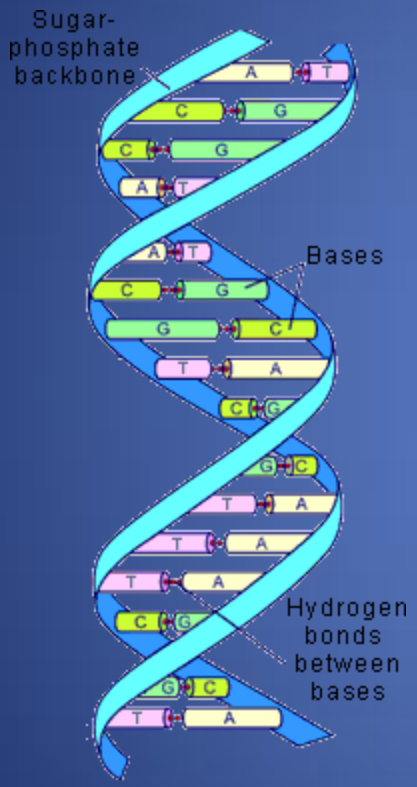
Marshall Nirenberg and the discovery of the Genetic Code

The Coding Problem

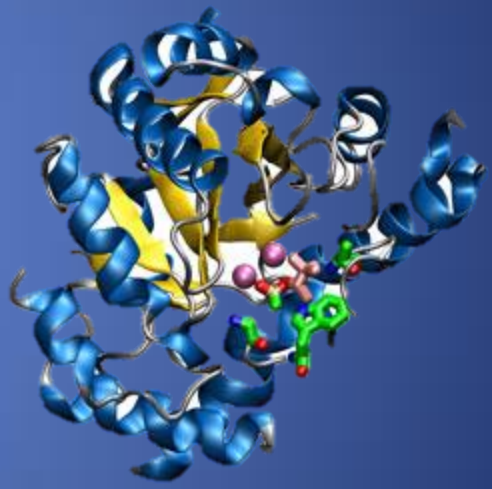
- Once the *function* of DNA as the genetic substance was shown by Avery et al in 1944
- And once the double helical *structure* of DNA was described by Watson and Crick in 1953
- The mystery still remained, how was the sequence of bases in DNA translated and expressed into the sequence of amino acids in proteins?
- This was known as the *coding problem*

Protein biosynthesis



?

in the cell



DNA

In nucleus

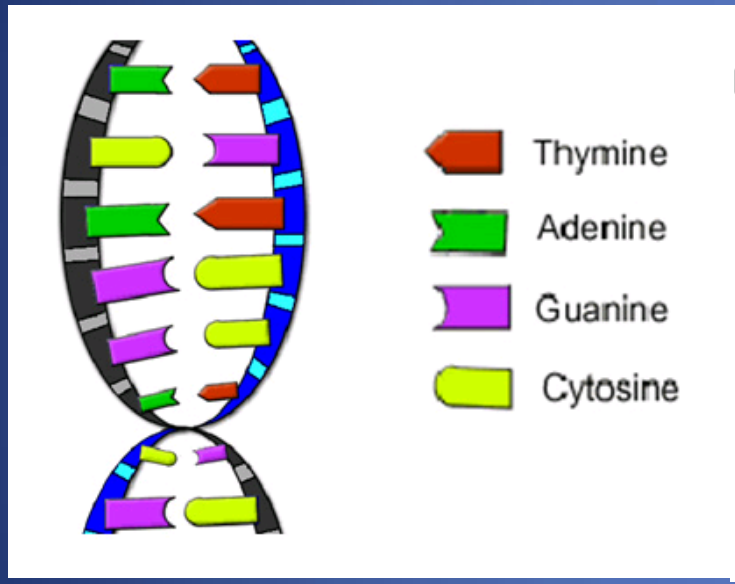
Protein

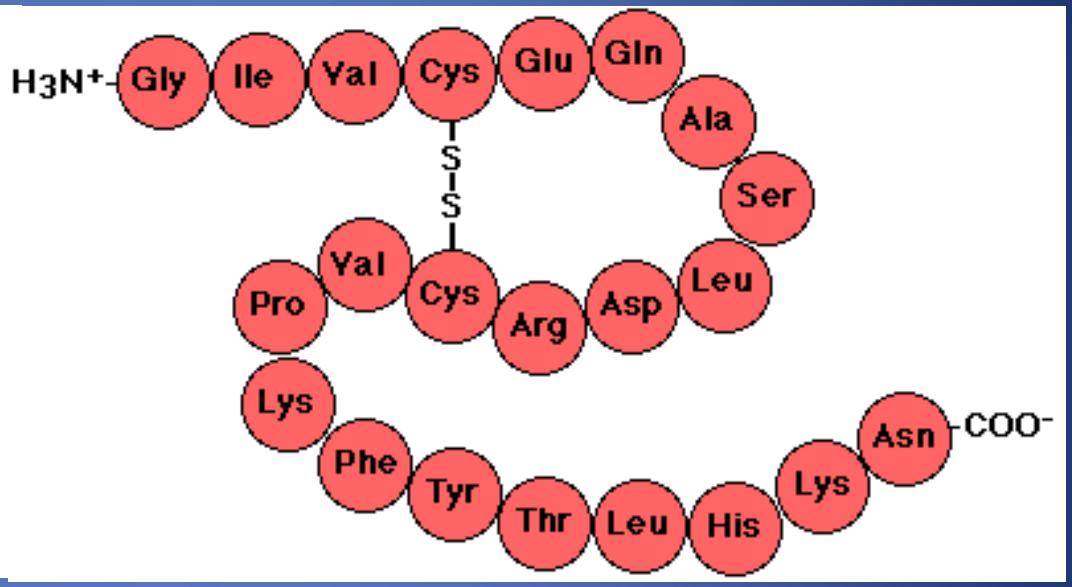
In cytoplasm

The Coding Problem

- It had been found that there are only 4 bases in DNA, abbreviated to A, G, C, T
- Of course, they can be found in any sequence in DNA, giving great variability as needed for genetic expression
- But, there are 20 common amino acids (aa's) in proteins
- So how could 4 bases code for 20 aa's? This was the coding problem

Linear polymers





DNA

Protein

The Coding Problem - Translation

- Clearly the code could not be one base per aa since there were only 4 bases
- Neither could it be two bases, since there are only 16 combinations of 4 bases
- But, three bases per aa was possible since a triplet code gave 64 possible combinations
- Some people assumed a triplet code, but at that time there was no evidence for it
- Also, what happened to the “extra” 44 possible combinations, were they not expressed into aa or was there a degenerate code (i.e more than one combination per aa)
- The code could also be more than 3 bases or could be an over-lapping code

The Coding Problem - Expression

And how was the code expressed?

- Was the DNA itself the template for the synthesis of the peptide bonds making up proteins? Some people devised complex combinations of DNA and proteins.
- Or was there an intermediate that transferred the code from the DNA in the nucleus to the protein synthesis apparatus (the ribosome) in the cell's cytoplasm (this was known as the "adaptor hypothesis")

The RNA Tie Club

- After solving the structure of DNA and realizing that the next major problem was the coding problem Watson and Crick formed the “RNA Tie Club” in 1953
- This was restricted to 20 members (the no. of aa’s) and consisted of the greatest names in molecular biology, including Ochoa, Gamow, Luria, Orgel, Delbruck, Feynman etc., six of whom would win a Nobel Prize
- They were sure they would solve the coding problem between them
- However, most of their methods were theoretical and even involved assumptions, such as the “adaptor hypothesis”

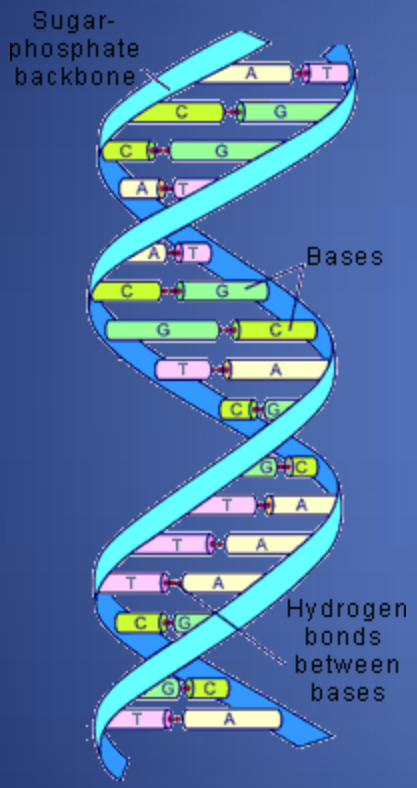
Marshall Nirenberg

- Marshall Nirenberg was born in NY in 1927
- When he was 10 years old his family moved to Orlando Fl, because he had rheumatic fever and his father was bankrupt
- Marshall grew up in a rural area where he was able to roam and developed a keen interest in biology
- He went to Univ Minn to do his PhD in Biochemistry and then received a Fellowship to the NIH in 1957
- Why was he considered "the least likely man" to solve the coding problem: he was young (30), he was unknown, he was obscure and he had many famous rivals and he was late on the scene

Protein biosynthesis

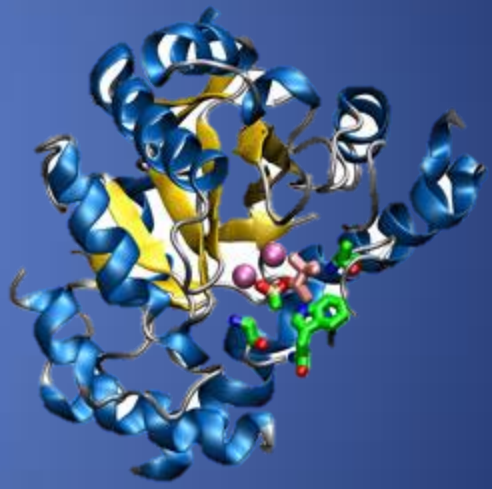
- It was known that the synthesis of proteins occurred on ribosomes, microgranules that are present in the cytoplasm, and not in the nucleus
- Also, degradation of DNA by an enzyme called DNAase did not stop the synthesis of proteins
- So it was concluded that DNA could not directly be involved in the synthesis of proteins
- So there had to be an intermediate that transferred the code from the DNA to the ribosome, that was proposed by Jacob and Monod to be an RNA
- This intermediate was later discovered by Nirenberg & Matthaei and Jacob, Brenner & Meselson in 1961 to be messenger RNA (mRNA)

Protein biosynthesis



DNA
In nucleus

mRNA



Protein
In cytoplasm

RNA

- RNA is a cousin to DNA, it is also a nucleic acid, but it has a different sugar component
- This gives it a different structure to DNA
- Three kinds of RNA were then known –
- Ribsomal or rRNA that made up the ribosomes in combination with some proteins
- Soluble RNA or sRNA that was later identified as transfer or tRNA that delivered each amino acid to the ribosome for synthesis
- And finally mRNA, the intermediate that provided the code for the synthesis of the protein

Cell free system

- But understanding the process of protein biosynthesis did not solve the coding problem
- This was solved by Nirenberg using a *cell free system (cfs)*
- A cfs consists of all the necessary ingredients from lysed cells that can be used to synthesize protein if the appropriate mRNA is present or is added
- The cfs contains: ribosomes, tRNAs (20 kinds), salts (Mg, Na), enzymes (protease)

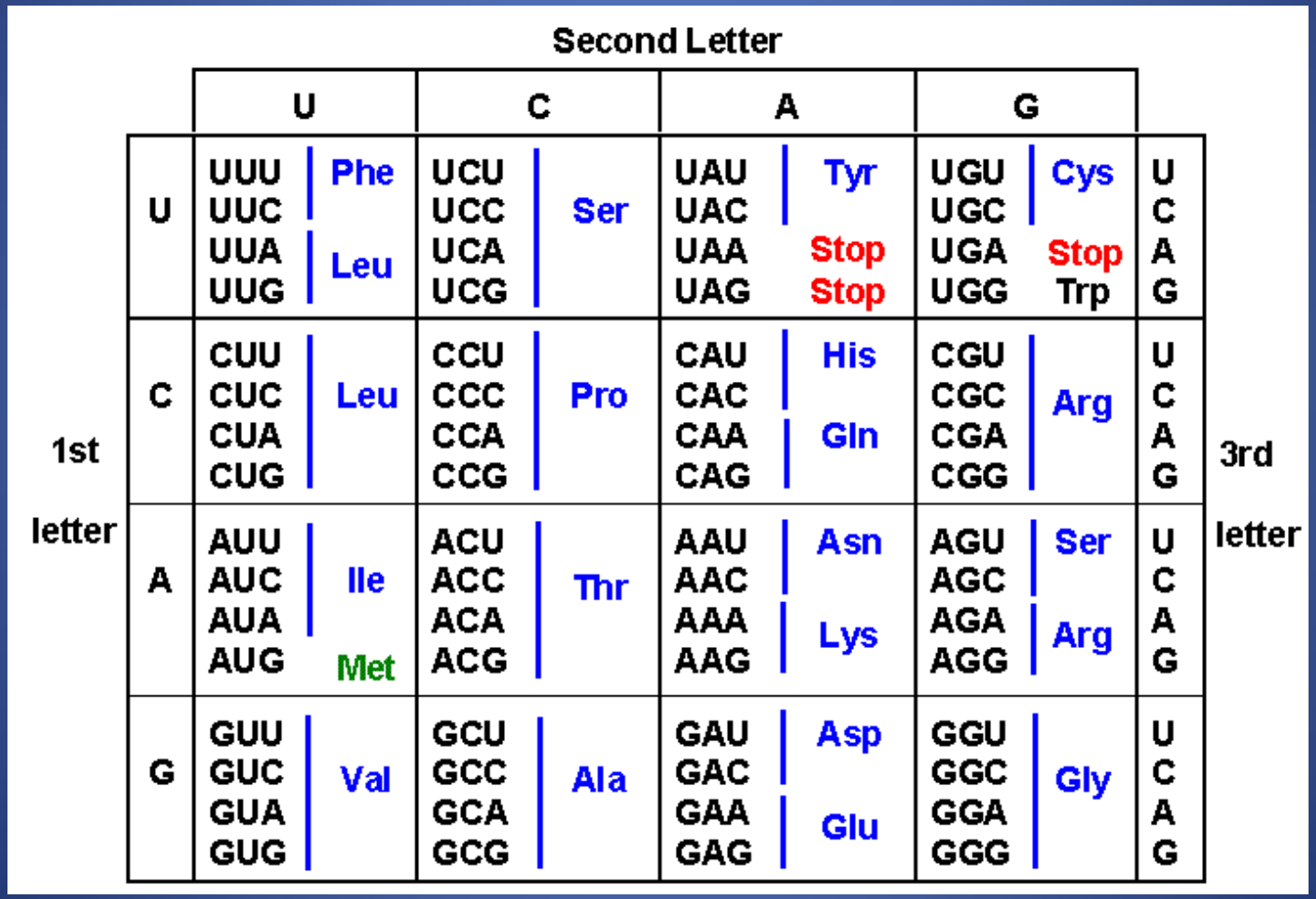
First success

- Once Nirenberg had a working cfs, he realized that he could add synthetic analogs of mRNA to see what polypeptide would be produced
- In 1961 he and Matthaei added poly-rU (prepared using the enzyme polynucleotide phosphorylase) and obtained a large amount of poly-Phe
- The conclusion was that the codon UUU or a multiple of it was the code for the aa Phe
- At the Intl. Biochem. Conf. in Moscow in 1961 Nirenberg was asked to address the plenum and immediately became famous
- But, this alerted others to the possibilities and there was suddenly a lot of competition, from Ochoa in NY, from Watson and Crick and their associates, including Brenner

The coding problem solved

- But the problem was not yet solved, only one codon had been identified and the other possibilities could only be guessed at by using different proportions of the bases in the mixture to synthesize a polymer, e.g. if the amount was 2 U: 1 A the product would be predominantly UUA, while if it was 1U: 2A it would be predominantly UAA. But, not exclusively, there would be all possible mixtures present (e.g. UAU, AUA, etc.)
- In an attempt to see what was the minimum length that would work, Nirenberg synthesized the trimers, the actual triplets, e.g. UUA, UAU, AUU, etc.
- By radioactively labeling these the 20 tRNAs one at a time he was able to show that only one triplet bound together to the ribosome complex at a time and could be separated by filtration (this required a huge amount of work).
- This enabled the specific triplet codons to be identified

The Genetic Code



Conclusions

- The code is a triplet code
- The code is degenerate with a third base "wobble" for many aa's
- The start code is AUG for Met
- There are three stop codes UAG, UAA and UGA
- The code is universal
- From the discovery of the genetic code has come all subsequent applications of DNA sequencing, including forensics, family and ethnic identity, as well as major advances in cancer and all biological research

Nobel Prize

Nirenberg was awarded the Nobel Prize in 1968

This lecture was based on the book: "*The Least Likely Man: Marshall Nirenberg and the Discovery of the Genetic Code*" By Franklin Portugal MIT Press (2015)

This picture was taken on the day that Nirenberg received the telegram from the Karolinska Institute informing him that he had won the 1968 Nobel Prize