3 Rules Govern the Genetic Code

Codons are read in a 5′ to 3′ direction
eg. the coding sequence of the dipeptide
NH_2 - Thr - Arg - COOH could be

(a) 5′- ACGCGA - 3′ ∴ MRNA is translated in a 5′ to 3′ direction
Thr Arg

OR

(b) 3′- GCAAGC - 5′ [Codons written in the same order but opposite to their original orientation]
↓
if this is translated
then the peptide would be

NH_2 - Arg - Thr - COOH

Codons are non overlapping i.e message contains no Gaps

Message is translated in a fixed reading frame which is set by the initiation codon.

∵ codons are non overlapping and consists of 3 consecutive nucleotides
∴ a stretch of nucleotides could be translated in principle in any of 3 reading frames.

eg: 5′ ACGACGACGACGACG .... 3′ could be translated a
(i) series of threonine (5′ACG-3′)
(ii) series of arginine (5′CGA-3′) OR
(iii) series of aspartate (5′GAC-3′)

Genetic Code

(i) **Degenerate** - many aa specified by more than one codon

codons specifying same aa - synonyms.

UUU, UUC } Phenylalanine

UCU, UCC, UCA, UCG } serine

Often A & G are interchangeable } III rd place equivalence
C & U

But degeneracy is not based on equivalence of 1st 2 ~~bases~~ nucleotides.

e.g. Leucine { UUA, UUG

also by CUU, CUC, CUA, CUG

Mutation in 1st position → similar aa if not the same

2nd " → replace 1 aa with a very (transition) Unlike one

3rd " → rarely a diff. aa (transition)

Each transversion do not have

Another consistency

If 1st 1so Nts are G or C each of the 4 Ntds. in 3rd position specifies the same aa

eg: Proline Alanine, arginine, glycine

CCU	GCU	CGU	GGU
CCC	GCC	CGC	GGC
CCA	GCA	CGA	GGA
CCG	GCG	CGG	GGG

If 1st two positions are occupied by A or U the third Nt does make a differe

G:C > A:T ∴

stronger Mismatches in pairing the third codon base are often Tolerated if 1st 1so positions make a G:C bp.

so any of the 4 Nt at the third position specify the same aa.

NOBBLE 1966, Wobble Concept Francis Crick

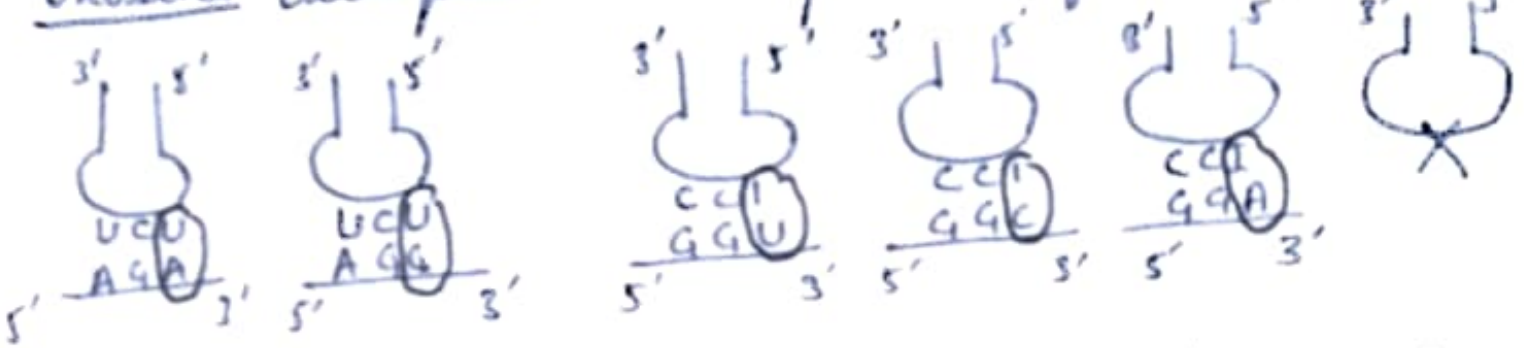
Pairing combinations:

Anticodon	Codon
G	U or C
C	G
A	U
U	A or G
I	A, U, or C

Pairing that gives ribose - ribose distance close to that of standard A:U, G:C

Purine : Purine / Pyr : Pyr have ribose - ribose distance

Inosine occupies the 1st position of the anticodon



Inosine is a nucleoside with a ribose and base hypoxanthine

Three codons can be recognised only when I is present at the I^st position (5') of anticodon

3 tRNAs exist for 6 SERINE codon

UCU UCC UCA UCG AGU AGC

LEUCINE	ARGININE
CUU	CGU
CUC	CGC
CUA	CGA
CUG	CGG
UUA	AGA
UUG	AGG

different tRNAs for sets of codons that differ in the 1st and II^nd position

Wobble is seen at the 1st 5' base of anticodon and not at 1st 5' base of codon b'cos the 1st 5' base of anticodon is least restricted as it is at the end of the stack whereas the 3rd anticodon base is exactly in the centre followed by a bulkier purine (modified residue) a restriction in its movement.

Chain Termination

Ochre UAA
Amber UAG
Opal or UGA
Umber

} NOT read by special tRNA but are read by RF_1 & RF_2 in bacteria
eRF1 in eukaryotes

Release factors enter A-site of ribosome and trigger hydrolysis of peptidyl tRNA occupying the P-site resulting in the release of newly synthesised protein

R Epstein and C Steinberg named the phage mutants after their friend H. Bernstein his last name mean 'amber' in german

UGA → selenocysteine (21st aa)
UAG → pyrrolysine (22nd aa)

non-standard aa

Achaeal prokaryote
Acetohalobium arabaticum can expand its genetic code from 20 to 21 aa (pyrrolysine) under different conditions of growth.

UGA as tryptophan in Mycoplasma

CUG as serine in yeast
rather than leucine

GUG or UUG start codons in bacteria and archaea

Triplet nature of Genetic code

1960s Sydney Brenner argued 2 letter 4^2 provides 1
c

4 letter 4^4 " 256
Whereas 3 " 4^3 " 64 co
↓
sufficient
known 20 amin

st Expt al evidence provided by
Leslie Barnett
Brenner
R. J Watts-Tobin
f Francis Crick

↓ Induced

del (Insertion and Deletion mutations) in rII locus of phage

↓ using acridine dye proflavin

Wild type phage → lyse and plaque formation in E. coli (Both strain B & K12)

→ rII mutants do not infect in strain K12

a single nucleotide insertion(+) causes frameshift mutation due to shifting of reading frame

If these mutants are retreated with proflavin another insertion(+) or deletion(-) will occur

if (-) occurs near to the insertion (+) original reading frame would be restored and this mutant (revertant) may display a wild-type behaviour & infect K12 strain of E. coli

with 2+ followed by 2− occurred in the same sequence

↓

NO infection in K12 strain

i.e correct reading frame not reestablished

(Argument against 2 letter code)

But 3(+) / 3(−) or 3(−)/3(+) reestablished the reading frame

(Supported Triplet nature of code)

Crick and Brenner also suggested that genetic code is non-overlapping.

Crick et al also reasoned to conclude that the code is commaless and Degenerate

Deciphering of the Code

1961, Marchall Nirenberg
J. Heinrich Mathaei

Invitro cell-free (in a test-tube)

$n[rNDP] \xrightarrow{\text{polynucleotide phosphorylase}} [rNMP]_n + n[P_i]$

Ribonucleoside diphosphates → RNA; Inorganic phosphates

Under normal cond^n it favours degradation of RNA nucleoside diphosphates

* However it can be made to catalyze the formation of internucleotide 3′-5′ phospho-diester bonds & thus makes RNA by using high concentration

by 1960

Elucidation of aa sequence was a practical though laborious process.

1960 – mRNA participation in protein syn was established.

Made possible to CRACK the CODE
- artificial mRNA use was possible
- cell free systems for carrying out protein synthesis began.

Polynucleotide phosphorylase breaks down RNA into nucleoside diphosphates.

By use of high nucleoside diphosphate conc. it can catalyse the reverse reaction. i.e fort^n of internucleotide 3'→5' phosphodiester bonds and thus RNA.

No template is required with this enz.

∴ Possible to make RNA of choices

eg: only adenosine diphosphate results in RNA containing Adenylic acid & called Polyadenylic acid or poly-A

Likewise poly U, poly C, poly G

* Nirenberg Matthaei Ochoa, etal

If 2 diff. diphosphates are mixed

Copolymers* poly AU, poly AC, poly CU etc

but with random base sequences, with near neighbour frequencies determined by relative conc.

eg: 2X A that U in Poly AU will have

seq. UAAUAUAAAUAAUAAAAUAUU - - - -

High Mg^{2+} conc was used. It circumvents the need of initiation factors and special initiator fMet-tRNA

Poly U – Polyphenylalanine
Poly C – Proline
Poly A – lysine
* Poly G – formed multistranded triple helices that do not bind to ribosomes.

Poly A C – contained CCC, CCA, CAC, ACC
CAA, ACA, AAC, AAA

The proportion varied with the copolymer A/C ratio.

e.g. more A than C
promotes more asparagine than histidine residues.

∴ conclusion.
asp is coded by 2 As & 1 C
his " " by 2 Cs & 1 A
but order can't be predicted
2C + 1A can be CCA, CAC, ACC.

964 → TRINUCLEOTIDE BINDING TECHNIQUE
ORDER was known by Trinucleotide effect.

specific aacyl-tRNA bind to ribosome-mRNA complexes in absence of all other factors required for protein synthesis.
Binding of a trinucleotide is sufficient.
5'-GUU-3' – promotes valyl-tRNA binding
5'UUG-3' – " leucyl-tRNA
UGU – " cysteinyl-tRNA

BUT some trinucleotides bind less efficiently to Ribosomes ∴ impossible to know what

were being used to prepare synthetic
polyribonucleosides with repeating sequences

Ribosome start protein syn at random points
along these regular copolymers
yet they incorporate specific aa into
polypp. eg: CUCUCUCU
leu - ser alternate.
UGUGUG
Cys - Val
ACACAC
thr - His

Repeat of 3 nucleotide seq. AAG
gives 3 pp: polylysine, polyarginine,
polyglutamic acid

PolyAUC behaves in same way giving
polyisoleucine, polyserine, polyhistidine.

Repeat of tetranucleotide

A stretch of Nts could be translated in any
of three reading frames. eg 5'- ACGACGACGACGA-
as a series of threonine 5'ACG3' } depending upon
Arg 5'CGA-3' } the frame of

Gobind Khorana deciphered the genetic code by using Repeating Copolymers in early 1960s

First, he created the individual short sequences (e.g di, tri- then he replicated them many times tetranucleotide and finally joined them enzymatically to form the long polynucleotides

Khorana reaffirmed the identity of triplets that had already been deciphered and filled in the gaps left from other approaches.

Repeating sequence	Polynucleotide	Repeating triplet
eg: dinucleotide UG	5' UGUGUGUGUGUG 3' ↑ initiation	UGU & GUG
trinucleotide UUG	5' UUGUUGUUGUUG 3'	UUG, UGU, GUU
tetranucleotide UAUC	5' UAUCUAUCUAUC 3'	UAU, AUC, UCU, CUA

Using 2 tetranucleotide sequences GAUA & GUAA Khorana reached a conclusion that atleast two triplets were TERMINATION CODONS because neither of these repeating sequences directed the incorporation of more than a few aas into a polypeptide, too few to detect.

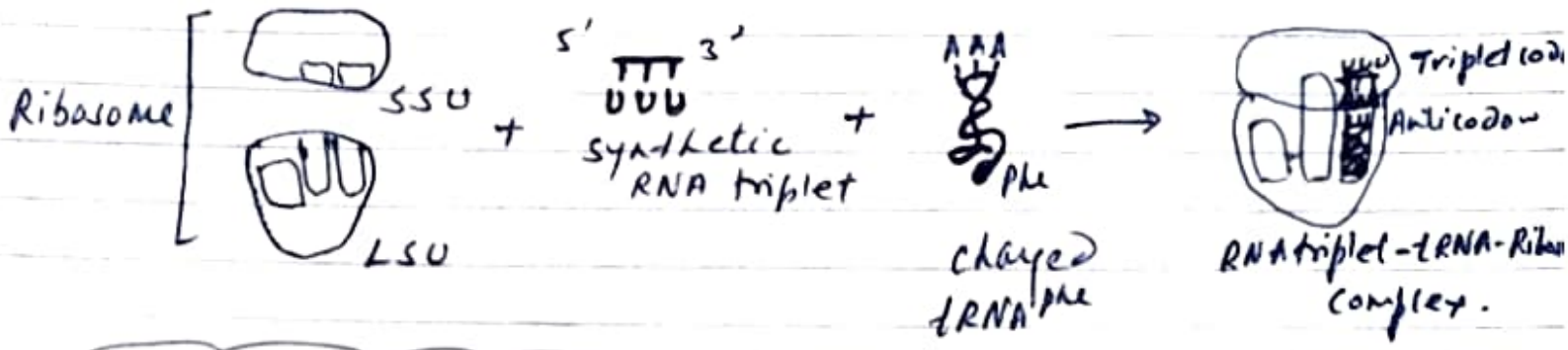
TRIPLET BINDING ASSAY

Developed in 1964, Nirenberg & Philip Leder

leading to specific binding/assignment of triplet codons.

Based on

Ribosomes when present *in vitro* with an RNA sequen as short as 3 Ribonucleotides will bind to it and forms complexes similar to the ones found *in vi*



The amino acid to be tested was made radioactive & combined with its tRNA creating a charged t-RNA.

This tRNA, the RNA triplet & Ribosomes were incubated together on a nitrocellulose filter which retains ribosomes but not other smaller components such as charged tRNA.

If radioactivity is not retained on the filter an incorrect aa has been tested.

If retained, it does so because the charged tRNA has bound to the RNA triplet associated with the ribosome, which itself remains on the filter.

This way specific codon assignment could be made.