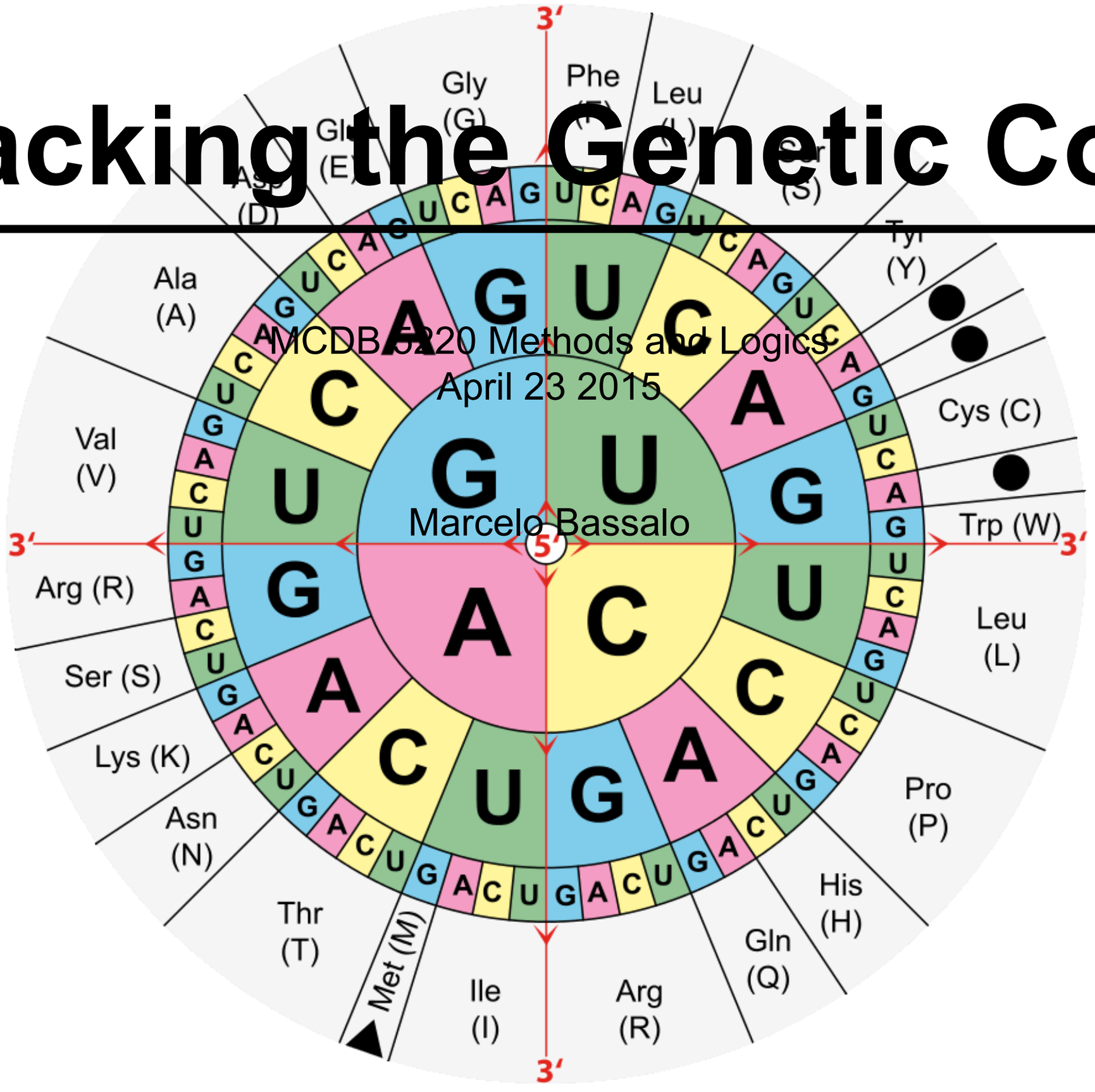
Cracking the Genetic Code

MCDB 5220 Methods and Logics
April 23 2015

Marcelo Bassalo



Last Tuesday…

Nirenberg and Matthaei: RNA is the template for protein synthesis (poly-U —> phenylalanine)

GENERAL NATURE OF THE GENETIC CODE FOR PROTEINS

By Dr. F. H. C. CRICK, F.R.S., LESLIE BARNETT, Dr. S. BRENNER and Dr. R. J. WATTS-TOBIN

Medical Research Council Unit for Molecular Biology, Cavendish Laboratory, Cambridge

Thursday!

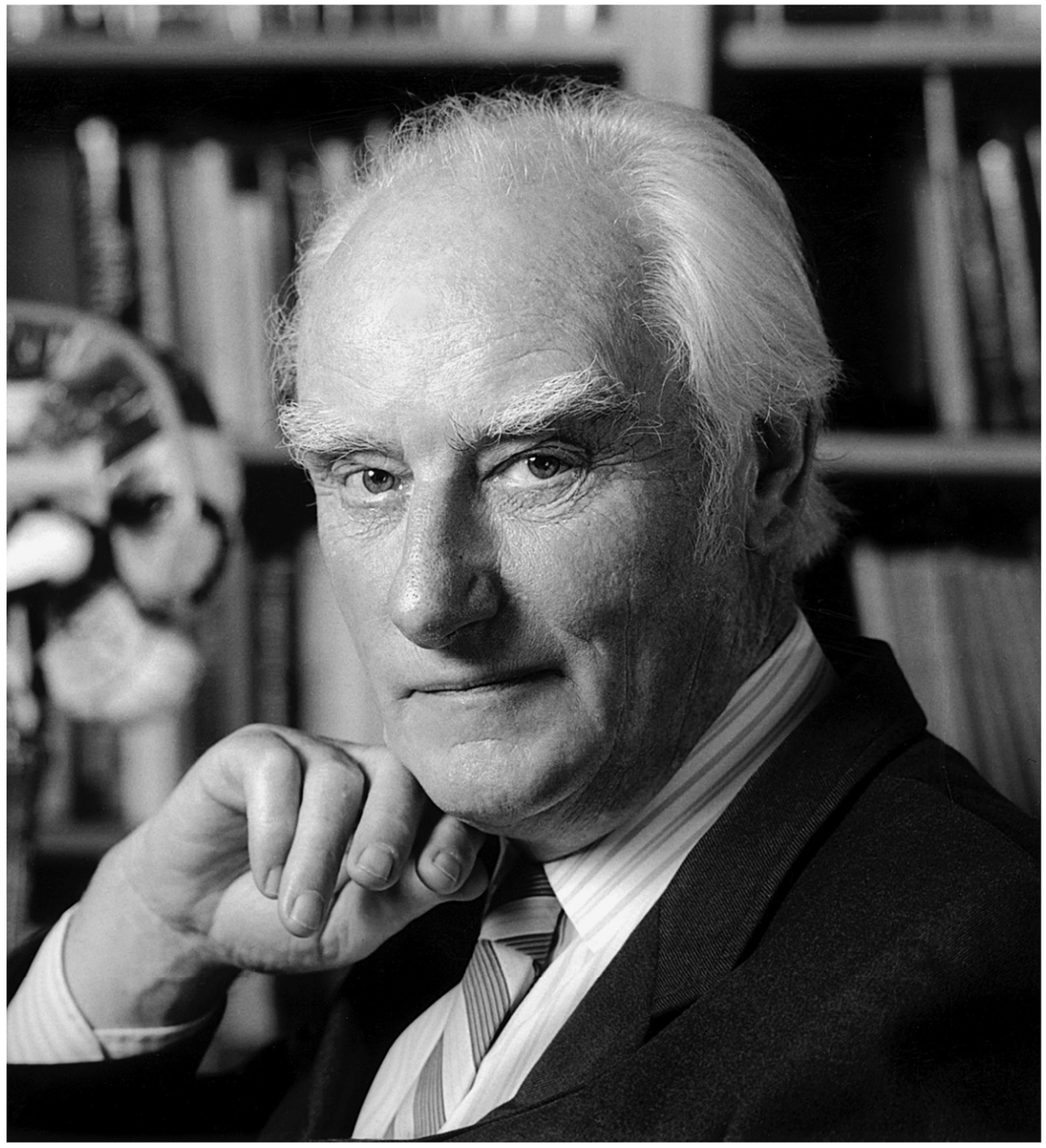
*CHARACTERISTICS AND COMPOSITION OF RNA CODING UNITS**

By J. Heinrich Matthaei,† Oliver W. Jones, Robert G. Martin, and Marshall W. Nirenberg

NATIONAL INSTITUTE OF ARTHRITIS AND METABOLIC DISEASES, BETHESDA

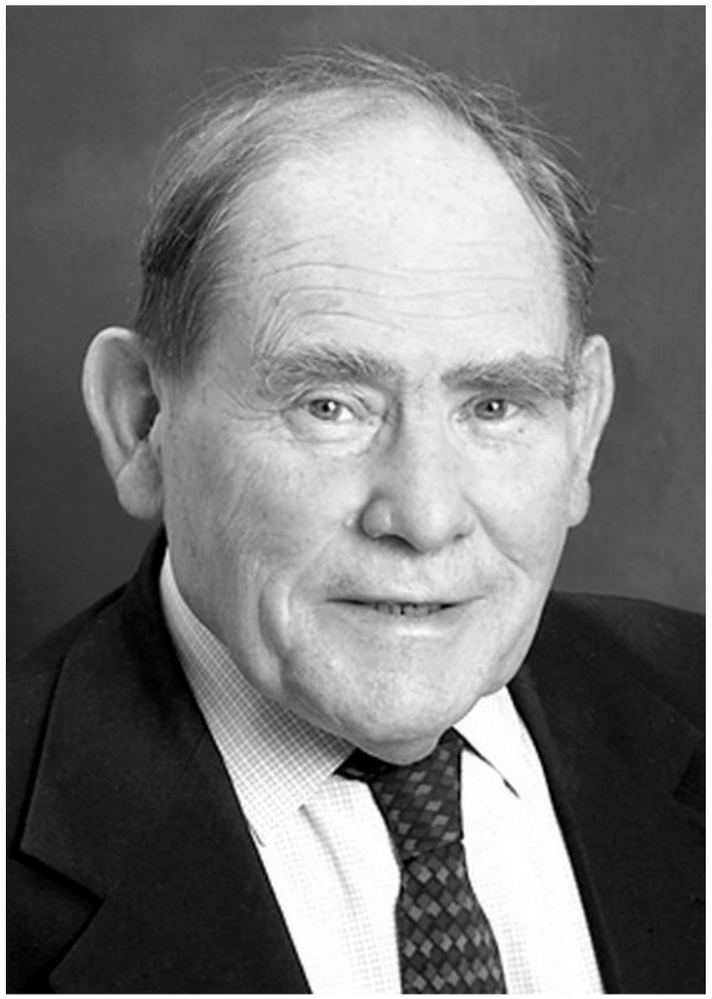
Communicated by Richard Roberts, February 27, 1962

Francis Crick



Sydney Brenner

- 1927 - today
- Born in South Africa
- BS in Anatomy and Physiology
- MS in Cytogenetics
- PhD in Physical Chemistry from Oxford
- Joined Salk Institute in 1976
- Established *C. elegans* as model organism for developmental biology
- 2002 Nobel Prize Physiology or Medicine



Leslie Barnett

- 1920 - 2002
- Born in London
- BS in Dairying
- Worked with Brenner for most part of her life

Leslie Barnett, Francis Crick and Sydney Brenner at the MRC Laboratory, Cambridge in 1986. Original photograph held by Cold Spring Harbor Laboratories. Library reference no. SB/8/3/23.

Cracking the Genetic Code

No. 4809 December 30, 1961 NATURE 1227

GENERAL NATURE OF THE GENETIC CODE FOR PROTEINS

By Dr. F. H. C. CRICK, F.R.S., LESLIE BARNETT, Dr. S. BRENNER
and Dr. R. J. WATTS-TOBIN

Medical Research Council Unit for Molecular Biology,
Cavendish Laboratory, Cambridge

Bacteriophage T4

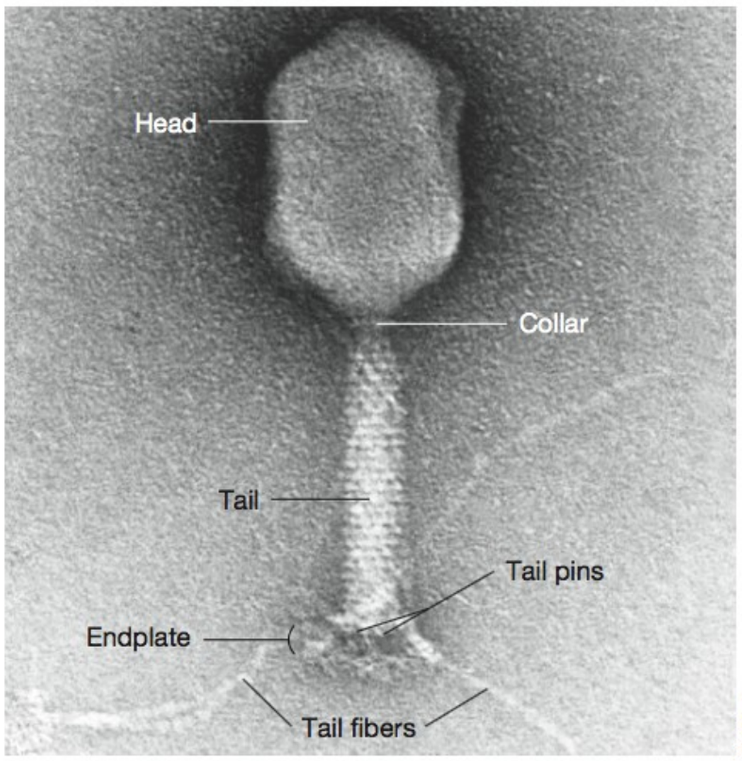
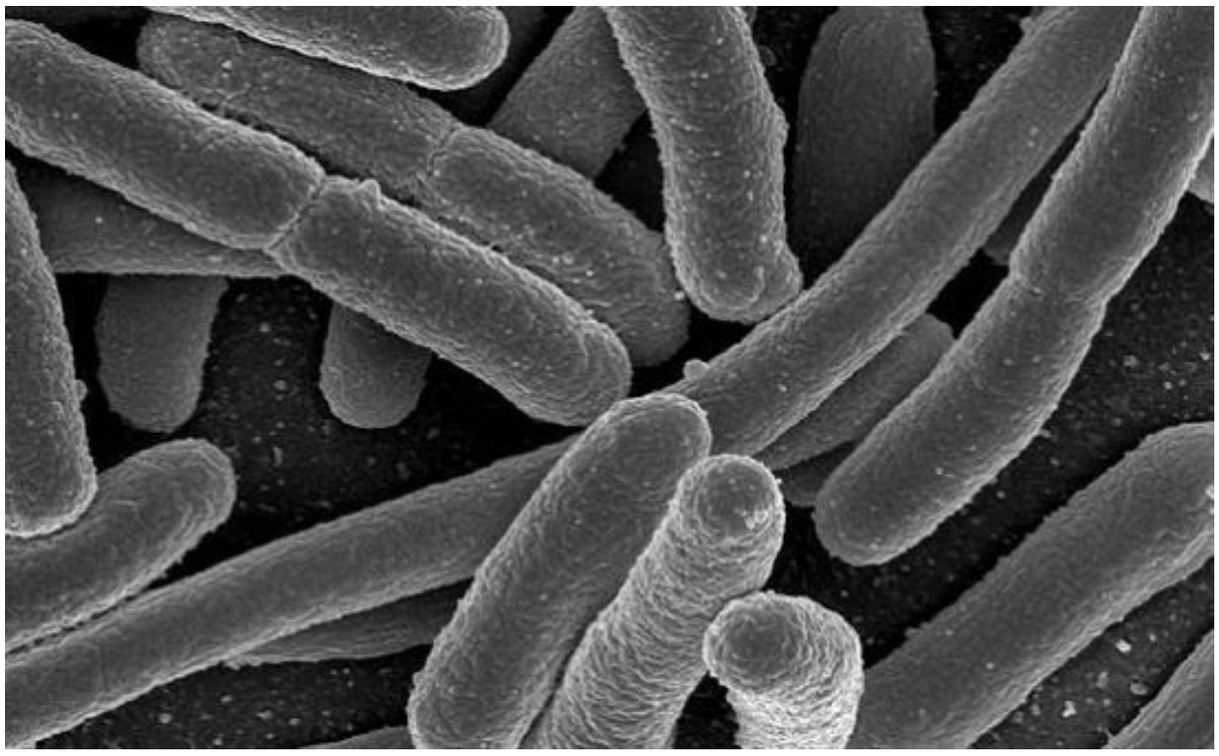
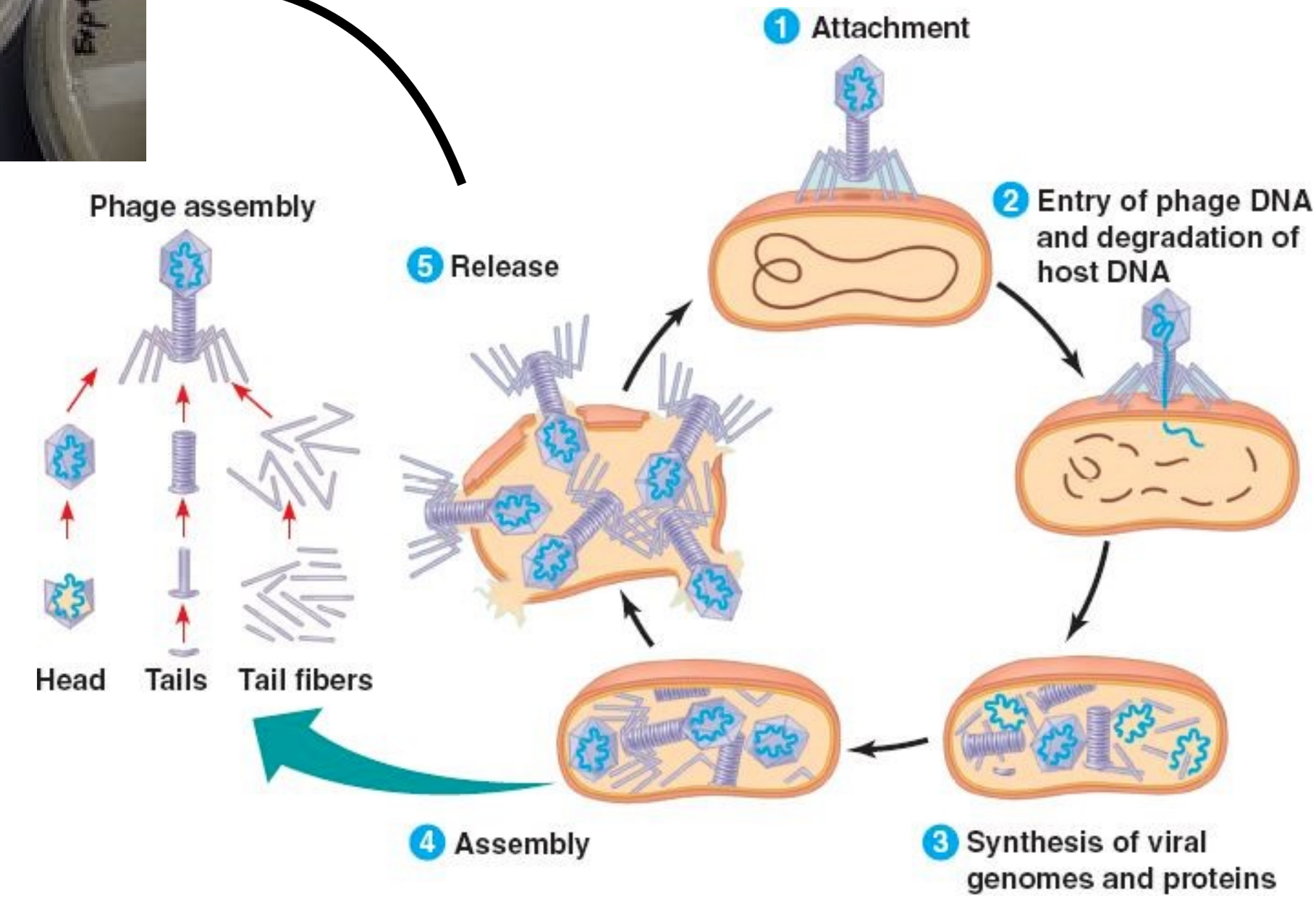


Figure from Brock Biology of Microorganisms

Escherichia coli



Cracking the Genetic Code

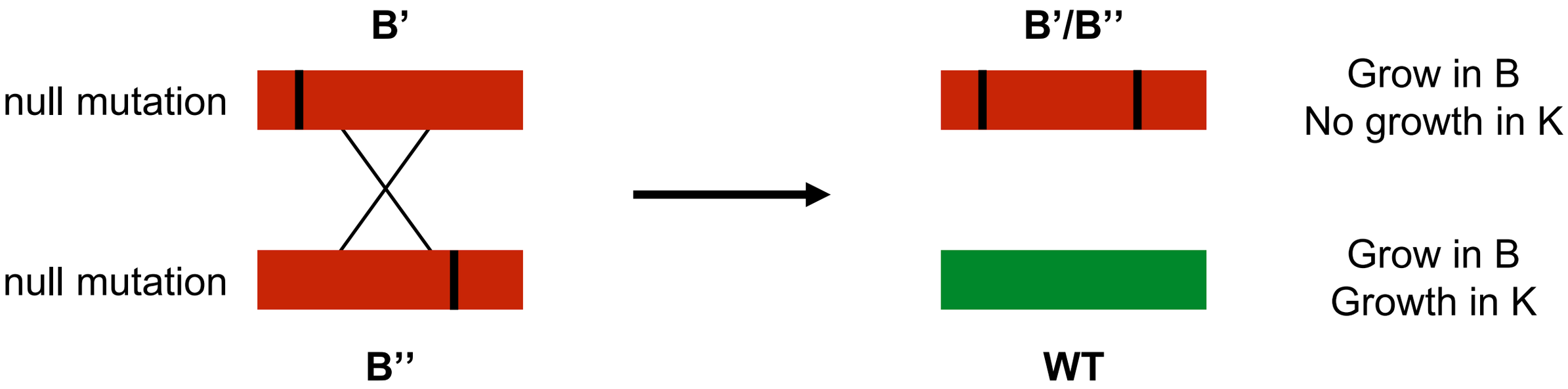


Why Phage T4?

- Plaques are an easy screening system
- Allows investigation of rare events (trillions of tries in a single LB plate)
- **rII locus**: phenotypes allows genetic mapping (Benzer)

	rII genes	***E. coli* K-12 (K)**	***E. coli* B (B)**
WT		✓	✓
non-leaky mutation (null)		X	larger/irregular plaques (r-plaque)
leaky mutation (partial function)		✓	larger/irregular plaques (r-plaque)

Why Phage T4?



Grow in B
No growth in K

$$\text{Distance (B' to B'')} = \frac{\text{\# plaques (K)}}{\text{\# plaques (B)}}$$

The Genetic Code is not overlapping

Evidence comes from previous studies:

- Tobacco mosaic virus RNA: mutations in RNA change only 1 amino acid (Tsugita et. al)
- Abnormal human hemoglobins shows only single amino acid changes (Watson et. al)

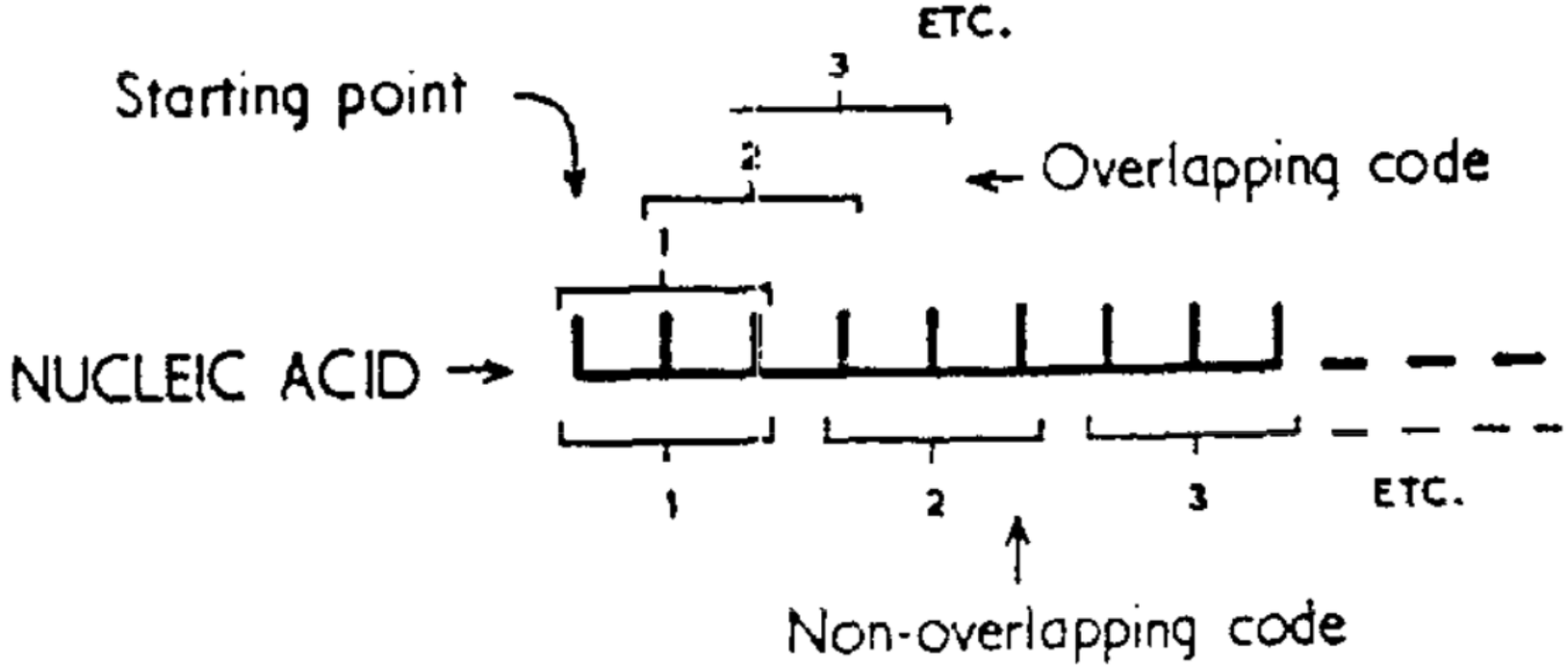


Figure 1

The Genetic Code is not overlapping

- How to find the code (reading frame) in a non-overlapping arrangement?

1) The comma hypothesis

CODE,CODE,CODE,CODE,CODE,CODE,CODE

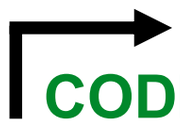
2) The comma-free hypothesis

CODE CODE CODE CODE CODE CODE CODE

ODEC ODEC ODEC ODEC ODEC ODEC ODEC

DECO DECO DECO DECO DECO DECO DECO

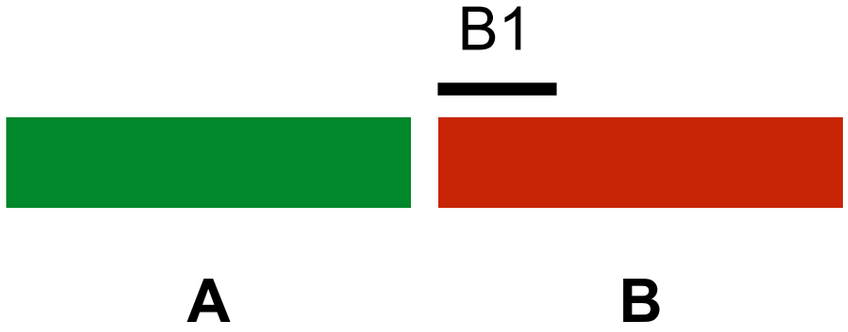
3) The fixed start hypothesis



CODE CODE CODE CODE CODE CODE CODE

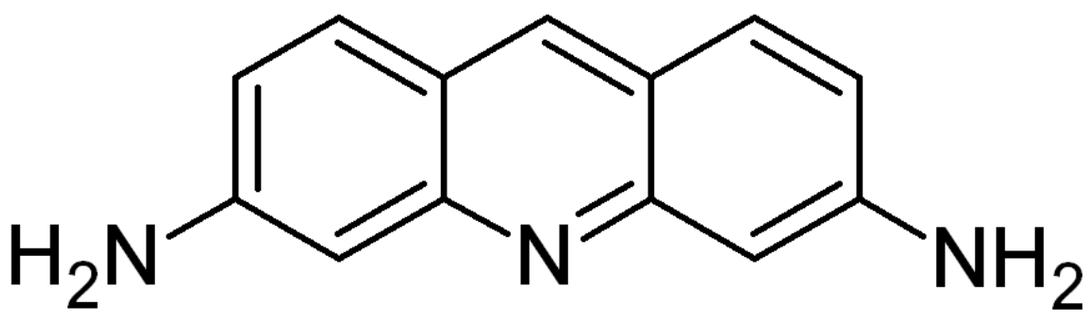
Experimental System

- FC0 mutant in the in the B1 segment of the “B cistron”



- Non-leaky (null) mutant
- No growth in K
- Growth (r-plaques) in B

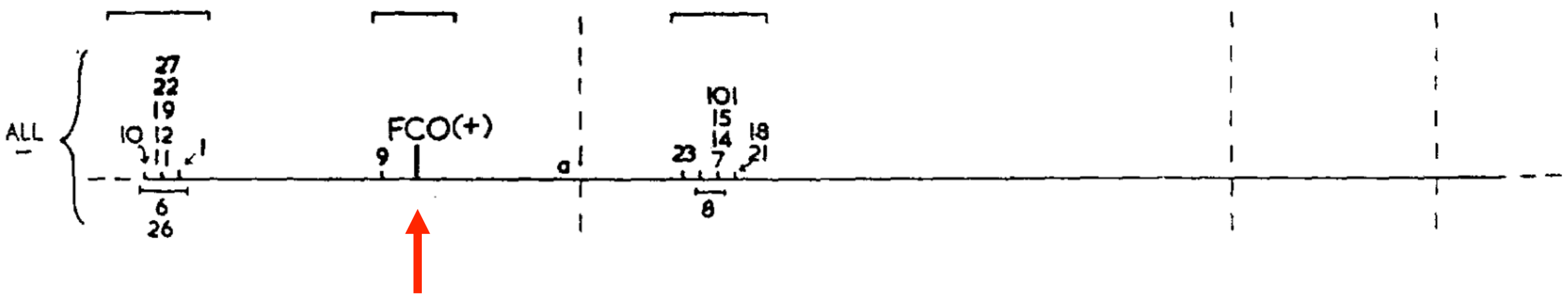
- FC0 mutant was produced by proflavin treatment



- Adds or deletes a base
- Generates mostly non-leaky mutants

Suppressors of FC0

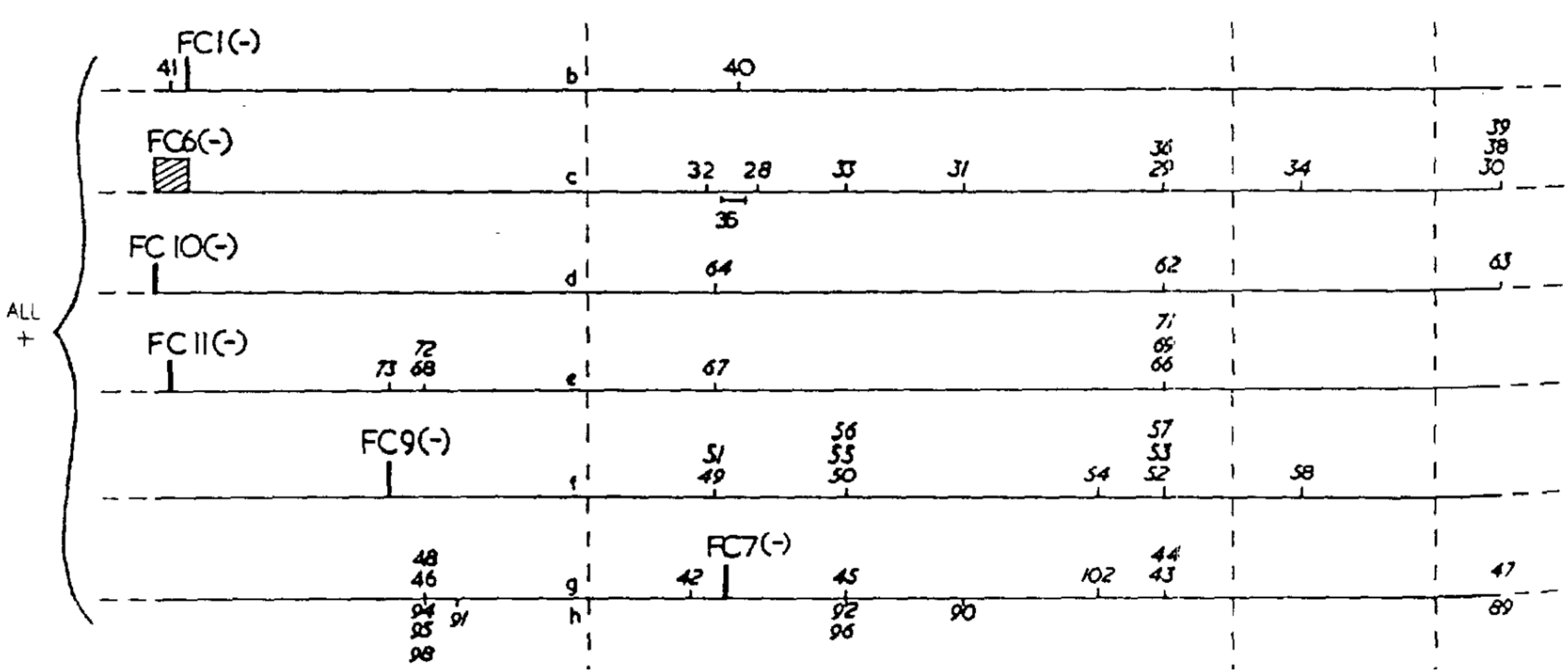
- Second mutation (supressor) restores wild-type phenotype
- Suppressor mutations by themselves are “non-leaky r” (null) mutants



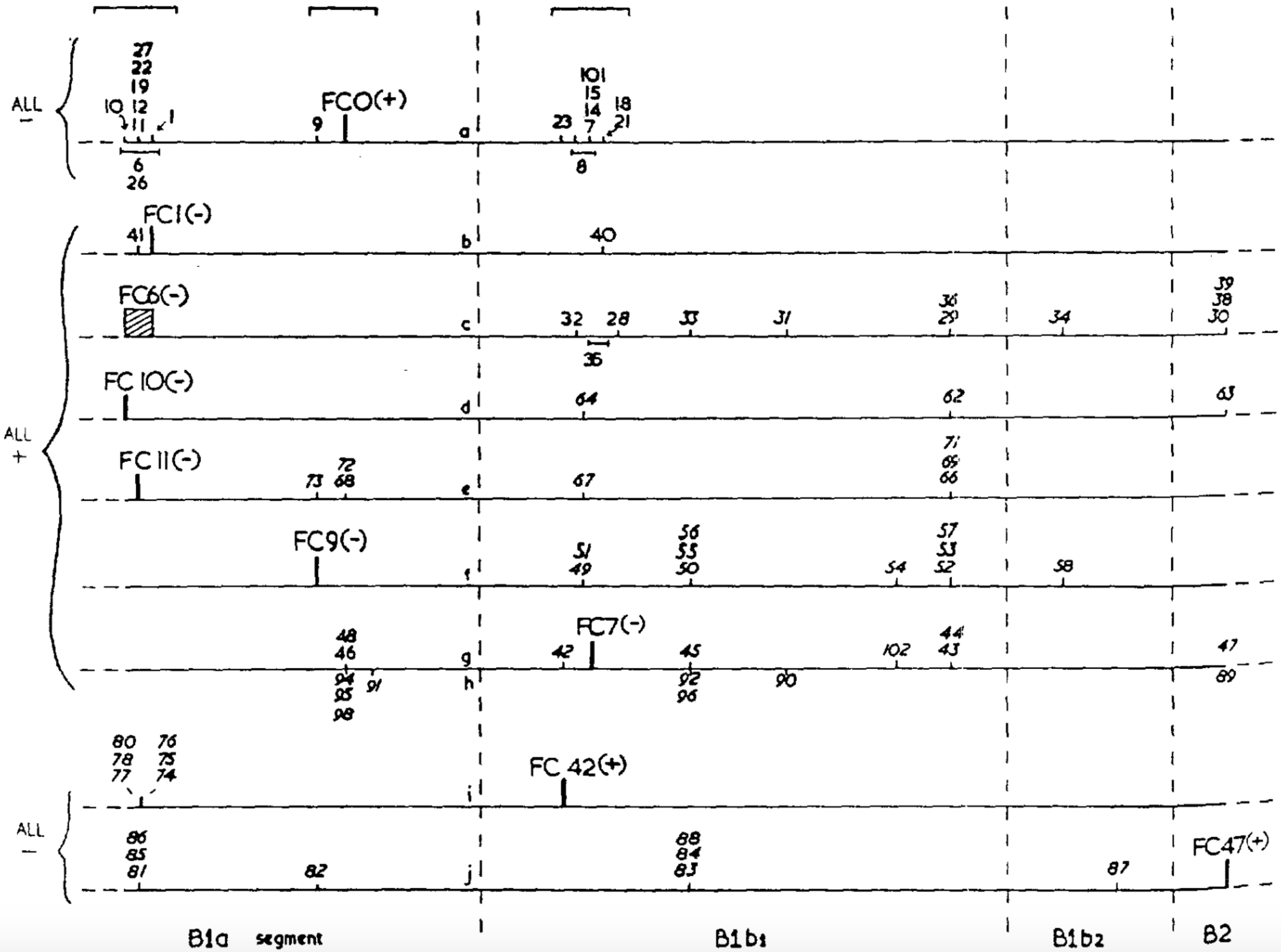
		Mutation	Phenotype
WT	A B C A B C A B C A B C A B C A B C A B C A B C	None	WT
FC0	A B C A A B C A B C A B C A B C A B C A B C	+	null
Supressor of FC0	A B C A A B C A B C A B C A B C A B C A B C C	+ and -	WT

Suppressors of suppressors

		Mutation	Phenotype
WT	A B C A B C A B C A B C A B C A B C A B C A B C	None	WT
FC0	A B C A A B C A B C A B C A B C A B C A B C A B	+	null
Supressor of FC0	A B C A A B C A B C A B A B C A B C A B C A B C	+ and -	WT
Supressor of Supressor	A B C A B C A B C A B A B C A B C A B C A B C	- and +	WT



Suppressors of suppressors of suppressors

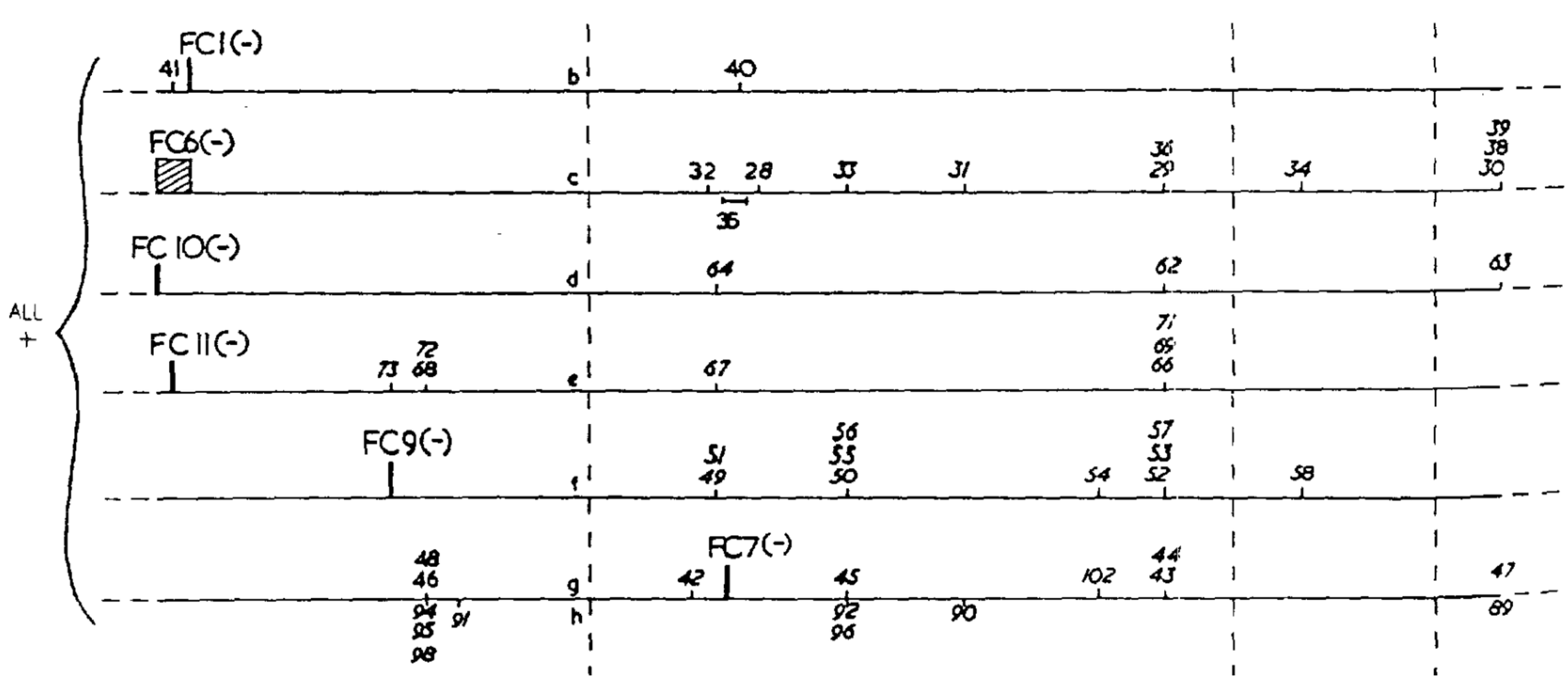


rII mutations

Collection of ~ 80 mutants

- All mutants are non-leaky r (null)
- All mutants (except FC0) were occurred spontaneously

- Non-spontaneous (proflavin) mutations are similar to spontaneous mutations (line **h**)

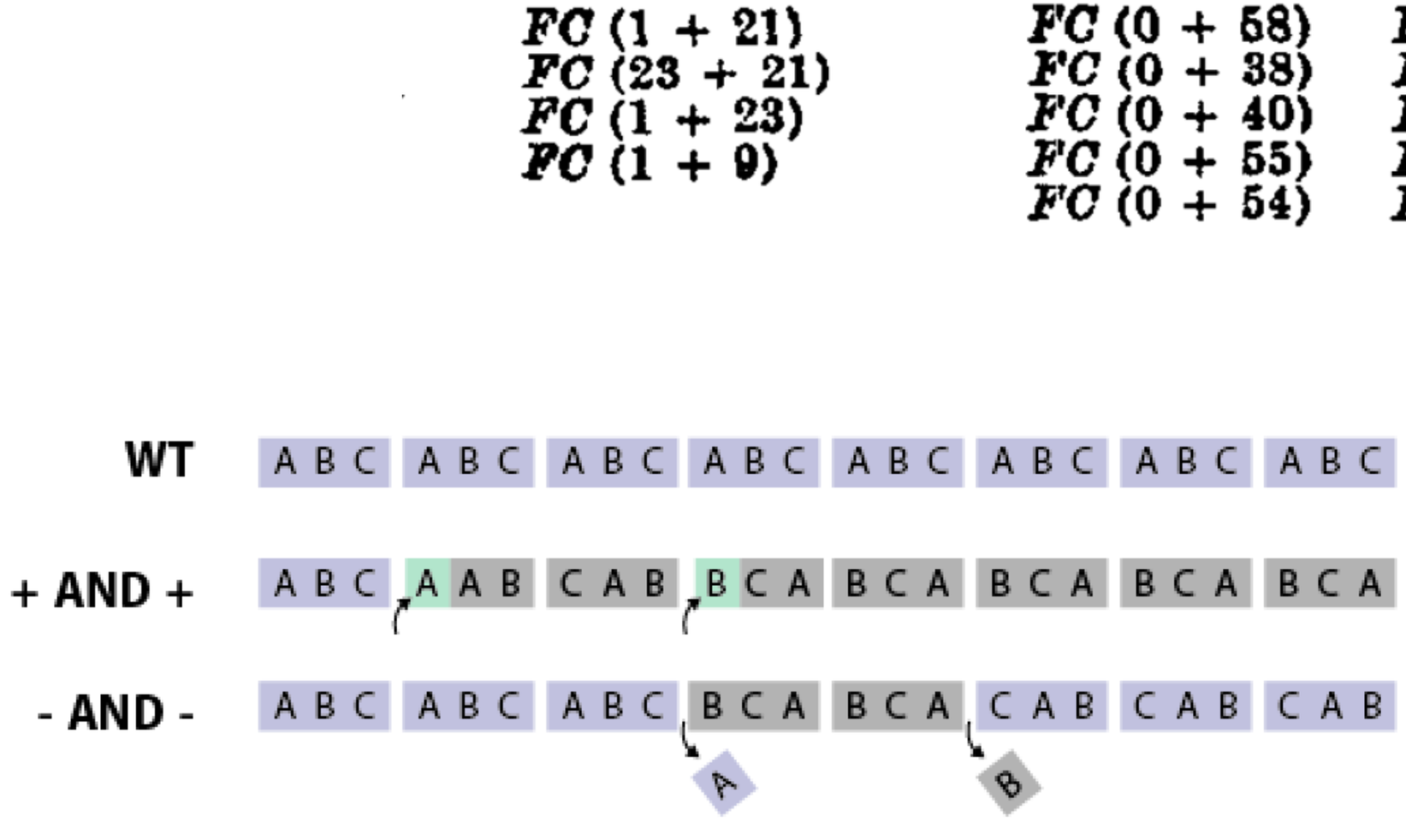


Generating double mutants

- **+** and **+**: r phenotype
- **-** and **-**: r phenotype

Table 1. DOUBLE MUTANTS HAVING THE r PHENOTYPE

− With −	+ With −	+ With +
FC (1 + 21)	*FC* (0 + 58)	*FC* (40 + 57)
FC (23 + 21)	*FC* (0 + 38)	*FC* (40 + 58)
FC (1 + 23)	*FC* (0 + 40)	*FC* (40 + 55)
FC (1 + 9)	*FC* (0 + 55)	*FC* (40 + 54)
	FC (0 + 54)	*FC* (40 + 38)



Generating double mutants

- **+** and **-**: Will all combinations be WT?

"Unacceptable codons might exits":
- nonsense
- end-of-chain
- complications in protein structure

"Convention" on frameshifts:

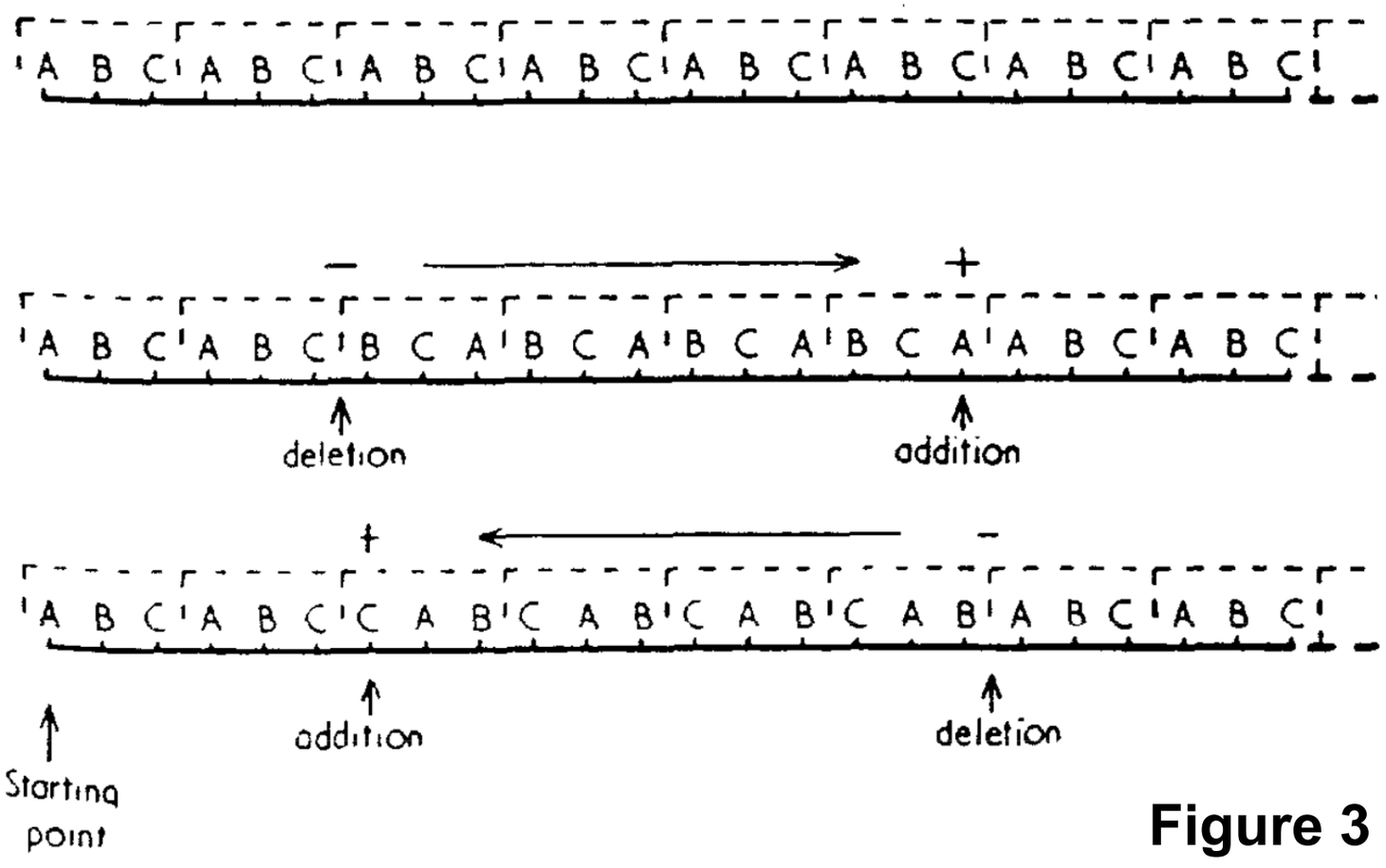


Figure 3

Generating double mutants

Forward shifts are more acceptable is this region
Reverse shifts may generate “unacceptable” triplets (stop codon) — “Unacceptable region”

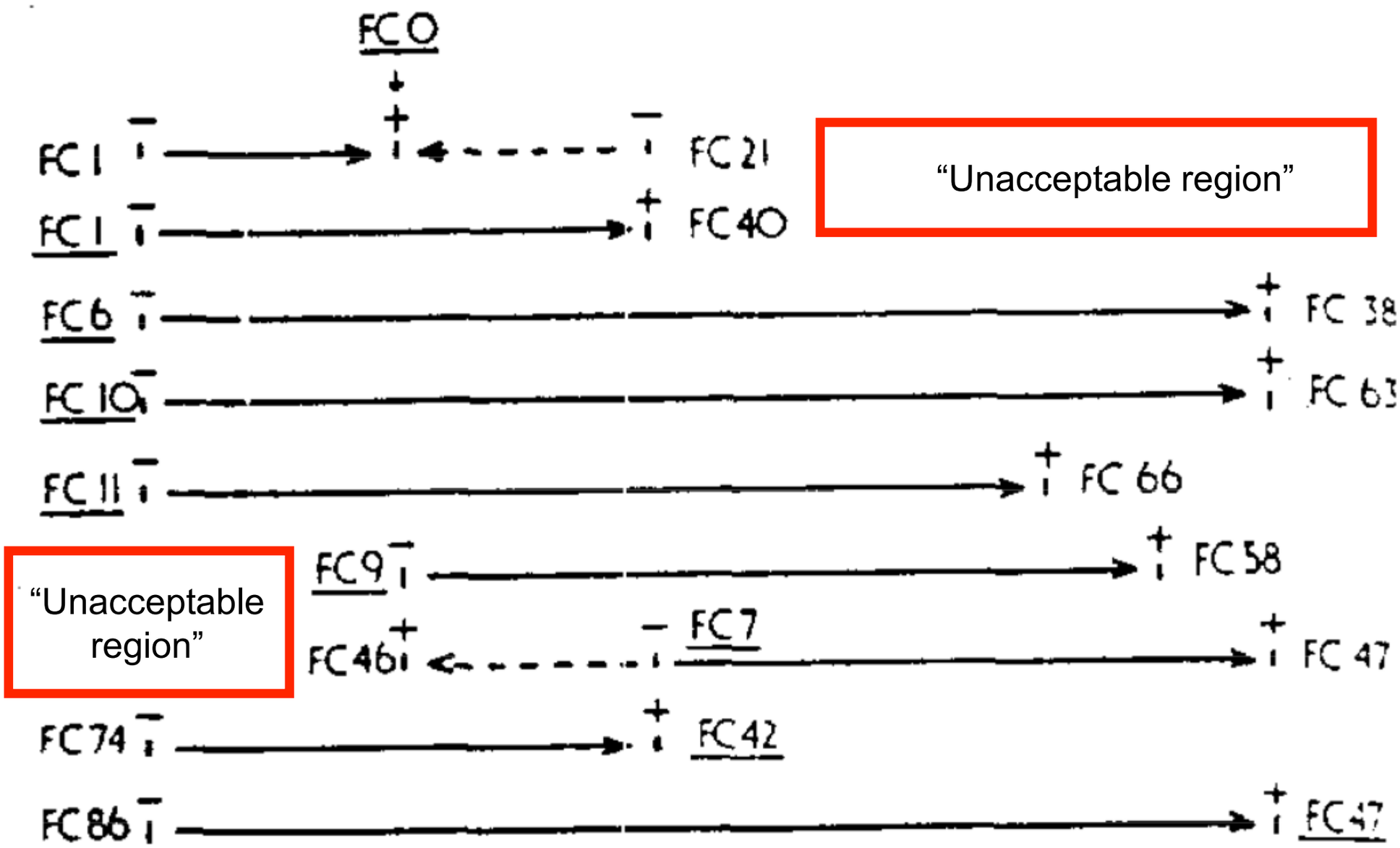


Figure 4

Generating double mutants

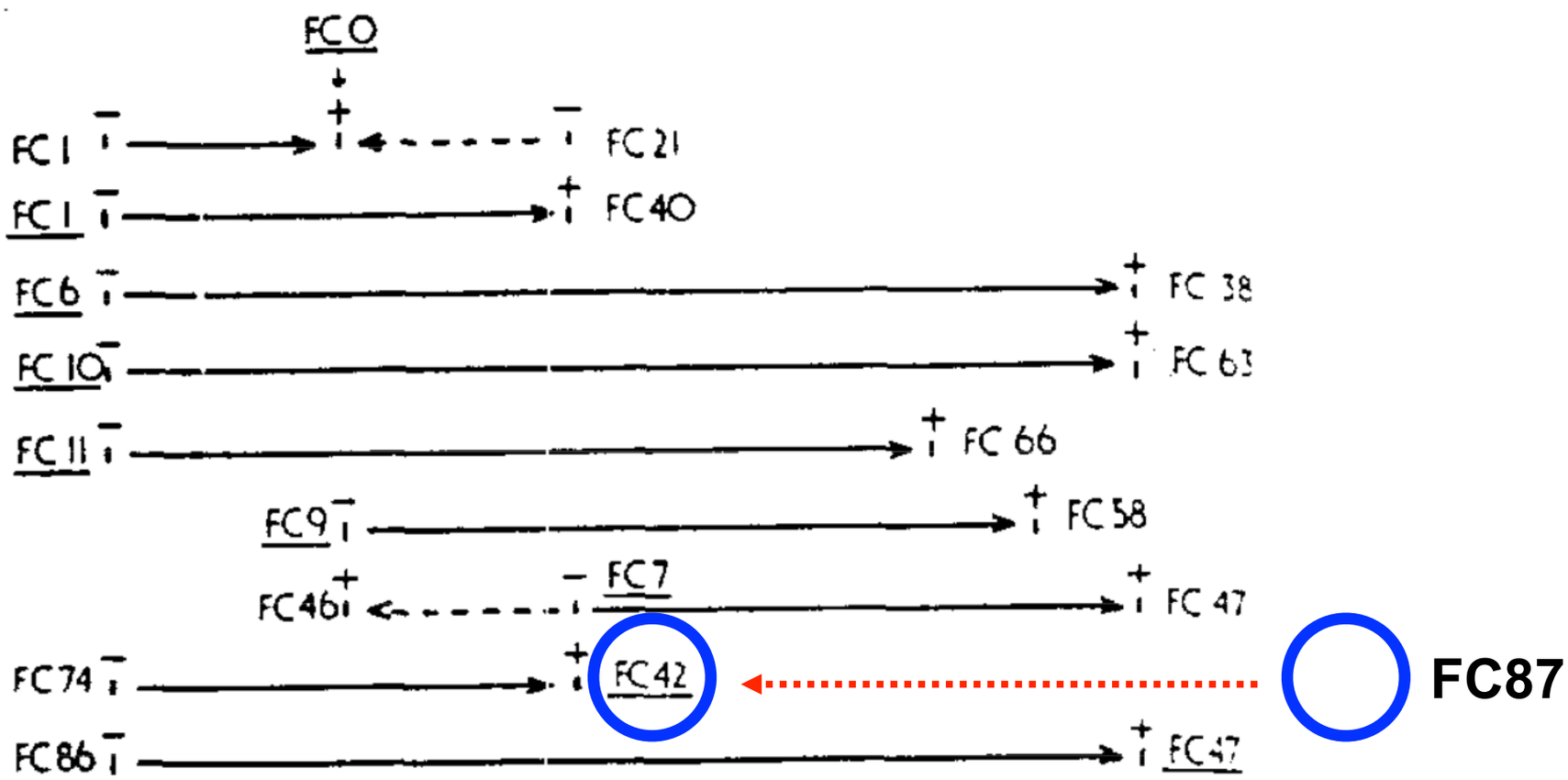


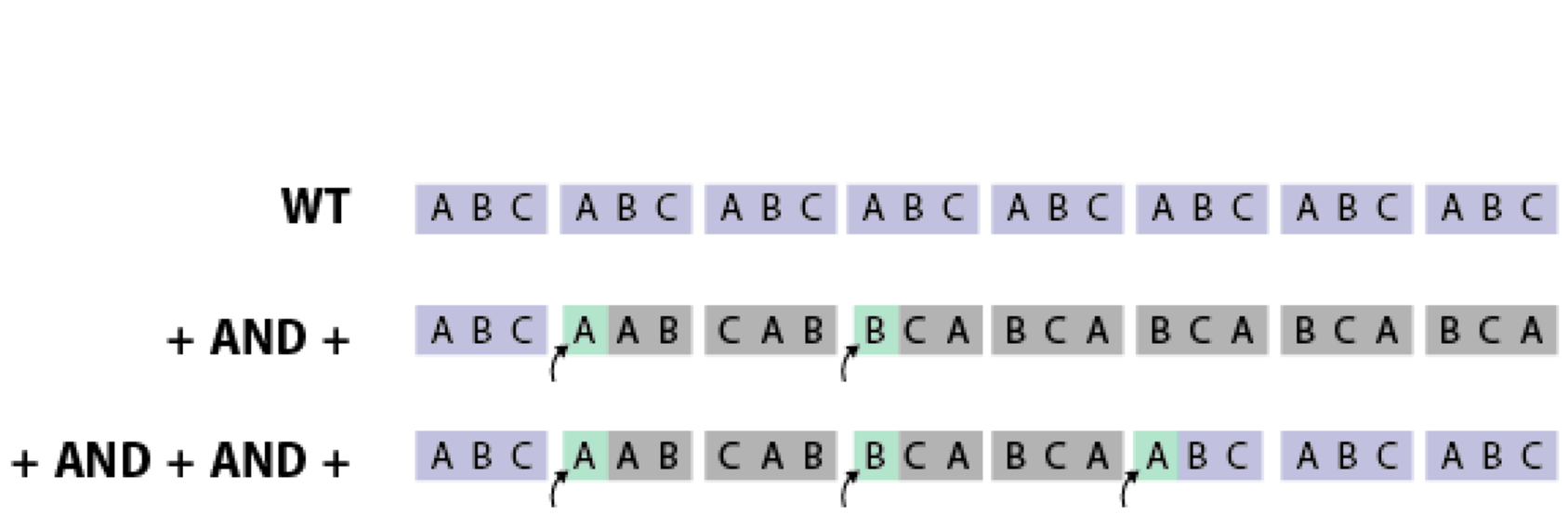
Table 2. Double Mutants of the Type (+ with −)

− \ +	*FC* 41	*FC* 0	*FC* 40	*FC* 42	*FC* 58*	*FC* 63	*FC* 38
FC 1	*W*	*W*	*W*		W		W
FC 86		W	W	W	W	W	
FC 9	*r*	*W*	W	W	W		W
FC 82	*r*		W	W	W	W	
FC 21	*r*	*W*			W		W
FC 88	*r*	*r*			W	W	
FC 87	*r*	*r*	*r*	*r*			W

Generating triple mutants

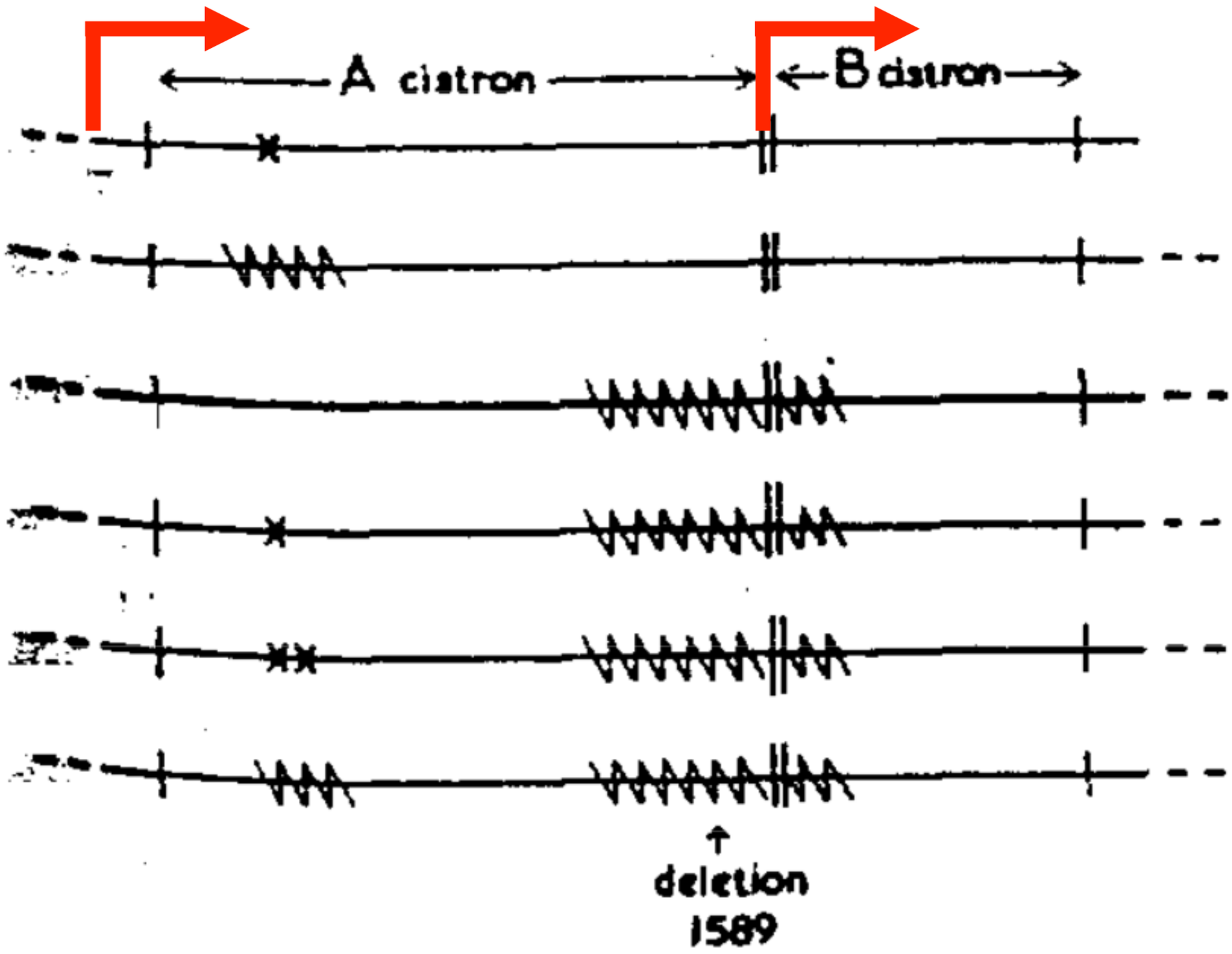
Table 3. TRIPLE MUTANTS HAVING A WILD OR PSEUDO-WILD PHENOTYPE

FC (0 + 40 + 38)	
FC (0 + 40 + 58)	
FC (0 + 40 + 57)	
FC (0 + 40 + 54)	
FC (0 + 40 + 55)	
FC (1 + 21 + 23)	



Genetic code is a triplet (or less likely a multiple of 3)

The start site hypothesis



Function of B cistron

✓

✓

✓

X

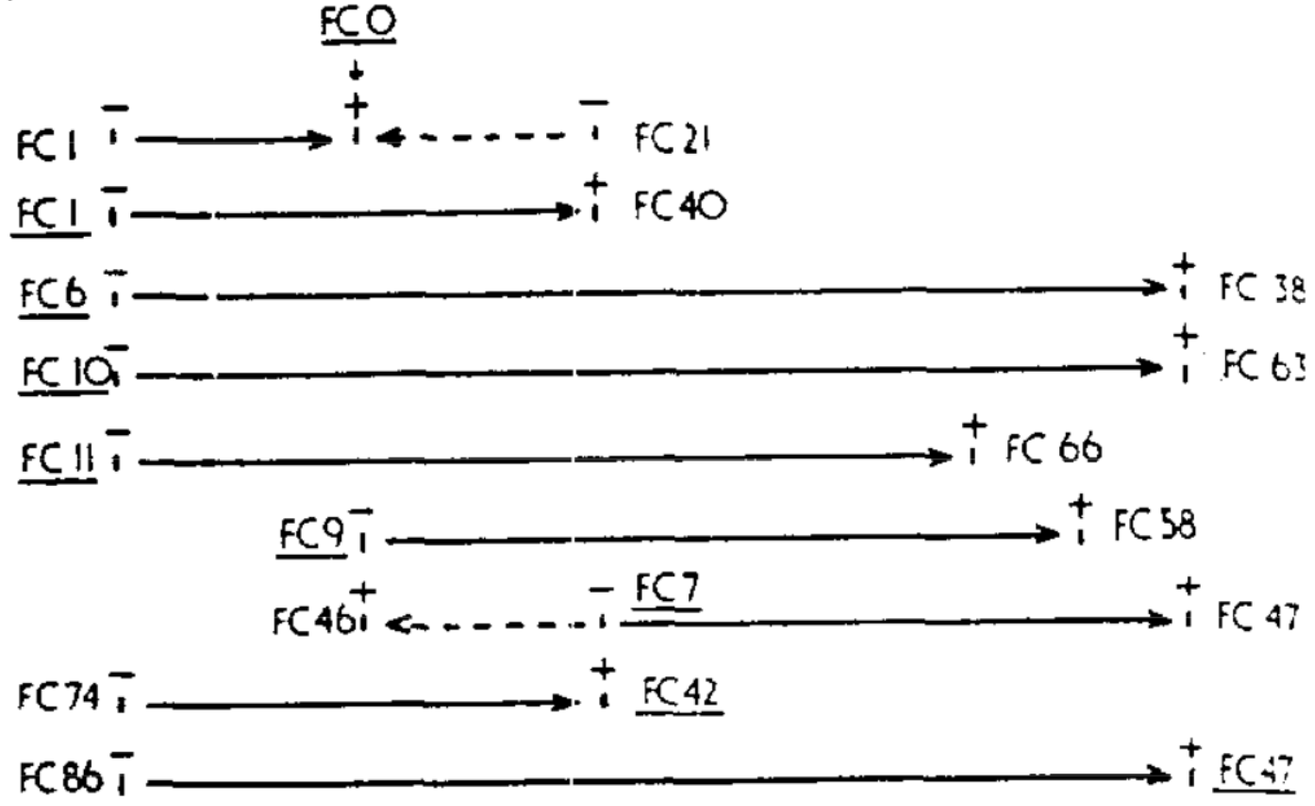
✓

✓ or **X**

Is the code degenerate?

- 4 common base pairs organized in triplets = 4^3 combinations = 64 triplets
- 20 common amino acids
- If the genetic code is not degenerate: there is 44 nonsense amino acids

Unlikely



Is the code degenerate?

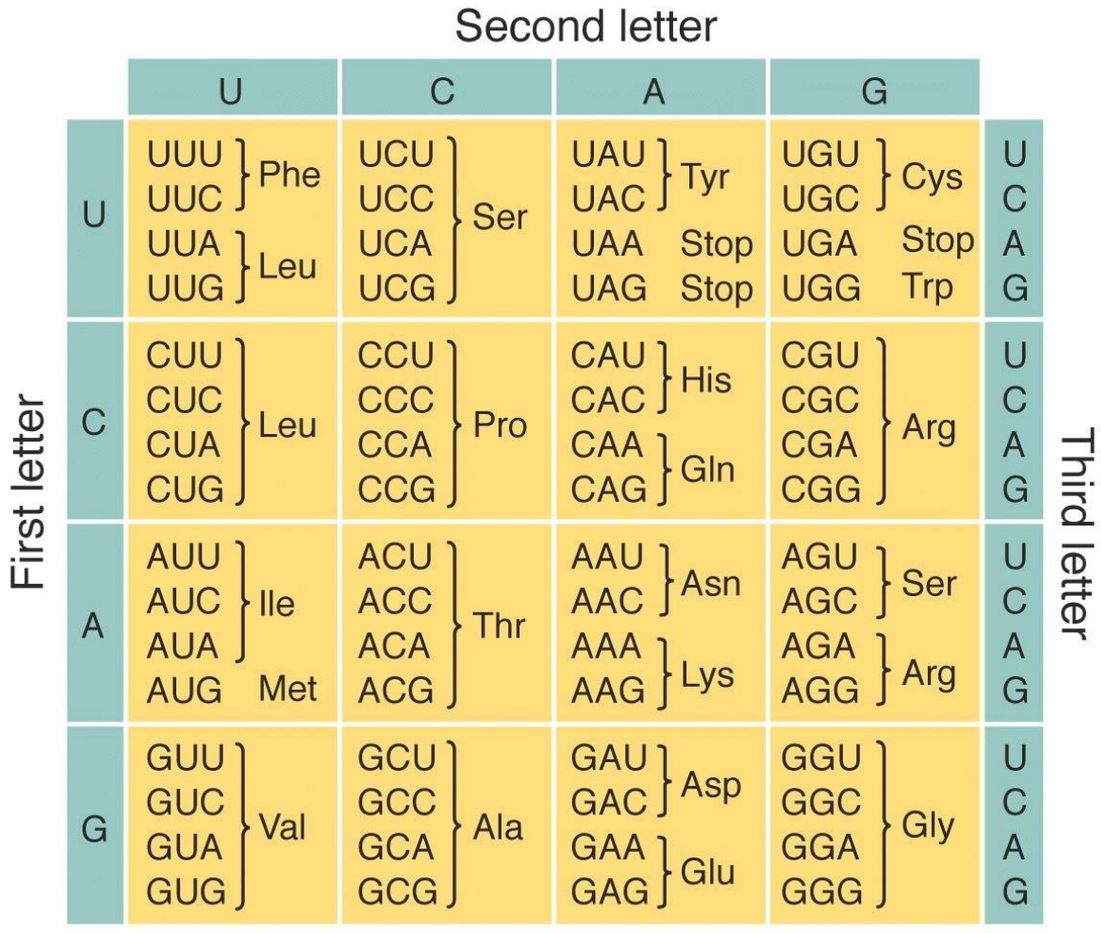
- 4 common base pairs organized in triplets = 4^3 combinations = 64 triplets
- 20 common amino acids
- If the genetic code is not degenerate: there is 44 nonsense amino acids
- The genetic code must be degenerate
- “*how many triplets code amino-acids and how many have other functions we are unable to say*”

Summary

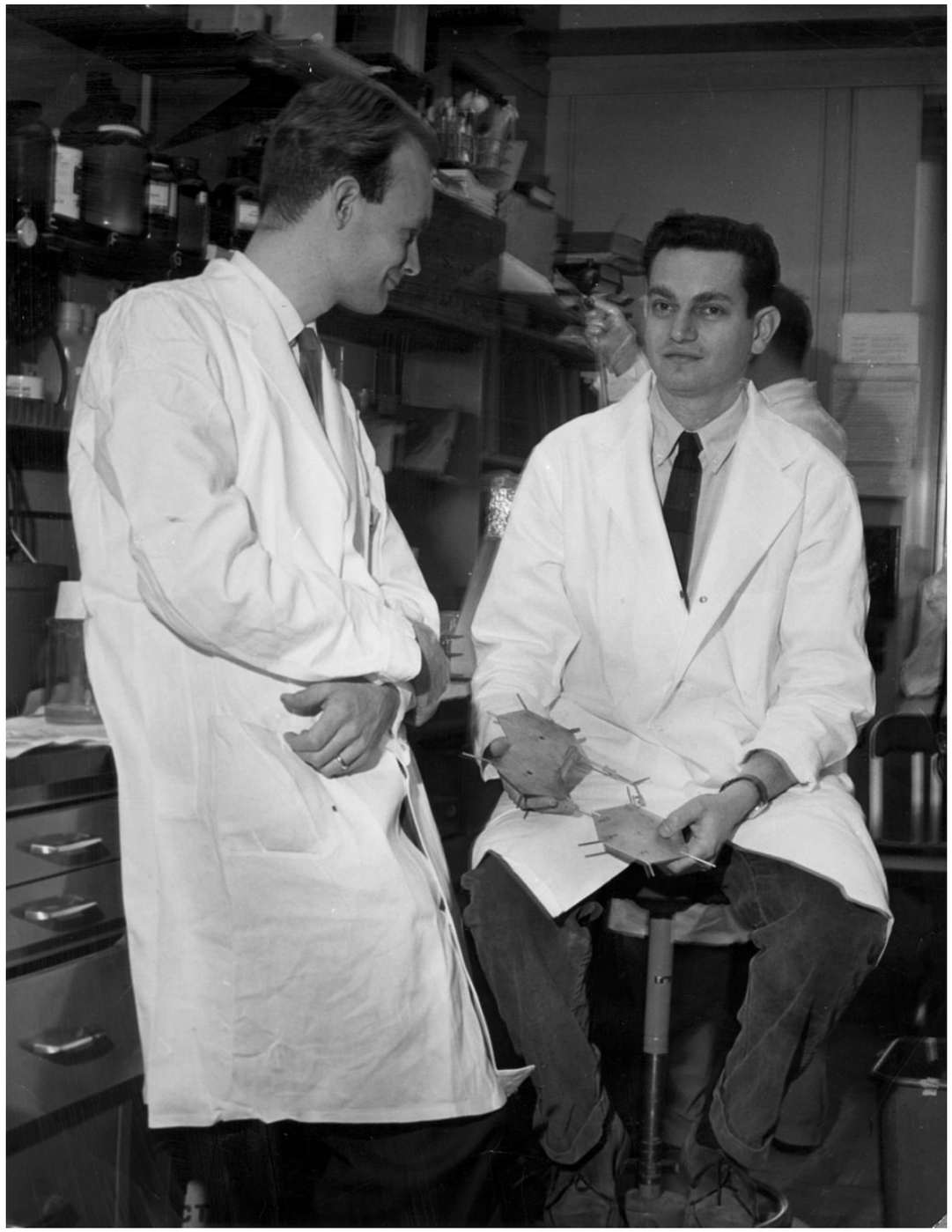
- The genetic code is of the following general type:

- A group of three bases (or, less likely, a multiple of three bases), codes one amino acid
- The code is non-overlapping
- Sequence of bases is read from a fixed point
- The code is probably degenerate

“the coding problem is wide open for experimental attack, and in fact many laboratories, including our own, are already working on the problem”



Cracking the Genetic Code



Marshall Nirenberg
J. Heinrich Matthaei

Cracking the Genetic Code

Har Gobind Khorana

University of Wisconsin, Madison

Mastered synthesis of RNA

*CHARACTERISTICS AND COMPOSITION OF RNA CODING UNITS**

BY J. HEINRICH MATTHAEI,† OLIVER W. JONES, ROBERT G. MARTIN, AND MARSHALL W. NIRENBERG

NATIONAL INSTITUTE OF ARTHRITIS AND METABOLIC DISEASES, BETHESDA

Communicated by Richard Roberts, February 27, 1962

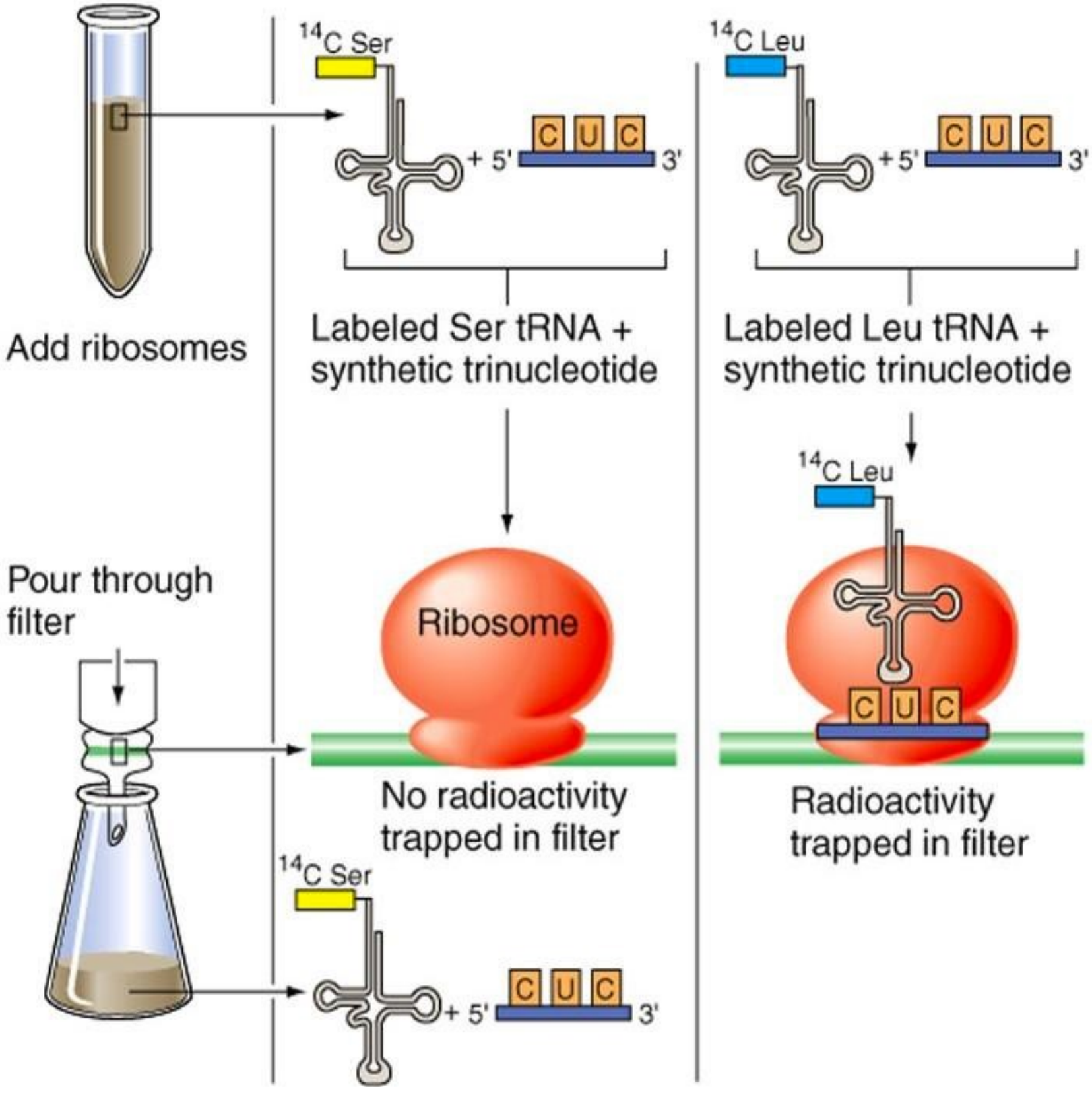
Cracking the Genetic Code

TABLE 3

AMINO ACID INCORPORATION INTO PROTEIN STIMULATED BY RANDOMLY MIXED POLYNUCLEOTIDES

Polynucleotide	UA	UC	UG	UAC	UGC	UGA	Composition of coding units*
Base-ratio	U = 0.87 A = 0.13	U = 0.39 C = 0.61	U = 0.76 G = 0.24	U = 0.834 A = 0.050 C = 0.116	U = 0.341 G = 0.152 C = 0.502	U = 0.675 G = 0.291 A = 0.034	
Probability of triplet relative to phenylalanine (UUU) = 100%	UUU—100 UUA— 13 UAA— 2.2 AAA— 0.3	UUU—100 UUC—157 UCC—244 CCC—382	UUU—100 UUG— 32 UGG— 10.6 GGG— 3.4	UUU—100 UUA— 6.0 UAA— 0.4 AAA— 0.02 UUC— 13.9 UCC— 1.9 CCC— 0.3 UAC— 0.8 AAC— 0.05 ACC— 0.12	UUU—100 UUG— 46.2 UGG— 21.0 GGG— 1.0 UUC—147 UCC—218 CCC—322 UGC— 68.1 GGC— 31.7 GCC—101	UUU—100 UUG— 43 UGG— 19 GGG— 8.1 UUA— 5.1 UAA— 0.26 AAA— 0.01 UGA— 2.2 GGA— 0.1 GAA— 0.01	
Amino Acid							
Phenylalanine	100	100	100	100	100	100	UUU...
Arginine	0	0	1.1	0	49.3	2.9	UCG...
Alanine	1.9	0	0	1.0	40.4	0.9	UCG...
Serine	0.4	160	3.2	3.6	170	2.3	UUC.. + UCG...
Proline	0	285	0	0	188	0	UCC...
Tyrosine	13	0	0	8	1.0	8.6	UUA...
Isoleucine	12	1.0	1.0	4.8	5.4	8.4	UUA...
Valine	0.6	0	37	0.4	29.8	75	UUG...
Leucine	4.9	79	36	5.1	157	44	UUC.. + UUG...
Cysteine	4.9	0	35	0	5.4	46	UUG... or UGG
Tryptophan	1.1	0	14	0	1.6	23	UGG...
Glycine	4.7	0	12	0.5	9.7	15	UGG...
Methionine	0.6	0	0	0.6	1.5	8	UGA...
Glutamic acid	1.5	0	0	1.2	0.44	6.2	UGA...
Lysine							UAA... (?)

Cracking the Genetic Code



Cracking the Genetic Code

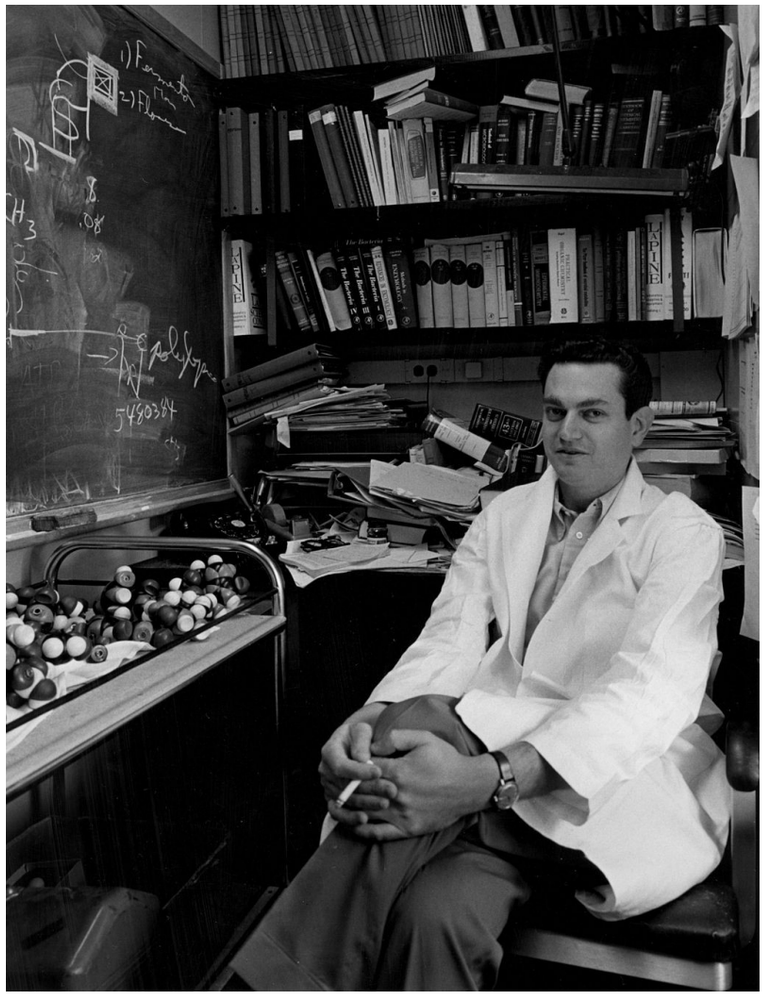
1966

Second letter

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

Cracking the Genetic Code

1968 Nobel Prize for Deciphering the Genetic Code



Marshall Nirenberg

Har Gobind Khorana

Robert Holley

Cracking the Genetic Code

11 August 1967, Volume 157, Number 3789 **SCIENCE**

Will Society Be Prepared?

"Science has reached a new frontier. A revolution far greater… then the atomic or hydrogen bomb"

- New York Times, 1961

"might lead to methods of tampering with life, of creating new diseases, of controlling minds, of influencing heredity, even perhaps in certain desired directions."

- Arne Tiselius, Nobel Laureate in Chemistry

Cracking the Genetic Code

Thank you!