



# **Lecture 3:**

## **History and Review (II)**

### **DNA the code**

**Course 485**

**Introduction to Genomics**

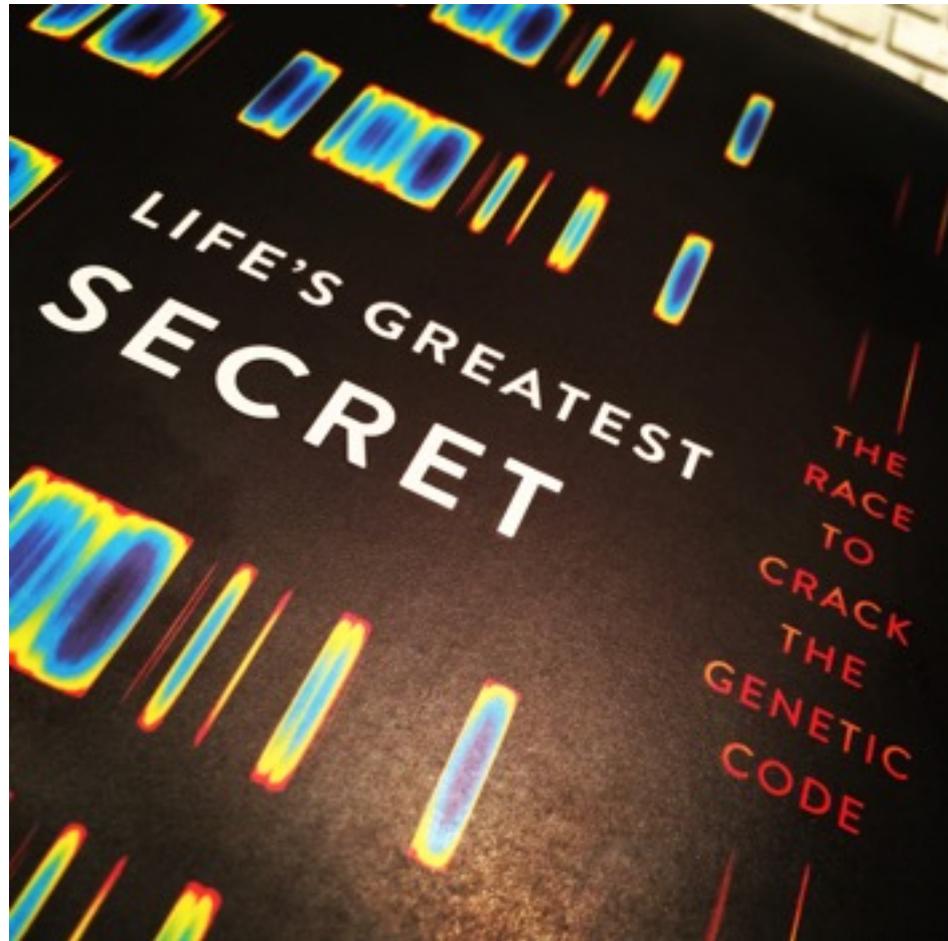


# AIMS

---

- Review concepts related to:
  - Chemical composition of DNA and RNA
  - Double helix DNA structure
  - Genome organizations
  - Histone modifications
  - Crick's Central Dogma
  - Decoding the code

# Good Reads



Life's Greatest Secret: The Story of the Race to Crack the Genetic Code by Matthew Cobb



**What is nucleic acid made of?**

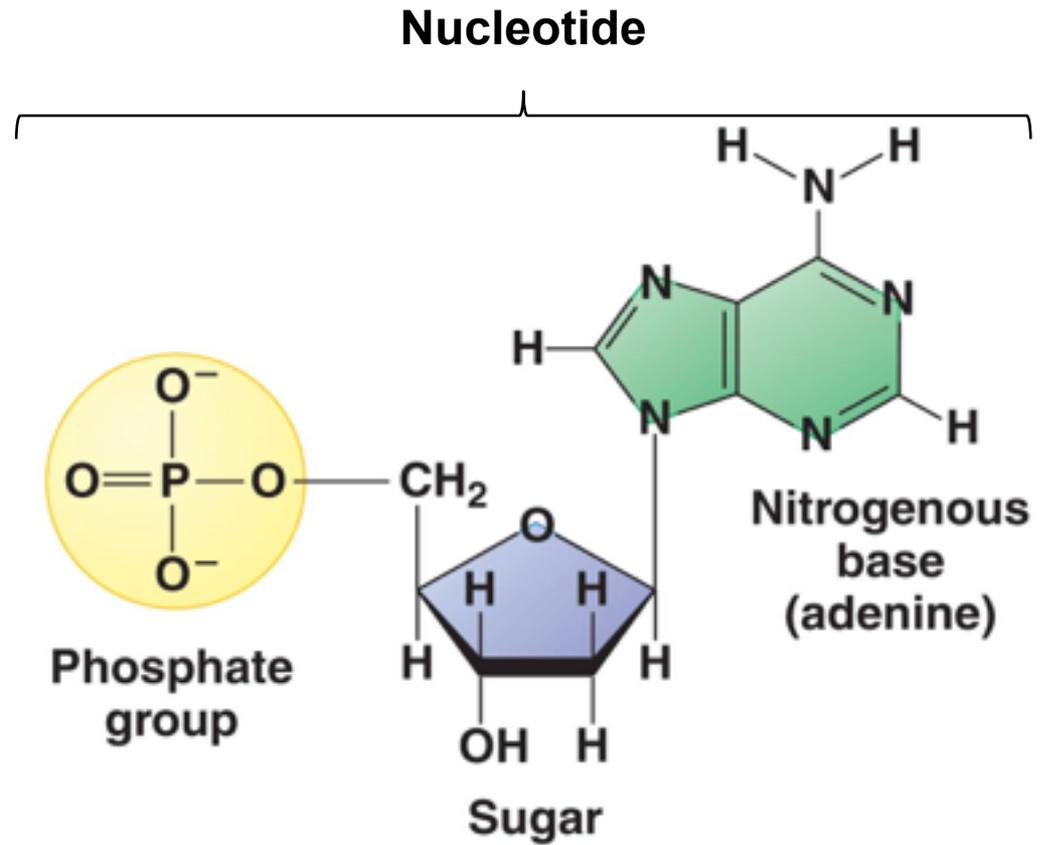
**What is the chemical composition of nucleic acids (DNA and RNA)?**

# DNA and RNA chemical unit

- The chemical unit that makes nucleic acids (DNA and RNA) is called **Nucleotide**.

- A **nucleotide** is composed of:

- Sugar
- Phosphate group
- Nitrogenous base



Copyright © 2009 Pearson Education, Inc.

# DNA and RNA chemical unit

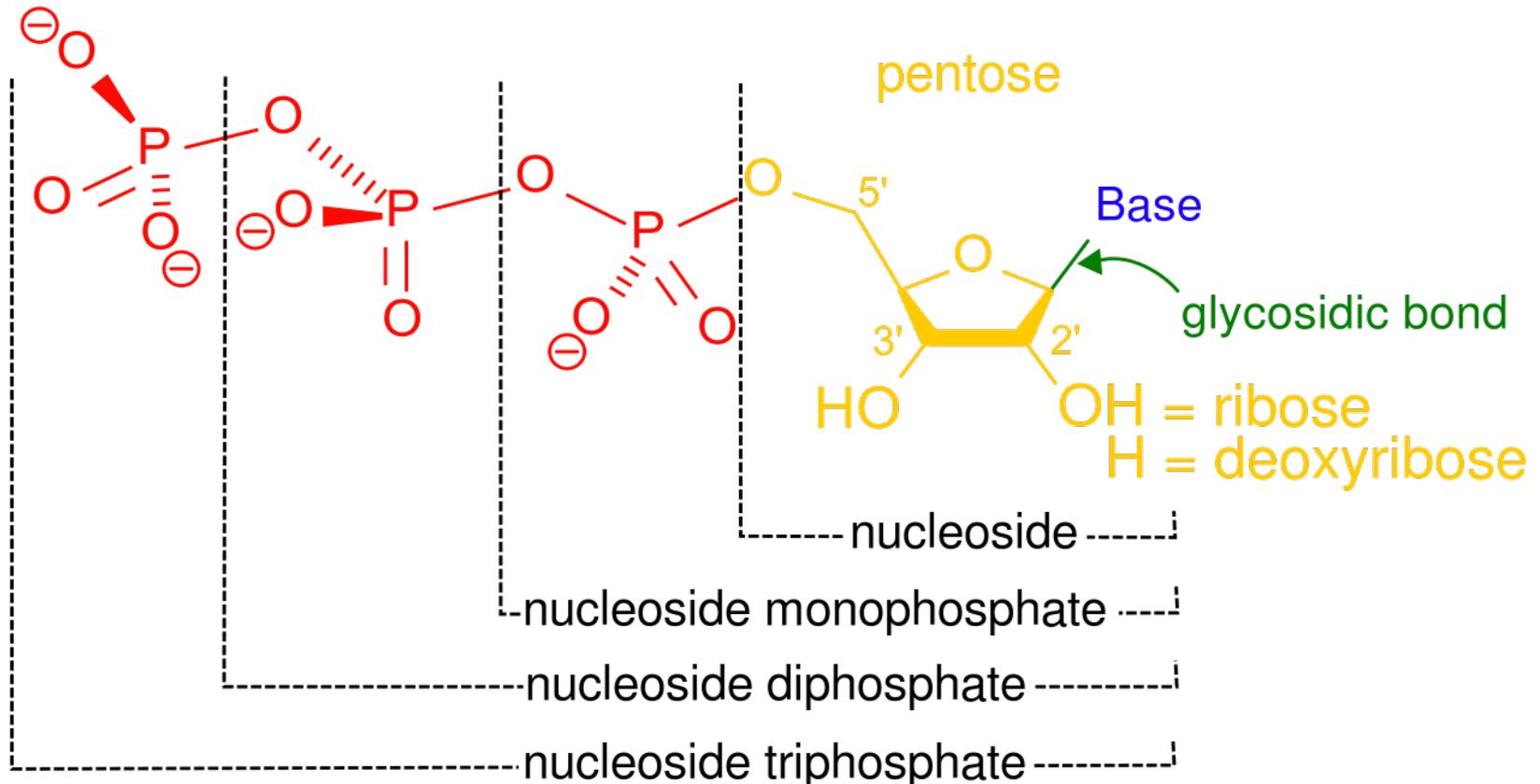


---

**Do you remember what a nucleoside is?  
only for fun :-)**

# The phosphate group

- A number of phosphate groups can be attached to carbon 5 of the sugar.
- NTPs? dNTPs?



# The phosphate group

---

What is the function of the phosphate groups?

Any relation to replication/sequencing?

# The bases

## Nitrogenous bases

Small bases

Large bases

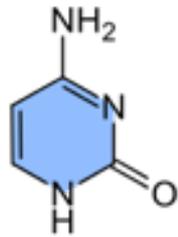
Pyrimidines  
One ring

Purines  
Two rings

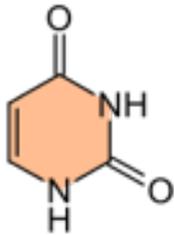
Found in DNA  
and RNA

Cytosine

**C**



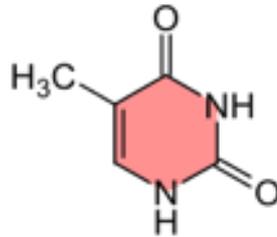
Uracil



**U**

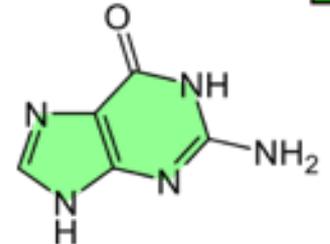
Thymine

**T**



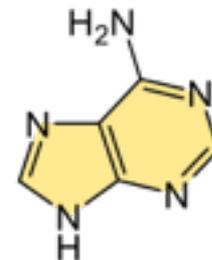
Guanine

**G**



Adenine

**A**

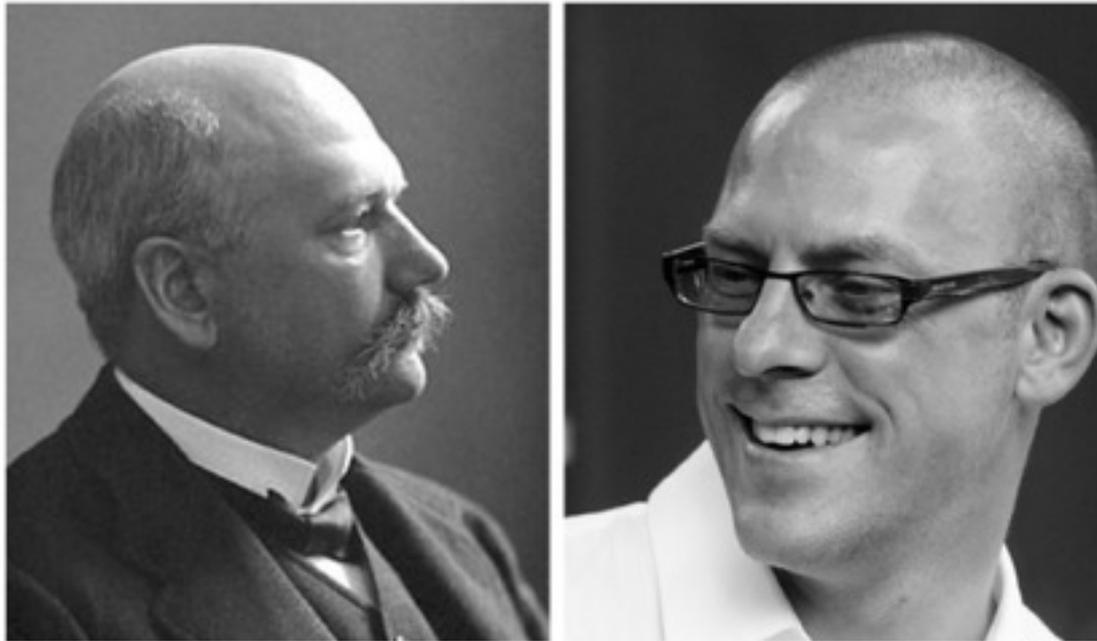


Found in RNA

Found in DNA

# Naming of DNA bases

Celebrating an unsung hero of genomics: how Albrecht Kossel saved bioinformatics from a world of hurt

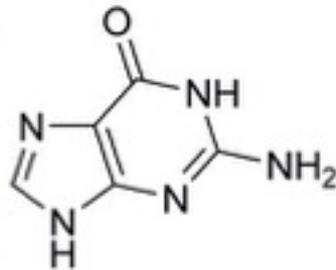


<http://www.acgt.me/blog/2014/3/3/celebrating-an-unsung-hero-of-genomics-how-albrecht-kossel-saved-bioinformatics-from-a-world-of-hurt>

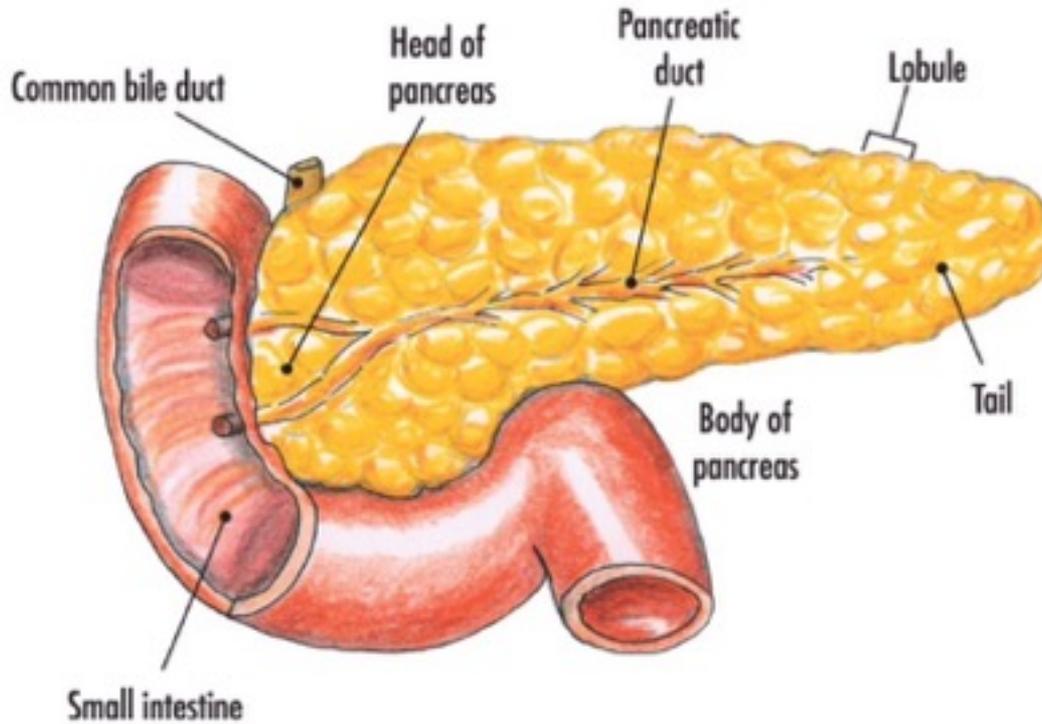
# Naming of DNA bases



Guanine had already been named based on where it had first been discovered (the excrement of seabirds known as **guano**).



# Naming of DNA bases

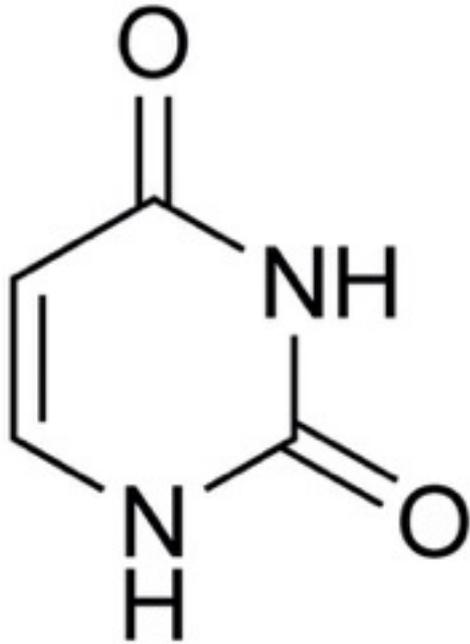


Adenine

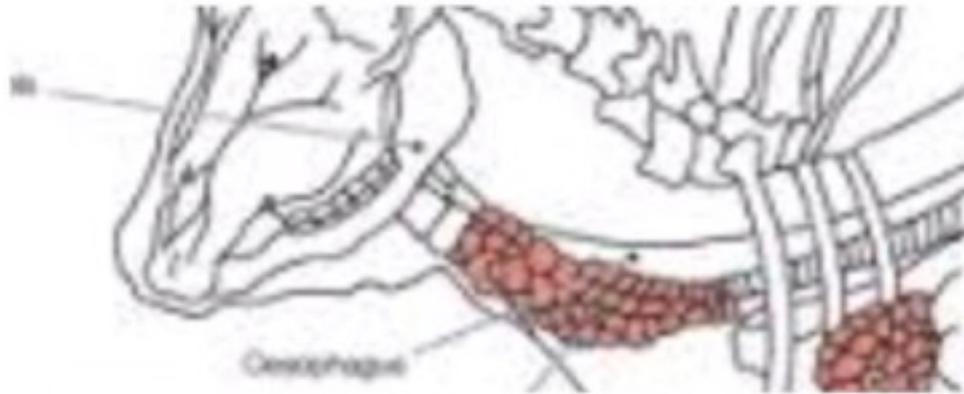
was so named by Kossel because it was isolated from the pancreas gland ('adenas' in Greek).

# Naming of DNA bases

Thymine was named because it was isolated in nucleic acids from the **thymus** of a calf.



# Naming of DNA bases



Cytosine — the last of the four DNA bases to be characterized — was also discovered from hydrolysis of the calf thymus. Its name comes from the original name in German ('cytosin') and simply **refers to** the Greek prefix for cell ('cyto').

# Naming of DNA bases

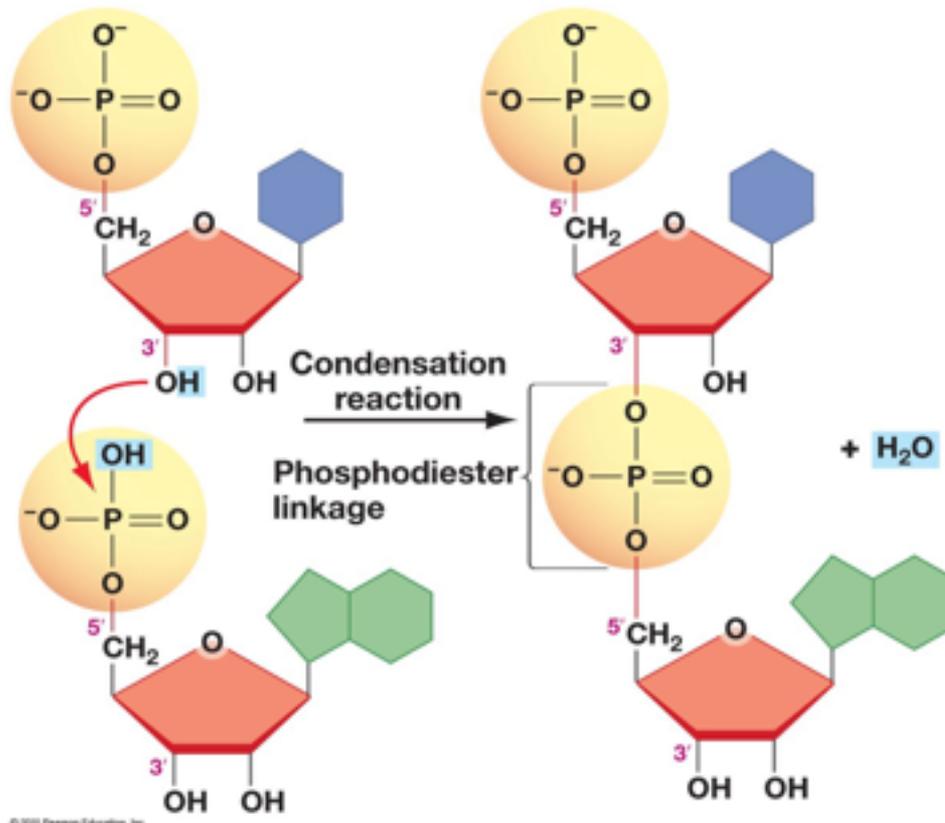


The naming of the DNA bases made it easier to abbreviate the names of the molecules into unique single letters (A,T,G,C).

The single letter representation of DNA enabled application of information theory via bioinformatics.

# Nucleotide linking

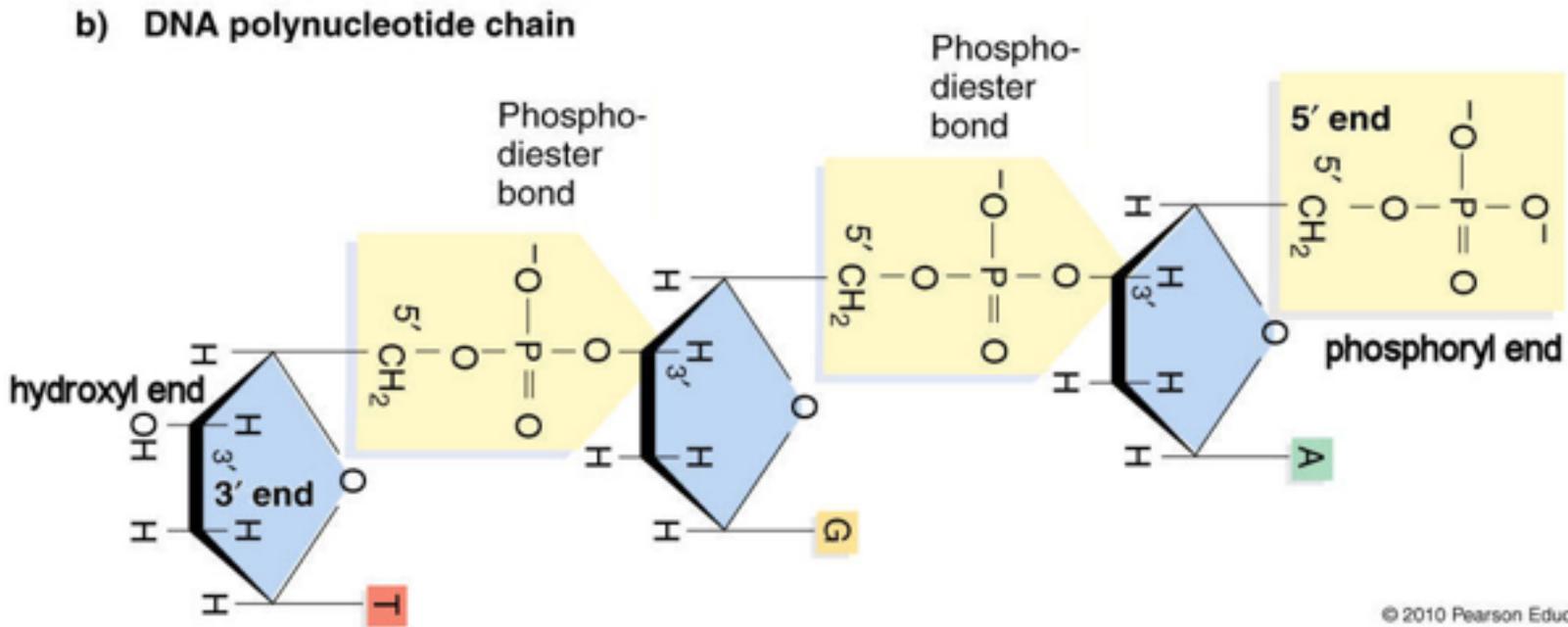
- Nucleotides are linked via **phosphodiester bond**.
- A covalent bond links the phosphate group of one nucleotide to the 3' carbon of the sugar of another.



© 2011 Pearson Education, Inc.

# Polarity

- **5' end:** where the 5' carbon at one end of the molecule has a phosphate group.
- **3' end:** where the 3' carbon at the other end of the molecule has a hydroxyl group.



# The Race to DNA structure



The story of the discovery of the double helix involves these key actors

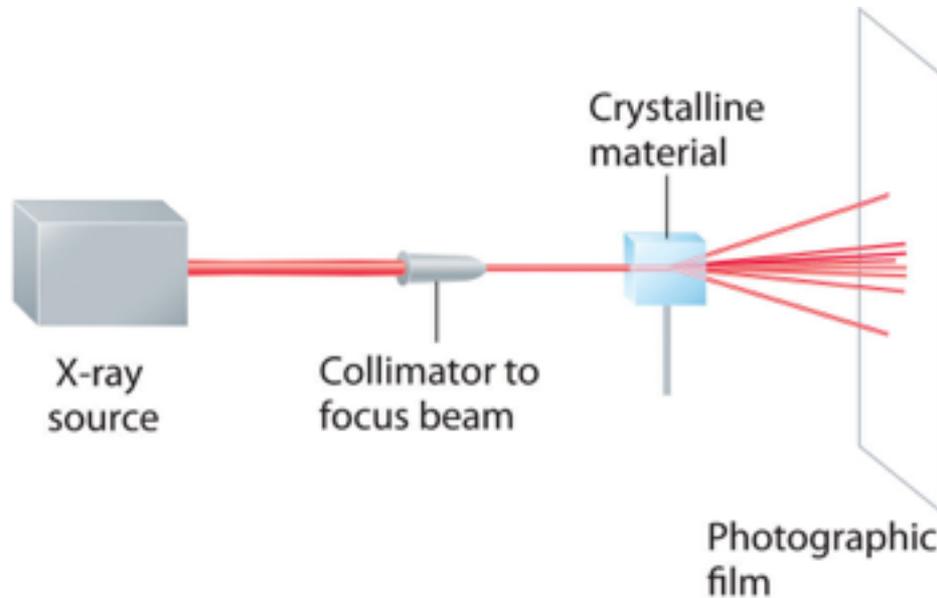
# Molecular structure

---

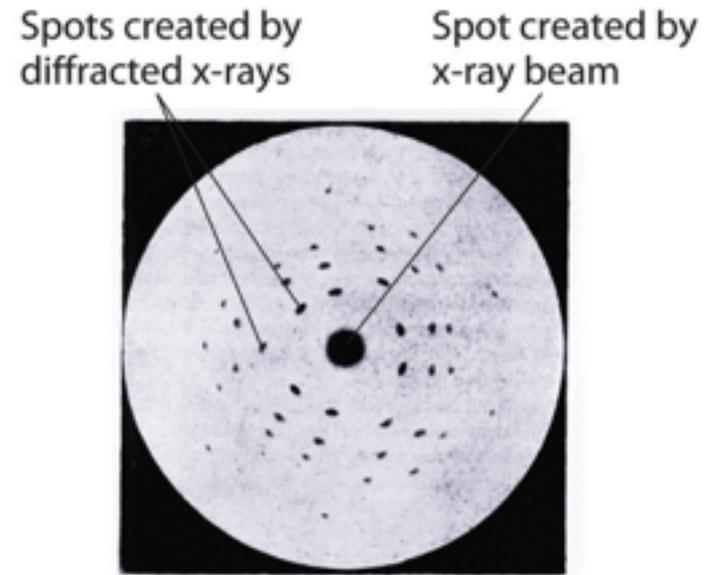
**How can we learn about the structure of molecules?**

# X-ray diffraction

X-ray diffraction was the method to study the fine structure of molecules. DNA was no different!



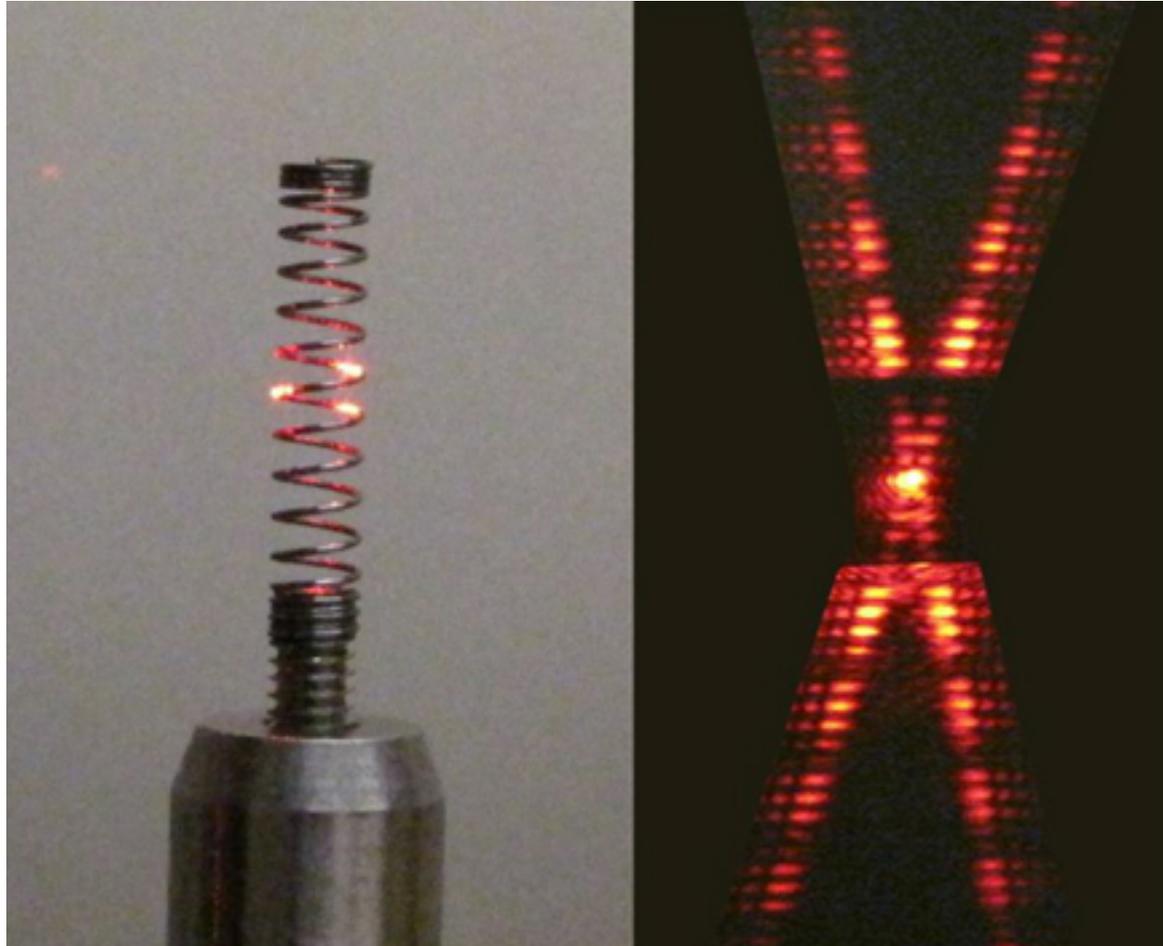
(a) X-ray diffraction



(b) X-ray diffraction pattern captured on photographic film

# X-ray diffraction

Look at the pattern resulting!



# New Method

## THE REVOLUTION WILL NOT BE CRYSTALLIZED

**MOVE OVER X-RAY CRYSTALLOGRAPHY. CRYO-ELECTRON MICROSCOPY IS KICKING UP A STORM IN STRUCTURAL BIOLOGY BY REVEALING THE HIDDEN MACHINERY OF THE CELL.**

BY EWEN CALLAWAY

In a basement room, deep in the bowels of a steel-clad building in Cambridge, a major insurgency is under way.

A hulking metal box, some three metres tall, is quietly beaming terabytes' worth of data through thick orange cables that disappear off through the ceiling. It is one of the world's most advanced cryo-electron microscopes: a device that uses electron beams to photograph frozen biological molecules and lay bare their molecular shapes. The microscope is so sensitive that a shout can ruin an experiment, says Sjors Scheres, a structural biologist at the UK Medical Research Council Laboratory of Molecular Biology (LMB), as he stands dwarfed beside the £5-million (US\$7.7-million) piece of equipment. "The UK needs many more of these, because there's going to be a boom," he predicts.

In labs around the world, cryo-electron microscopes such as this one are sending tremors through the field of structural biology. In the past three years, they have revealed exquisite details of protein-making ribosomes, quivering membrane proteins and other key cell molecules,

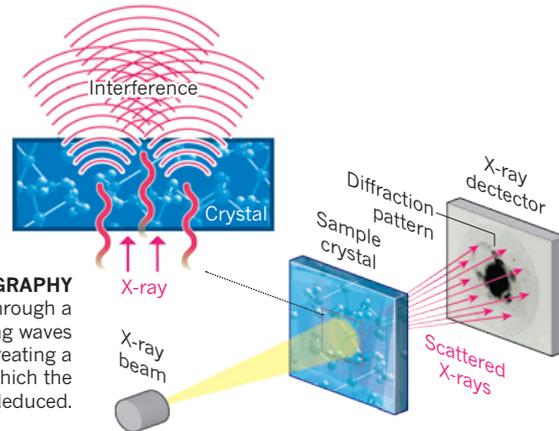
ILLUSTRATION BY VICTOR KOSTIN

172 | NATURE | VOL 525 | 10 SEPTEMBER 2015

### STRUCTURE SOLVERS

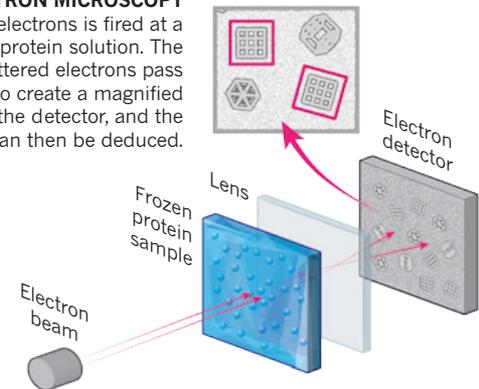
X-ray crystallography has long been the dominant method for deducing high-resolution protein structures, but cryo-electron microscopy is catching up.

**X-RAY CRYSTALLOGRAPHY**  
X-rays scatter as they pass through a crystallized protein; the resulting waves interfere with each other, creating a diffraction pattern from which the position of atoms is deduced.



### CRYO-ELECTRON MICROSCOPY

A beam of electrons is fired at a frozen protein solution. The emerging scattered electrons pass through a lens to create a magnified image on the detector, and the structure can then be deduced.



X-RAY IMAGE: SPL





# Watson and Crick

- They published a 900 words paper and Franklin and Wilkin also published on the same issue of Nature.



April 25, 1953 NATURE 737

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

<sup>1</sup>Yang, T. S., Stewart, P., and James, W., *Phil. Mag.*, **10**, 149 (1935).

<sup>2</sup>Levine, Miriam, M. S., *Ann. Ent. Soc. Am.*, **46**, 464 (1953).

<sup>3</sup>Lee, H. S., U. S. World War Paper in Phys. Chem., **57**, 111 (1943).

<sup>4</sup>Stewart, T. W., *Ann. Ent. Soc. Am.*, **46**, 464 (1953).

### MOLECULAR STRUCTURE OF NUCLEIC ACIDS

#### A Structure for Deoxyribonucleic Acid

WE wish to suggest a structure for the salt of deoxyribonucleic acid (DNA). This structure has several features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey<sup>1</sup>. They kindly made their manuscript available to us in advance of publication. Their model consists of three inter-twined 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 diagram 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.

We wish to put forward a radically different structure for the salt of deoxyribonucleic 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 joined 3'-to-5'-deoxyribose residues with P-O linkages. The two chains that 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 sequence of the atoms in the two chains run in opposite directions. Each chain locally resembles Pauling's model (Fig. 1), that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the atoms in the two chains run in opposite directions. It is close to Pauling's 'standard configuration', the angle being roughly perpendicular to the stacked base. There is a methyl on each chain every 3.4 Å, in the 2-direction. We have assumed an angle of 30° between adjacent methyls in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 Å. The distance of a phosphate atom from the fibre axis is 10 Å. As the phosphates are on the outside, outside links are easy access to them.

The structure is an open one, and its water content is rather high. As fewer water molecules we would expect the bases to fill so that the structure would become more compact.

The usual 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 *anti*-rotation. One of the pair must be a guanine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; guanine position 6 to pyrimidine position 6.

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 configuration) 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<sup>2,3</sup> that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribonucleic acid.

It is probably impossible to build the structure with a slight rigidity 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<sup>4</sup> on deoxyribonucleic 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 communication. We were not aware of the details of the results presented there when we devised our structure, which was mainly thought out entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific points we have postulated immediately suggest 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 coordinates for the atoms, will be published elsewhere.

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

# DNA structure



- 1) DNA is a double helix.
- 2) Two polynucleotides chains.
- 3) The two chains wind around right handedly - right handed double helix.
- 4) The two chains are in an anti-parallel orientation. One strand 5' – 3' orientation and the other 3' – 5').
- 5) Sugar-phosphate backbone is located on the outside of the helix.
- 6) The nitrogenous bases located on the inside of the helix.

# DNA structure



7) The bases are stacked flat and perpendicular to the axis of the helix. The bases are on top of each other following the twist of the helix.

8) The bases of the two polynucleotides are bonded together via hydrogen bonds on the inside of the helix.

9) Bases of the two polynucleotide chains are base-pairing in a combination that maintains similar diameter of the double helix.

10) A Pyrimidine always basepair with Purine forming **complementary base pairs**.

# DNA structure



11) Thymine (T) basepair with Adenine (A), and Cytosine basepair with Guanine (G).

12) Two hydrogen bonds involve the base-pairing of (A and T) and three hydrogen bonds between (G and C).

13) The sequence of one chain (strand) is enough to predict the complementary one in the other orientation.

14) A major and minor groove result from the unequal spacing of the phosphate-sugar backbone.





---

# Genome packaging and organization

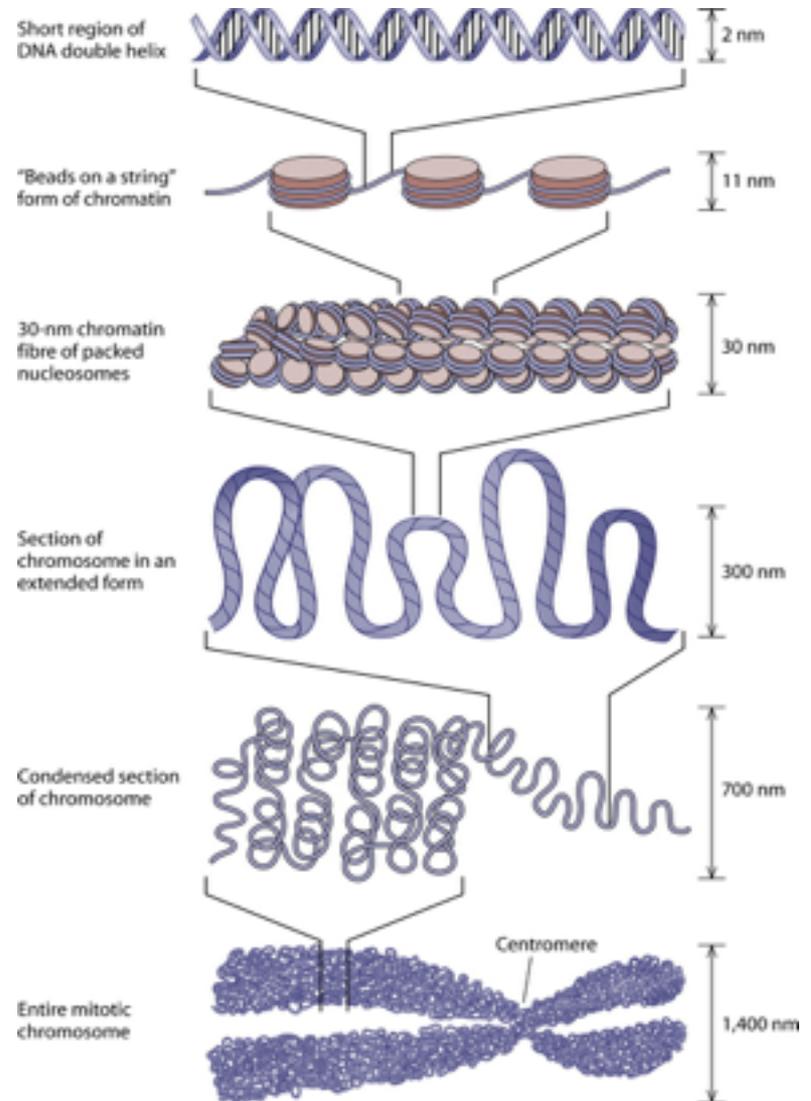


**Do you remember the organization of  
viral genomes?**

**prokaryotic genomes?**

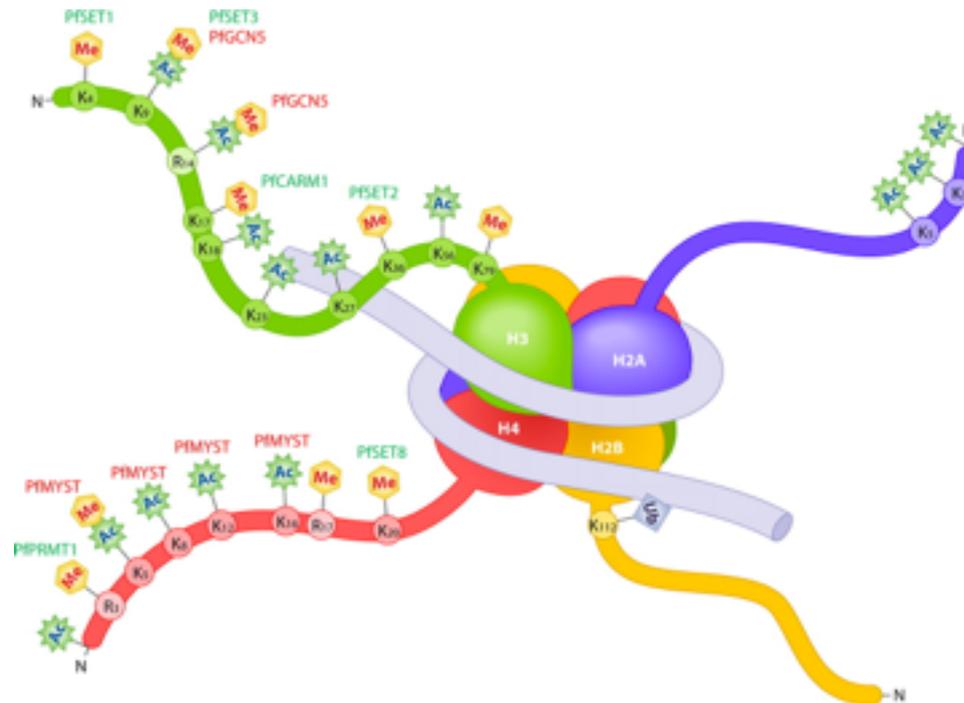
# Review of genome condensation

- Double helix.
- Nucleosomes (DNA histones).
- Beads on a string.
- 30 nm chromatin fiber/  
solenoid.
- Eu/Heterochromatin.
- Metaphase  
chromosome.

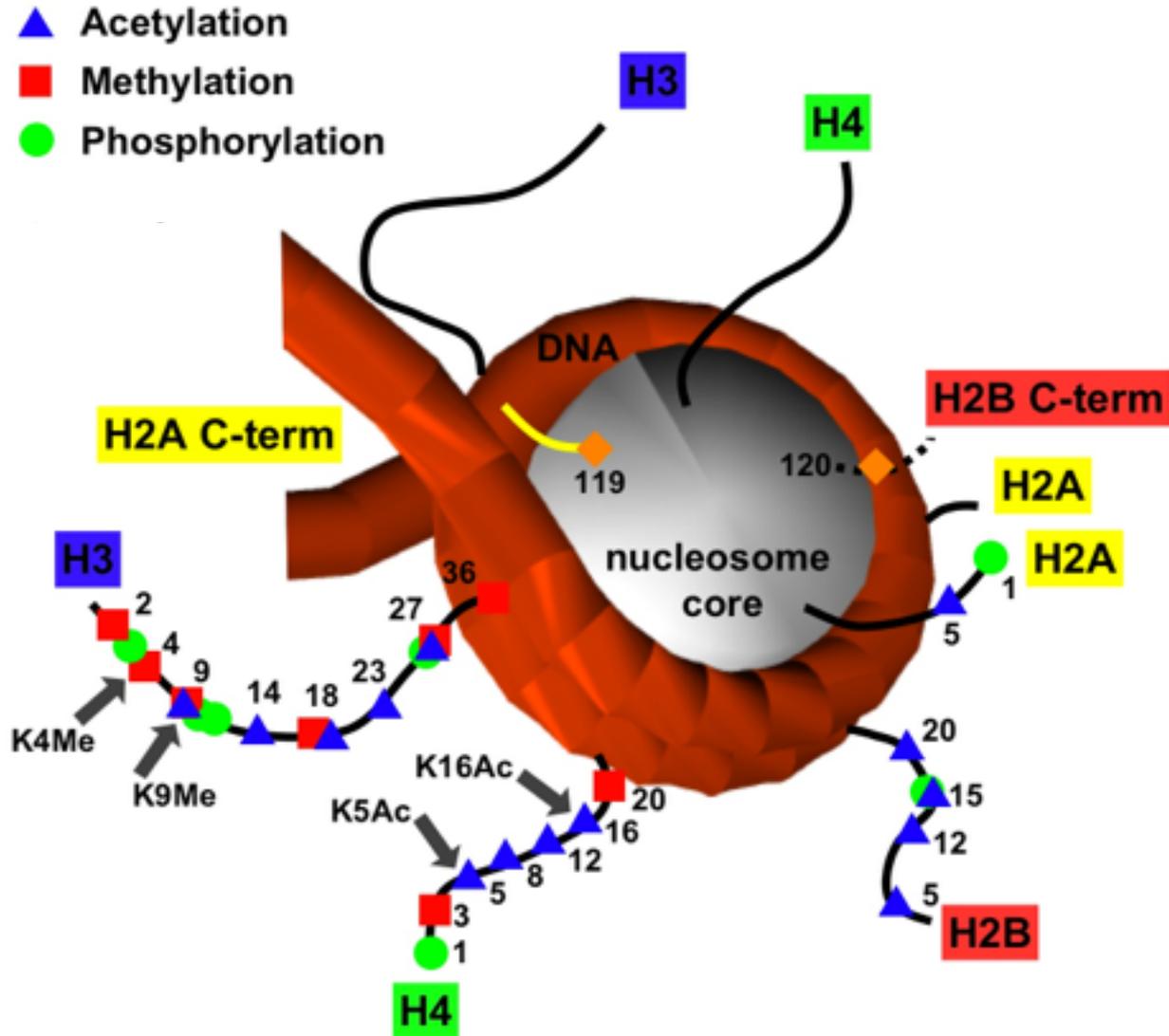


# DNA packaging

- A single nucleosome is composed of 8 histone proteins (octamer).
- Histone proteins contain tails of amino acids that can be modified.



# Histone tail modifications



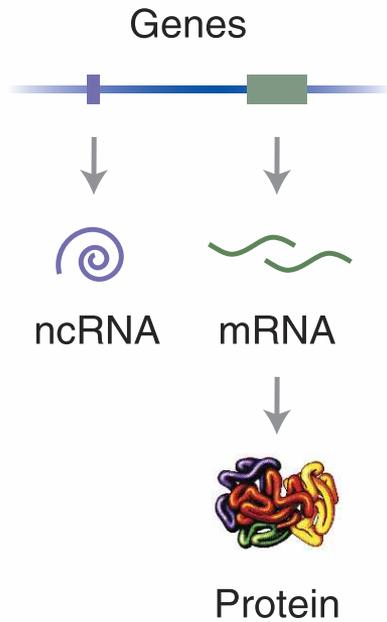
# Histone tail modifications



**Do you remember the functions/effects of histone modifications?**

**Are the histone modifications uniform across all chromosomes?**

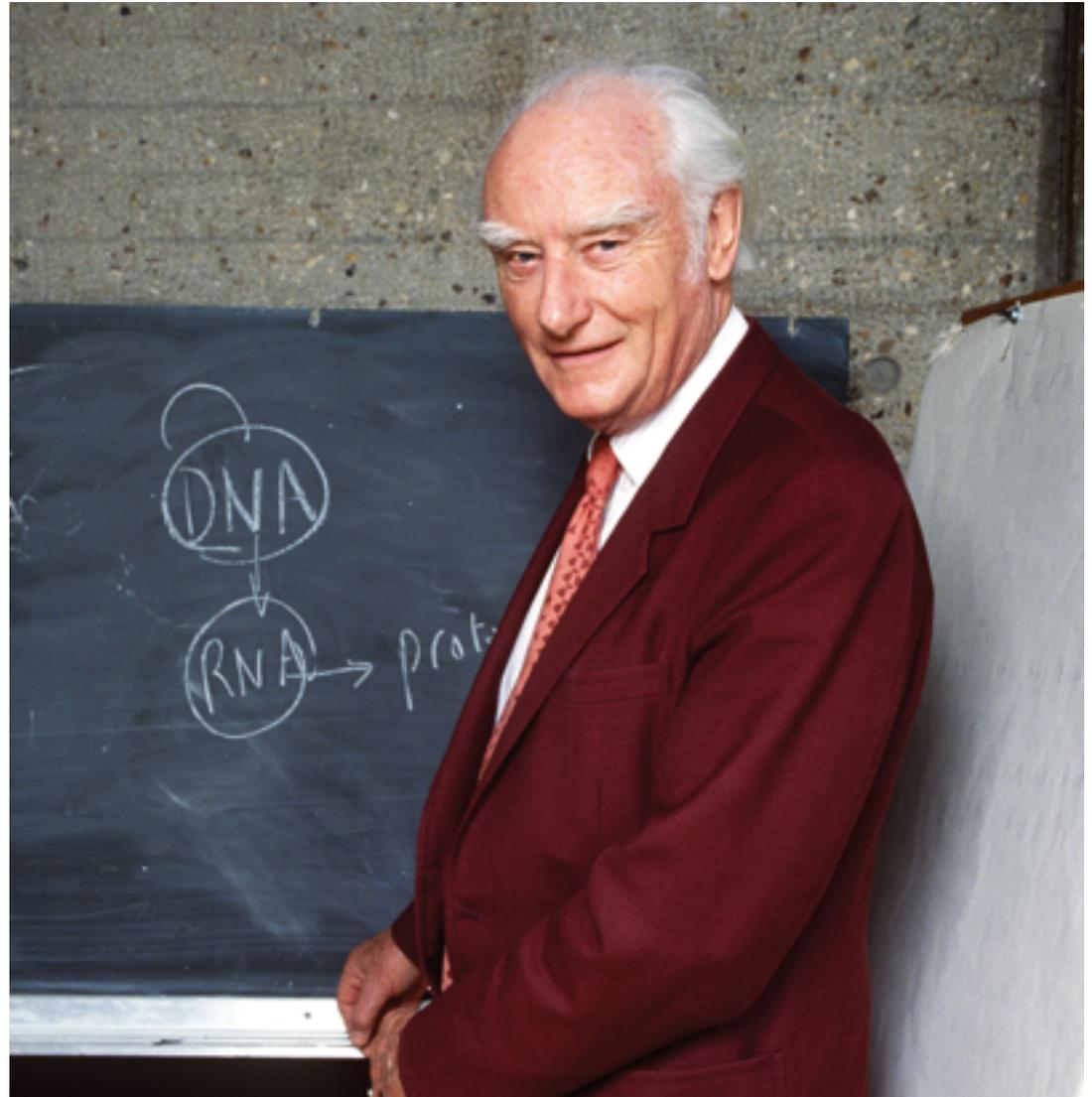
# The Central Dogma



The '**Central Dogma**' of molecular biology was proposed by Francis Crick.

⋮

1958



# The Central Dogma

---

**Do you what the word “Dogma” means?**

**Do we have dogmas in science?**

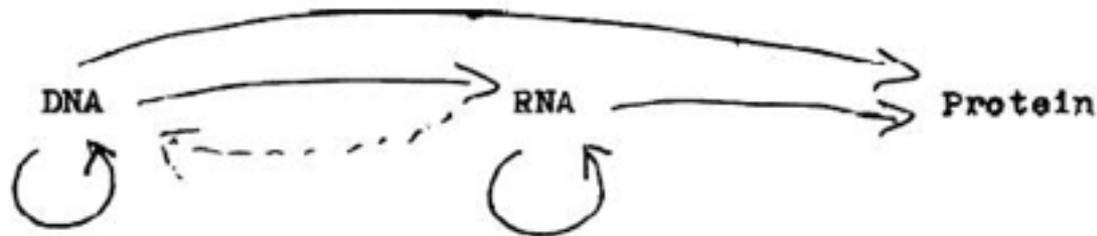
# The Central Dogma

Ideas on Protein Synthesis (Oct. 1956)

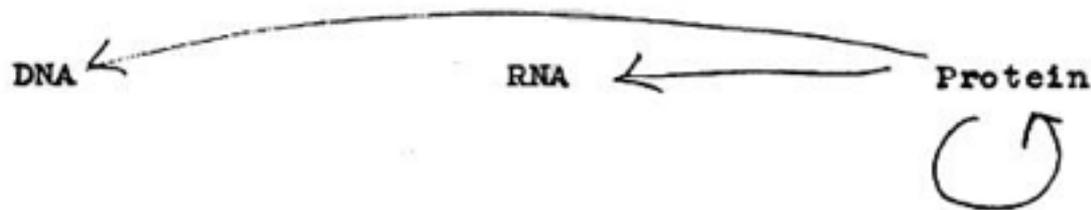
The Doctrine of the Triad.

The Central Dogma: "Once information has got into a protein it can't get out again". Information here means the sequence of the amino acid residues, or other sequences related to it.

That is, we may be able to have



but never



where the arrows show the transfer of information.

# The Central Dogma

Molecules as information and information flow

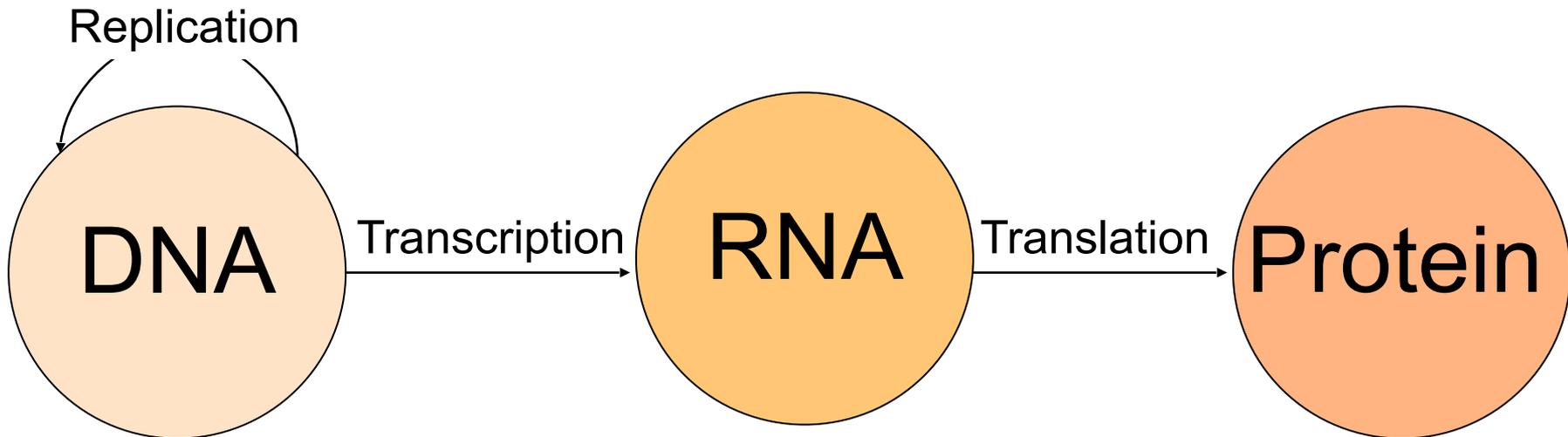


... article that appeared six weeks later, Crick and Watson proposed a hypothesis with regard to the function of the 'bases' – the four kinds of molecule that are spaced along each strand of the double helix and which bind the two strands together. **They wrote: 'it therefore seems likely that the precise sequence of the bases is the code which carries the genetical information.'**

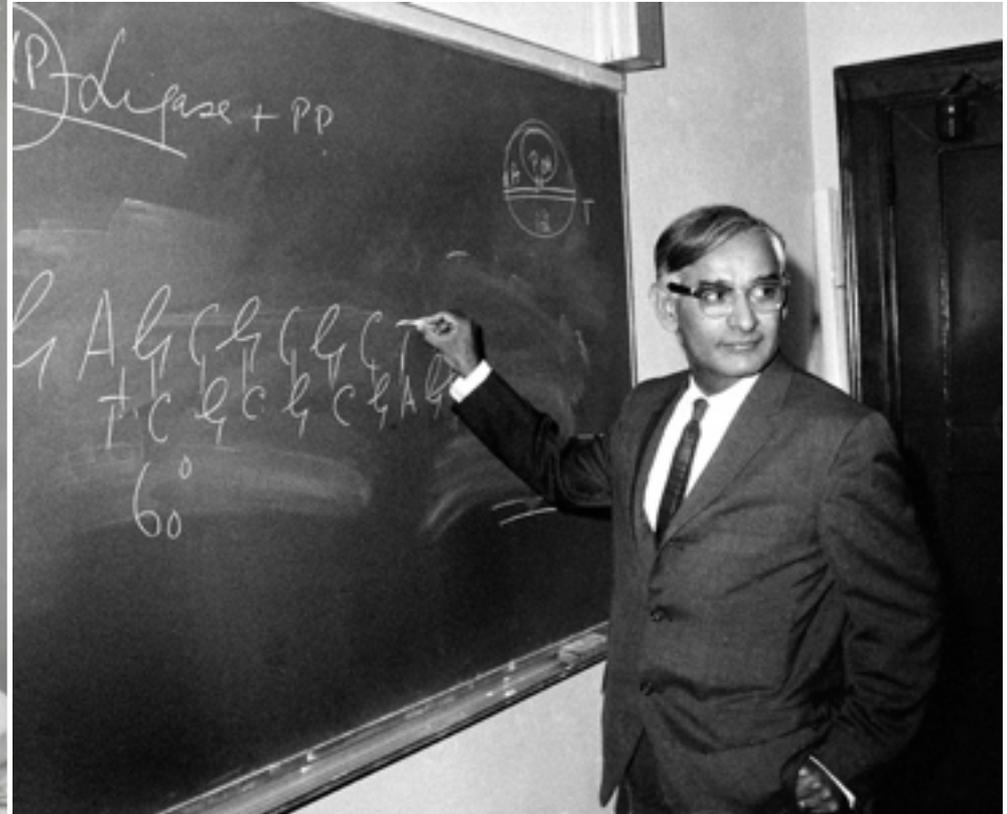
This phrase, which was almost certainly the work of Crick, must have seemed both utterly strange and completely familiar to those

# Terms and processes

The flow of information from DNA to proteins go through an intermediate molecule and involves two processes.



# The code



1965: Marshall Nirenberg and Har Gobind Khorana (and others)

# The code revealing experiments



What scientific developments enabled discovering the genetic code?

What were the experiments based on?

# The code revealing experiments



Major advances:

- (1) Cell free protein synthesis
- (2) Artificial (synthetic) m-RNA

# The code revealing experiments

## Decoding experiments

### Stage 1

Mononucleotide polymers  
Random copolymers

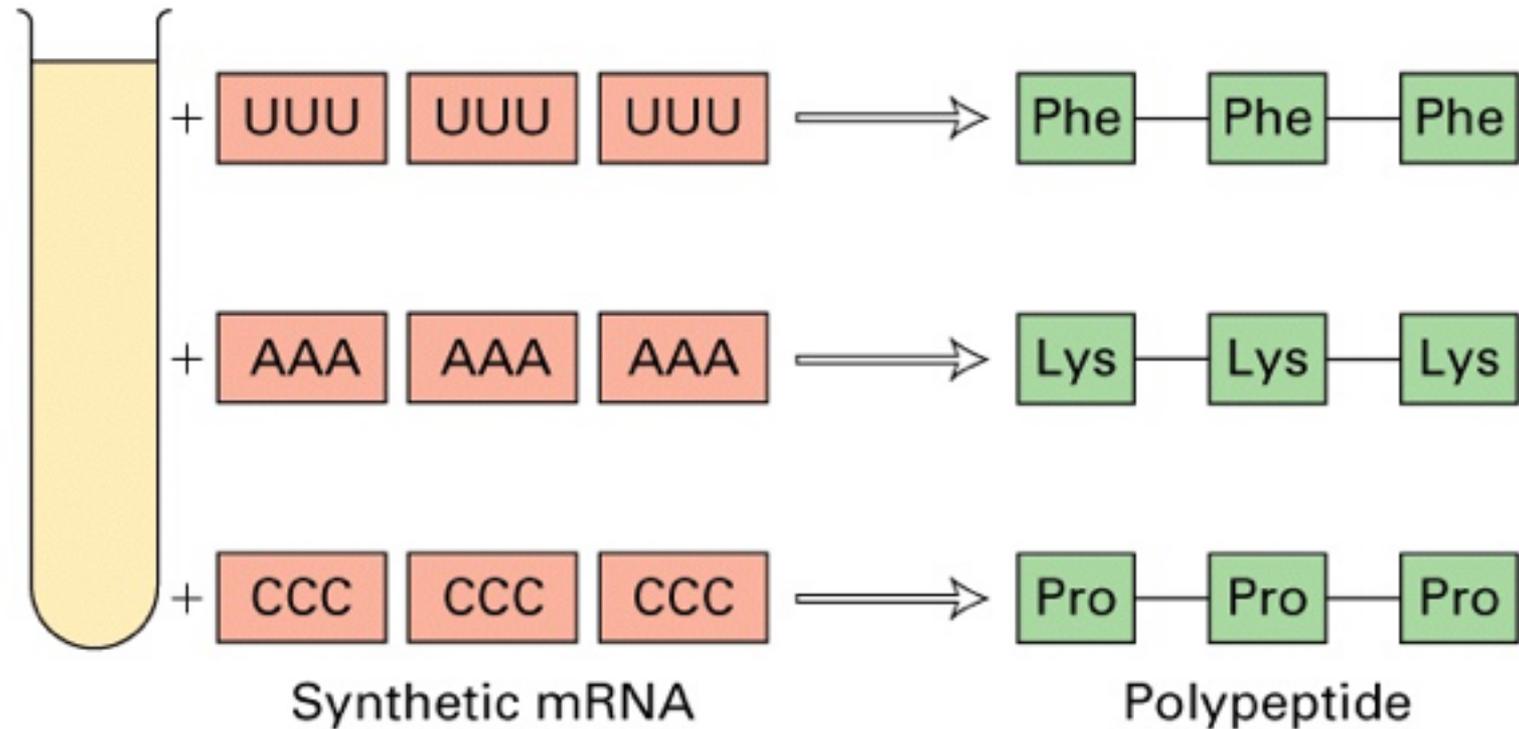
### Stage 2

Ribosome binding  
Assay

# Mononucleotide polymers

Translating a mononucleotide polymer results in a single amino acid polypeptide

Bacterial extract



# Random copolymers



## Historical review: Deciphering the genetic code – a personal account

Marshall Nirenberg

Polynucleotides	Amino acids			
U	PHE			
C	PRO			
A	LYS			
G	—			
UC	LEU	SER		
UA	LEU	TYR	ILE	ASN
UG	LEU	VAL	CYS	TRP
CA	HIS	THR	GLN	ASN
CG	ARG	ALA		
AG	ARG	GLU		
UAG	ASP	MET		
CAG	ASP	SER		

**Figure 6.** The specificity of randomly ordered polynucleotide templates in stimulating amino acid incorporation into protein in *Escherichia coli* extracts. Only the minimum kinds of bases necessary for template activity are shown, so many amino acids that respond to randomly ordered polynucleotides composed of two or more kinds of bases are omitted. The base compositions of RNA codons were derived from these experiments [49]. Reproduced from Ref. [49].



# Random copolymers

---

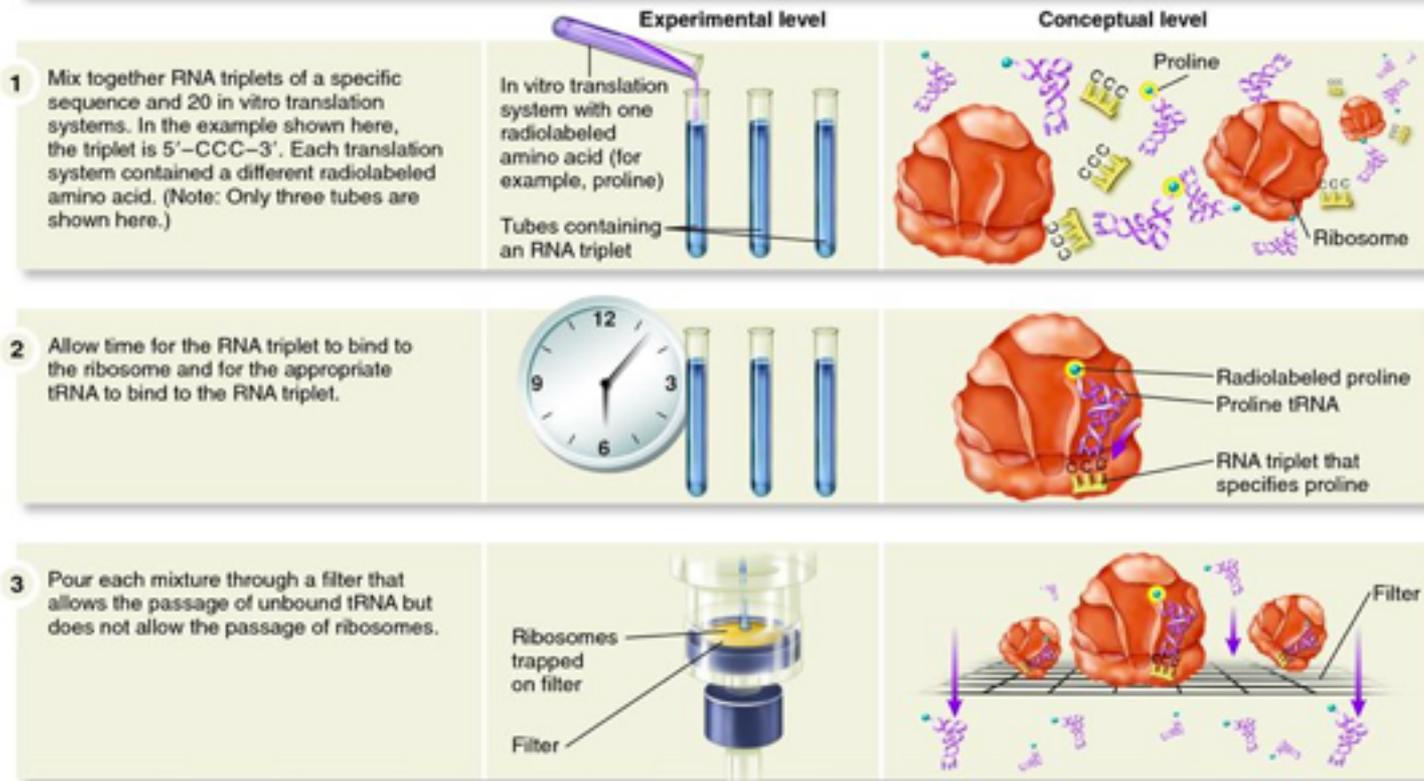
**Can random copolymers  
definitively identify the genetic  
code?**

# Ribosome binding assay

This technique revealed the language of the genetic material and the information hidden in it

**HYPOTHESIS** An RNA triplet can bind to a ribosome and promote the binding of the tRNA that carries the amino acid that the RNA triplet specifies.

**KEY MATERIALS** The researchers made 20 in vitro translation systems, which included ribosomes, tRNAs, and 20 amino acids. The 20 translation systems differed with regard to which amino acid was radiolabeled. For example, in one sample, radiolabeled glycine was added, and the other 19 amino acids were unlabeled. In a different sample, radiolabeled proline was added, and the other 19 amino acids were unlabeled. The tRNA preparation also contained the enzymes that attach amino acids to tRNAs.



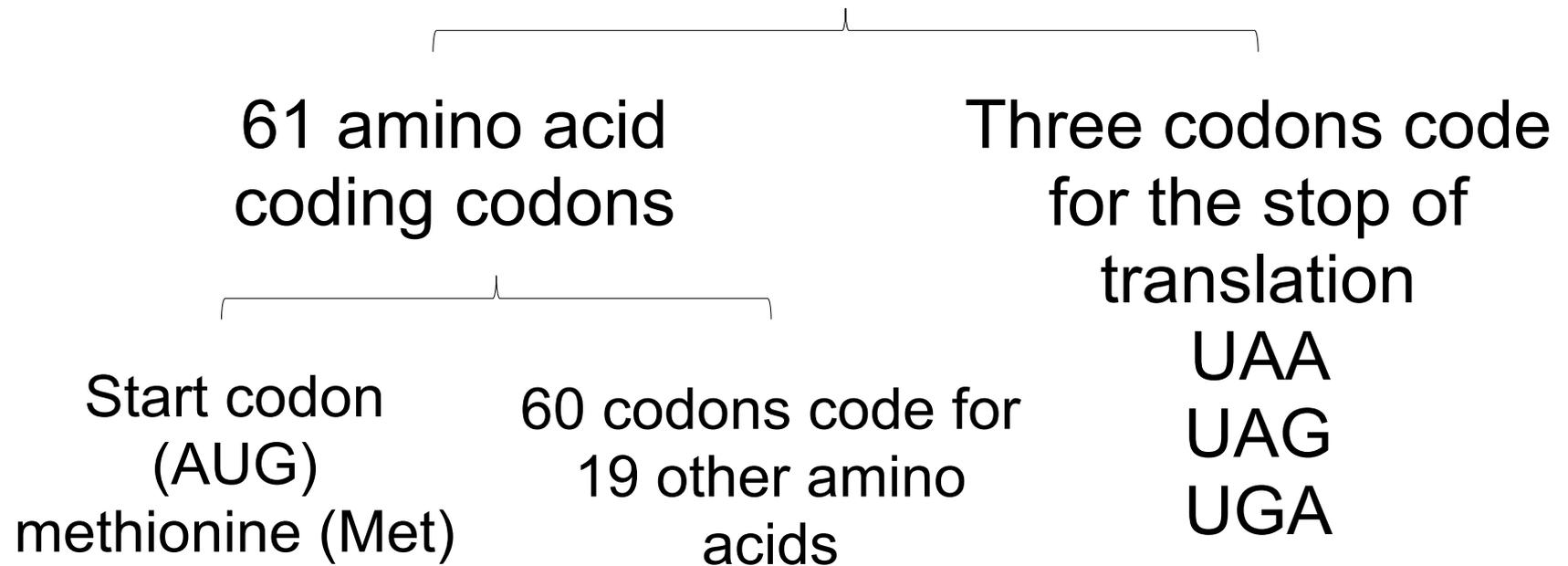


**What is the name of the protein coding unit  
of genetic information?**

**How many nucleotides it contains?**

# The genetic code

The genetic code is composed of 64 codons



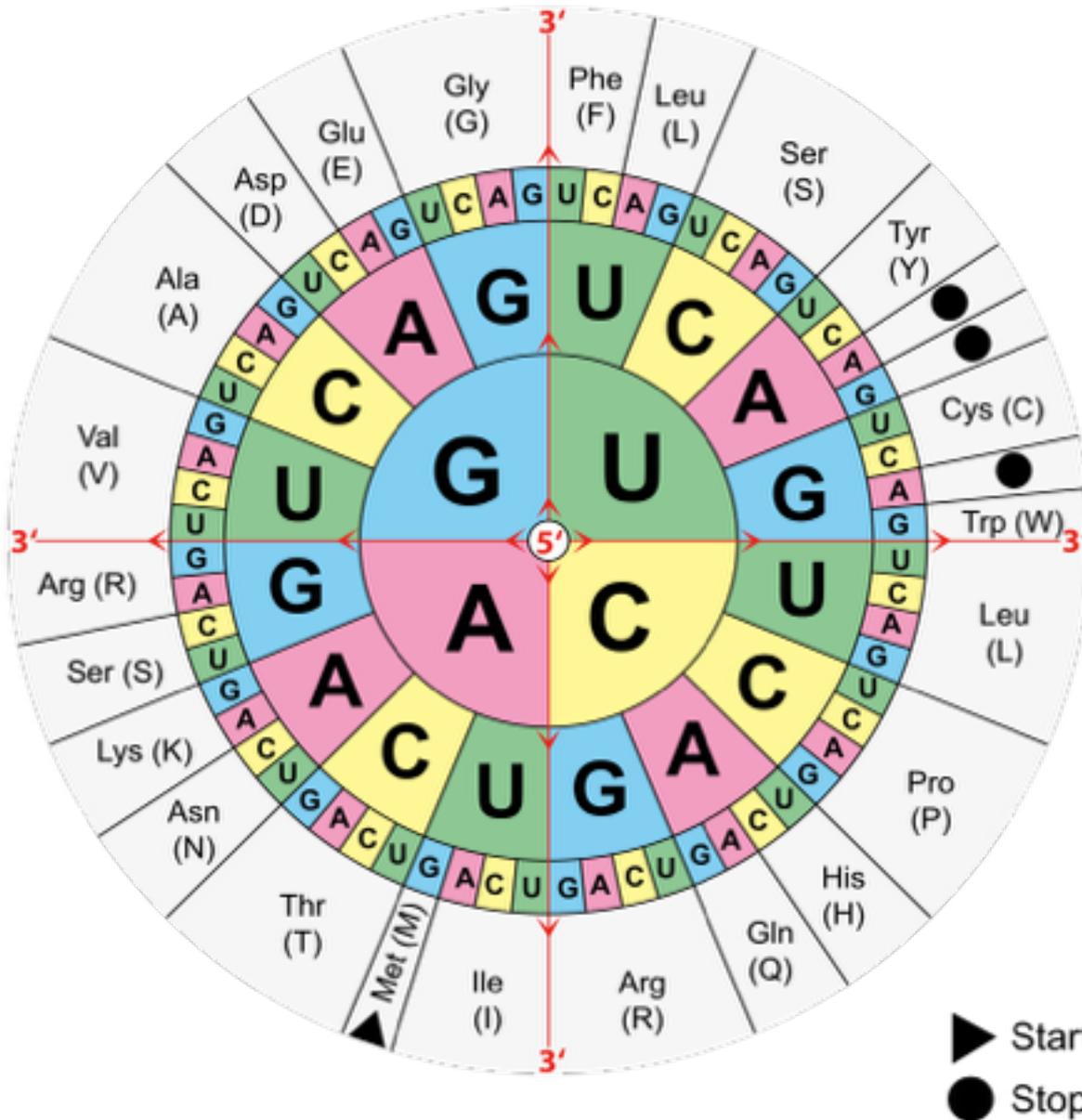
# Characteristics of the genetic code

1. The genetic code is made of triplets of nucleotides (3nts) called codons.
2. The genetic code is continuous (no skipping).
3. The code is not overlapping. Every three nucleotides in a sequence code for one codon.
4. The genetic code is universal (almost).
5. The code has specific signals for start of translation and stop of translation.
6. The genetic code is “degenerate”.
7. The Wobble effect of the third base in the codon.

# Characteristics of the genetic code

		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G

# The genetic code



# Expectations

---

- You are familiar, broadly, with the concepts covered.

# Disclaimer

---

Figures, photos, and graphs in my lectures are collected using google searches. I do not claim to have personally produced the material (except for some). I do cite only articles or books used. I thank all owners of the visual aid that I use and apologize for not citing each individual item. If anybody finds the inclusion of their material in my lectures a violation of their copy rights, please contact me via email.

[hhalhaddad@gmail.com](mailto:hhalhaddad@gmail.com)