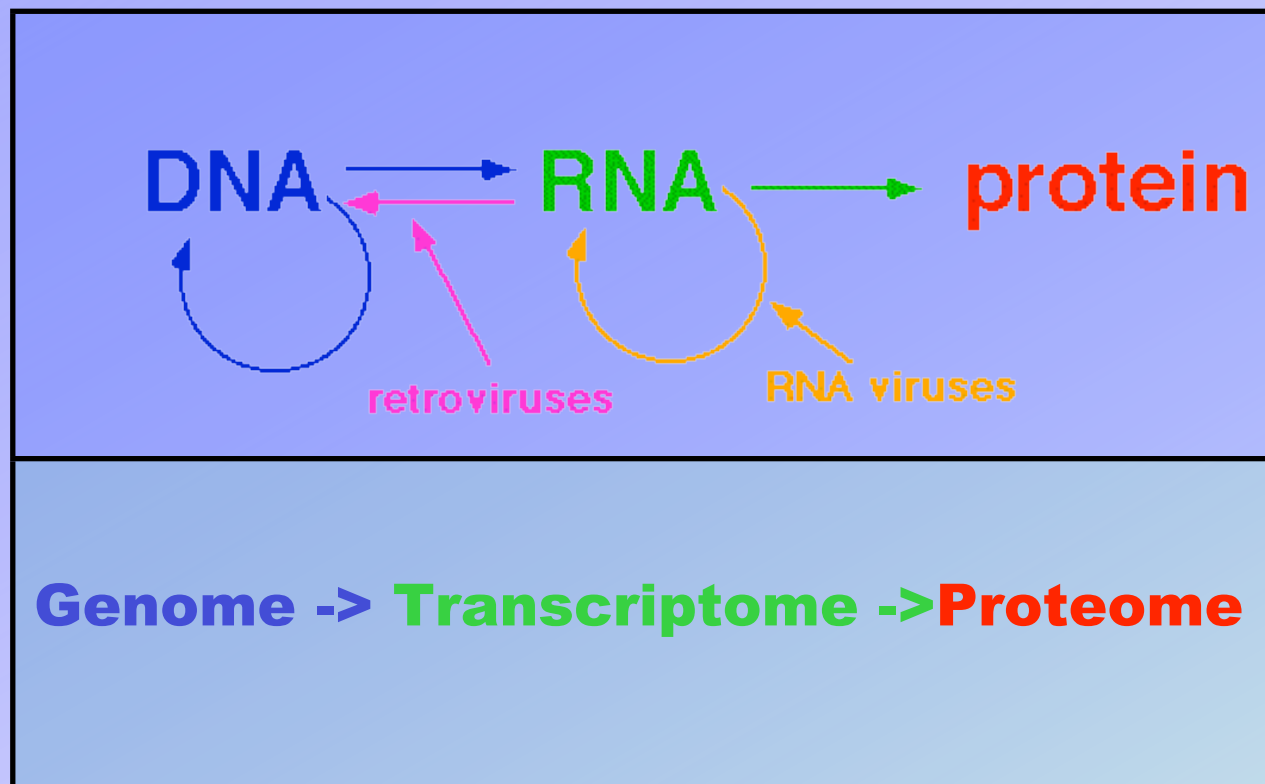
The General Idea

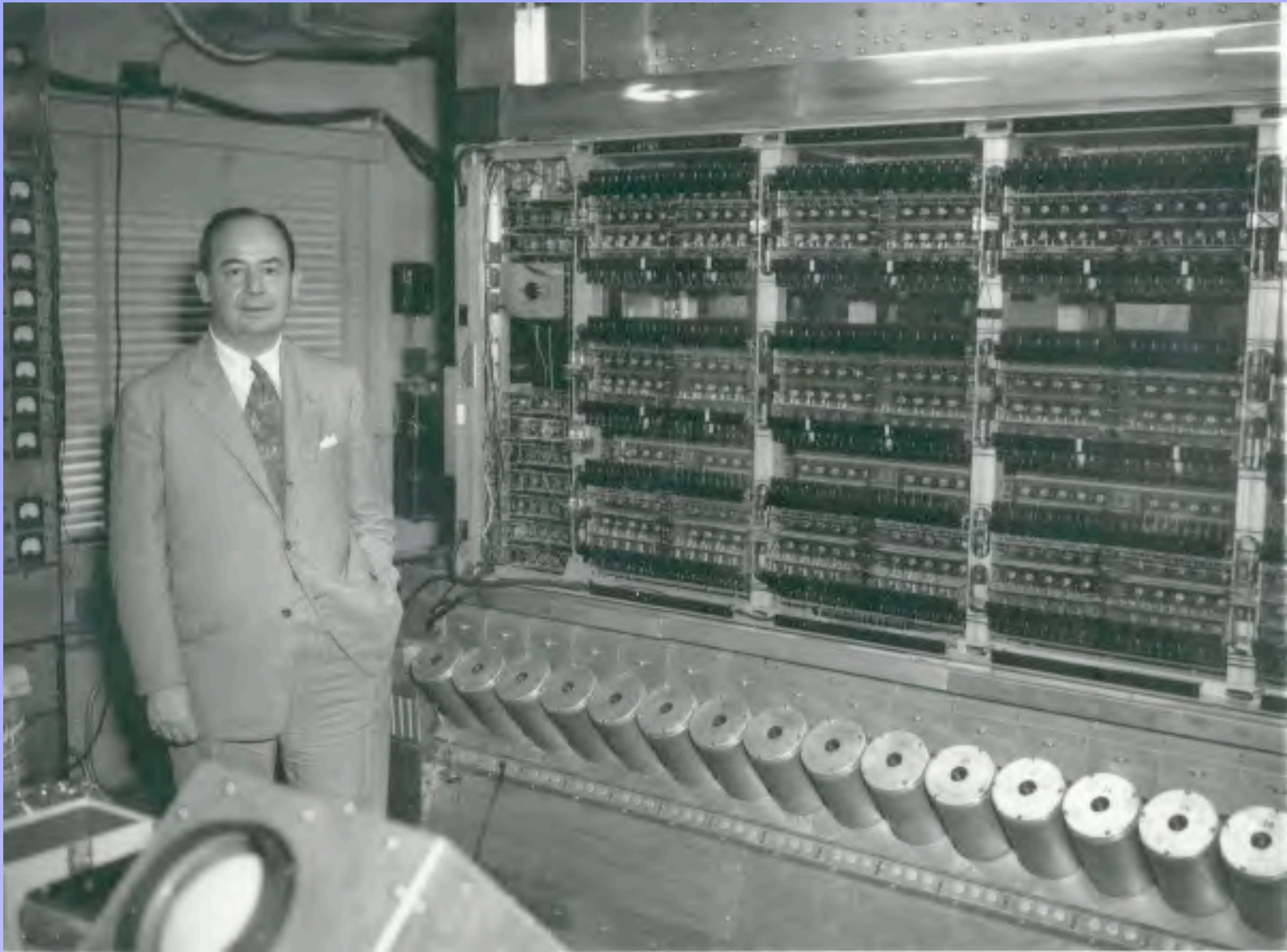
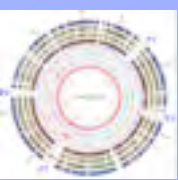
1. The Sequence Hypothesis

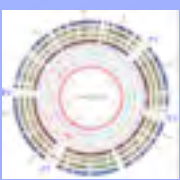
The amino acid sequence in proteins is specified from DNA and RNA.

2. The Central Dogma

Once information flows to protein, it cannot come back!





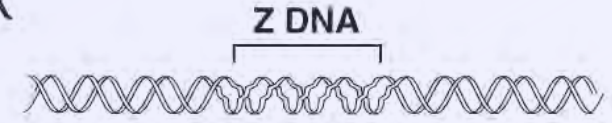
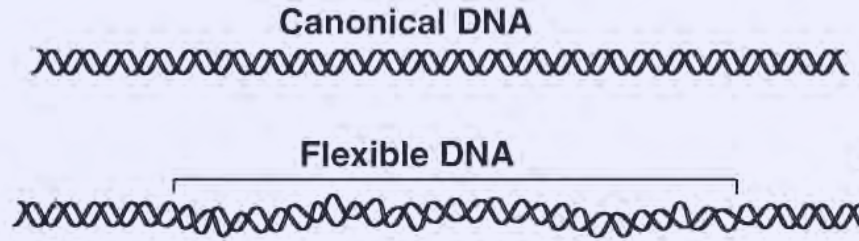
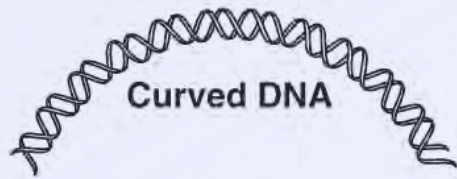
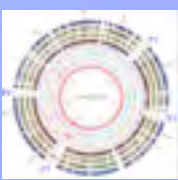


Biological Sequences as Information

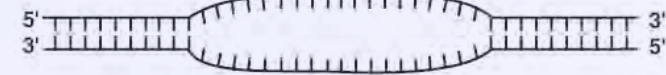
DNA sequences as information

1. DNA sequence can code for an amino acid sequences (mRNAs)
2. The DNA sequence can code for stable RNA sequences
 tRNA rRNA snRNA telomerase RNA
3. The DNA sequence can code for protein binding sites
4. The DNA can code for architectural information
 intrinsic DNA curvature
 nucleosome positioning
5. The DNA can code for structural / stability information
 transcription initiation
 origins of replication
 mutational "hot spots"

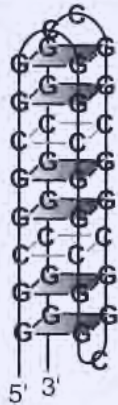




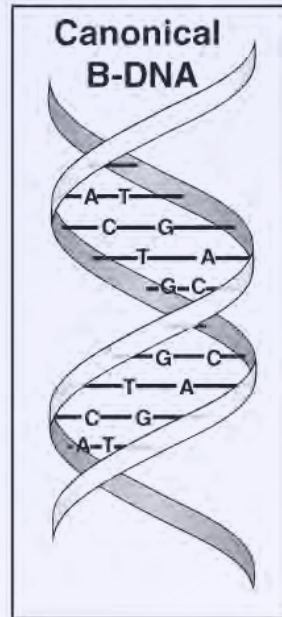
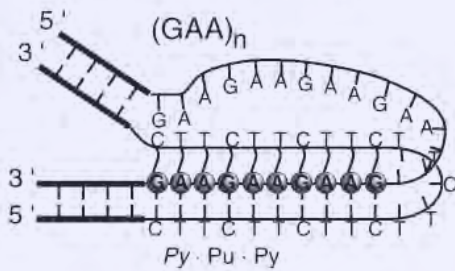
Unwound DNA



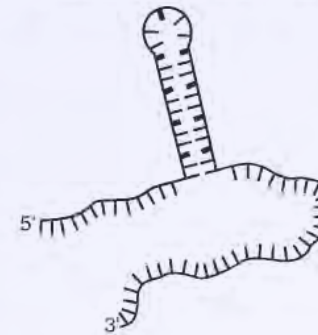
CGG quadruplex



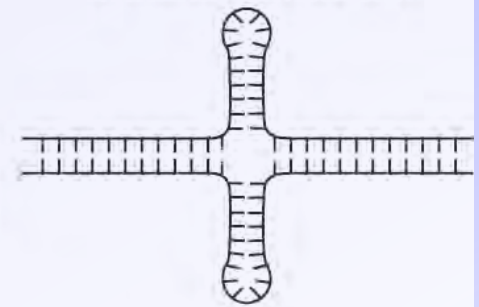
Hy3-type intramolecular triplex



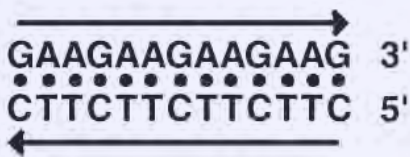
Hairpin structure



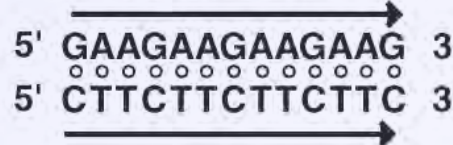
Cruciform structure



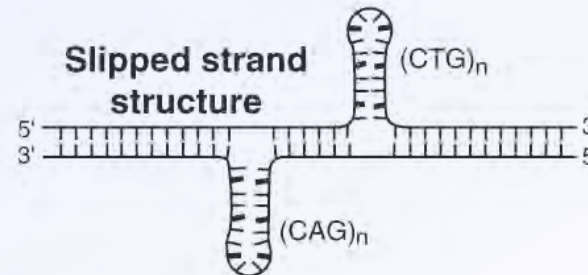
Anti Parallel DNA



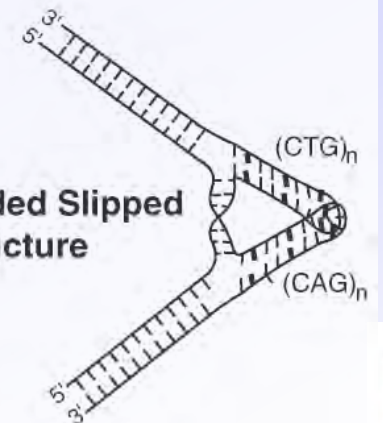
Parallel DNA

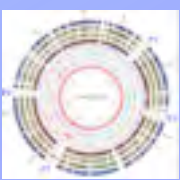


Slipped strand structure



Folded Slipped Structure

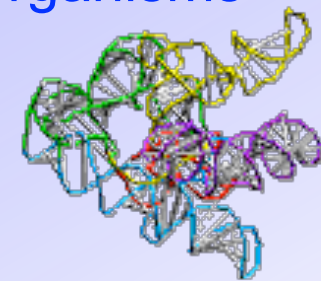


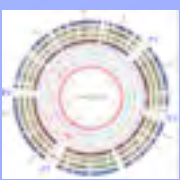


Biological Sequences as Information

RNA sequences as information

1. The mRNAs can contain several different levels of information:
 - specifies amino acid sequence for proteins
 - localisation signals for WHERE the protein will be made
 - stability signals to determine HOW MUCH protein is made
 - splice sites
 - editing sites
2. The tRNAs code for the genetic code - same in all living organisms (n.b. diff. in mitochondria)
3. The rRNAs code for the structures of ribosomes
4. Other RNA/protein complexes

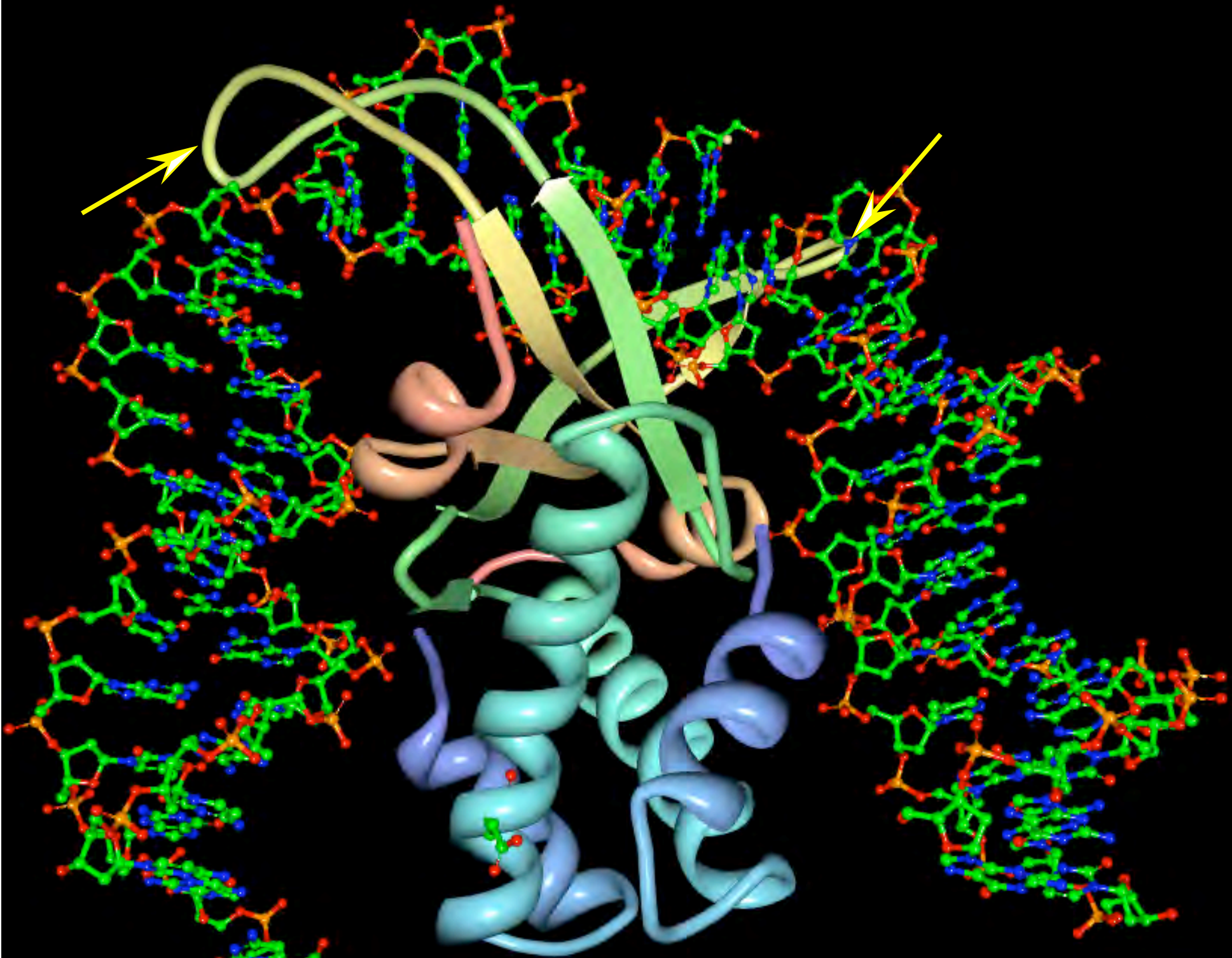


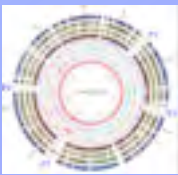


Biological Sequences as Information

Protein sequences as information

1. The protein sequence can code for an "active site" for enzymes
2. The protein sequence can code for structural roles: microtubules, myosin, collagen, *etc.*
3. The protein sequence can code for ion channels/pumps
4. The protein sequence can code for localisation information
5. The protein sequence can code for modification sites





Summary (so far!)

DNA
RNA
Protein



