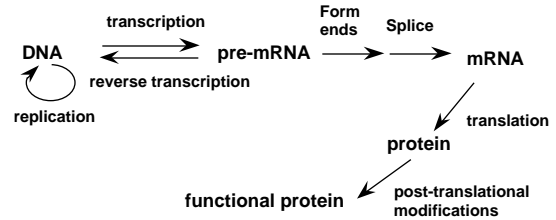


Genetic Code

Pathway for Gene Expression



3 nucleotides encode a single amino acid

- For 4 nucleotides to encode 20 amino acids, you need a coding unit of at least 3:
 - A coding unit of 2 nucleotides can only encode 16 amino acids (4x4)
 - A coding unit of 3 nucleotides can only encode 64 amino acids (4x4x4)
- Insertions or deletions of 1, 2, 4, 5, etc nucleotides cause a severe loss of function resulting from a change in the reading frame.
- But insertions or deletions of 3, 6, 9, etc have little effect on the phenotype, because the reading frame is not affected for most of the mRNA.

Experiments to decipher the code

Tools available

- Cell-free systems for translation
 - From bacteria, plants and animals
- Ability to synthesize polyribonucleotides
 - Polynucleotide phosphorylase can make RNAs from NDPs
 - Physiological function: reverse reaction for RNA degradation



Homopolymers of RNA direct synthesis of homopolypeptides

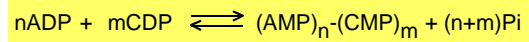
- Use of a single NDP as a substrate for polynucleotide phosphorylase will produce a homopolymer of that nucleotide
 - UDP as a substrate results in polyU as the product
- Addition of polyU to a cell-free translation system results in polyphenylalanine as the product.

Results of using homopolymers to program translation

- UUU encodes Phe
- AAA encodes Lys
- CCC encodes Pro
- GGG encodes Gly

Mixed co-polymers of RNA direct the incorporation of particular amino acids into polypeptides

- Mix two NDPs in a known ratio, and polynucleotide phosphorylase will catalyze synthesis of an RNA with those 2 nucleotides in that ratio, but random order.



e.g. ADP:CDP 5:1 $(\text{AMP})_5\text{-(CMP)}_1$

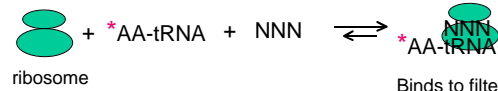
e.g. AACAAAAACAACAAAAAAAACAA

Frequency of incorporation of each amino acid correlates with frequency of occurrence of a particular trinucleotide

- For a 5:1 ratio of ADP:CDP → poly(AC) (5:1), A is present in the RNA 5/6 of the time, and C is present at a frequency of 1/6.
- Possible trinucleotides (codons) and resulting amino acids incorporated during translation:

Compo- sition	Num- ber	Relative Frequency	Relative Freq.	Amino Acid	Freq. of incorp.
3A	1	0.578	1.0	Lys	1.0
2A, 1C	3	3x0.116	3x0.20	Thr*, Asn, Gln	0.27*, 0.24, 0.24
1A, 2C	3	3x0.023	3x0.04	Pro**, His, Thr*	0.07, 0.07, (.04*)
3C	1	0.005	0.01	Pro**	(0.01**)

Defined trinucleotides stimulate binding of particular aminoacyl-tRNAs to ribosomes



Which trinucleotide will allow binding of a particular AA-tRNA to ribosomes?

AA-tRNA	pmoles AA-tRNA bound with:			
	no NNN	UUU	AAA	CCC
Phe-tRNA	0.34	1.56	0.20	0.30
Lys-tRNA	0.80	0.56	6.13	0.60
Pro-tRNA	0.24	0.20	0.18	0.73

Repeating sequence synthetic polynucleotides direct incorporation of particular amino acids

UCUCUCUCUCUCUCUCUCUC

↓ Ribosomes + AA-tRNAs

SerLeuSerLeuSerLeuSerLeu

→ UCU encodes Ser or Leu
CUC encodes Ser or Leu

Combine with other information. E.g. 5:1 random copolymer of C:U encodes Leu:Ser with a relative frequency of 0.20:0.04, then CUC encodes Leu and UCU encodes Ser.

Features of the Genetic Code

Table 3.4.4 The Genetic Code

1st	Position in Codon												3rd					
	U			C			A			G								
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
	UUA	Leu	UCA	Ser	UAA	Term	UGA	Term	A	UUA	Leu	UCA	Ser	UAA	Term	UGA	Term	A
C	UUG	Leu	UCG	Ser	UAG	Term	UGG	Trp	G	UUG	Leu	UCG	Ser	UAG	Term	UGG	Trp	G
	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC	Leu	CCG	Pro	CAC	His	CGC	Arg	C	CUC	Leu	CCG	Pro	CAC	His	CGC	Arg	C
A	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
G	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	AUG*	Met	ACG	Thr	AAG	Lys	AGG	Arg	G	AUG*	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
	GUG*	Val	GCG	Ala	GAG	Glu	GGG	Gly	G	GUG*	Val	GCG	Ala	GAG	Glu	GGG	Gly	G

* Sometimes used as initiator codons.

Degeneracy of the code

- Degeneracy refers to the fact that almost all amino acids are encoded by multiple codons.
- Degeneracy is found primarily in the 3rd position of the codon, i.e. the nucleotide in the 3rd position can change without changing the amino acid specified.
- In some cases, the 1st position is also degenerate.

Groupings of codons

- Of the 64 codons, 61 specify amino acids and the other 3 are signals to terminate translation
- 9 codon families.
 - E.g. encode Thr with
 - ACU
 - ACC
 - ACA
 - ACG
- 13 codon pairs
 - E.g. encode Asp with:

• GAU	Glu with
• GAC	GAA
	GAG

Codons for initiation of translation

- Major codon for initiation is AUG
- Regardless of codon used, the first amino acid incorporated in *E. coli* is formyl-Met.
- For the 4288 genes identified in *E. coli*:
 - AUG is used for 3542 genes.
 - GUG is used for 612 genes.
 - UUG is used for 130 genes.
 - AUU is used for 1 gene.
 - CUG may be used for 1 gene.

Codons for termination of translation

- UAA, UAG, UGA
- For the genes identified in *E. coli*:

UAA is used for	2705 genes.
UGA is used for	1257 genes.
UAG is used for	326 genes.

Genetic code is universal (almost)

- All organisms so far examined use the code as originally deduced (or something very close to it).
- The rare exceptions involve limited differences.
 - e.g. In RNA derived from mitochondrial DNA, UGA encodes Trp instead of serving as a stop codon.
 - Thus UGA and UGG form a codon pair in this case.

Differential codon usage

- Some codons are used much more frequently than others to encode a particular amino acid.
- The pattern of codon usage varies **between species** and even **among tissues** within a species.
- Correlates with **tRNA abundance**.
- Pattern of codon usage can be a predictor of level of expression of a gene.
- Preferred codon usage is a help in reverse genetics.

Wobble in anticodon-codon pairing

- Some nucleotides in the 1st position of the anticodon (in tRNA) can pair with >1 nucleotide in the 3rd position of the codon.

- G can pair with U and I can pair with U, C or A.

1st position anticodon 3rd position in codon

C	G
A	U
U	A or G
G	C or U
I (inosinic acid)	U, C or A

- Result: 61 codons can read by as few as 31 tRNAs

Types of mutations in coding regions

- Silent (synonymous)
 - Do **not** change the encoded amino acid
 - Occur in **degenerate** positions in the codon
 - Are often not subject to purifying selection and thus occur more frequently in evolution
- Nonsilent (nonsynonymous)
 - **Do** change the encoded amino acid
 - Occur in **non-degenerate** positions in the codon
 - Are more likely to be subject to purifying selection and thus occur less frequently in evolution

Changes that alter the encoded product

- **Missense**: cause a **replacement** of an amino acid
 - e.g. CAA (Gln) -----> CGA (Arg)
- **Nonsense**: cause a **termination** of translation
 - e.g. CAA (Gln) -----> UAA (term)
- **Frameshift**: insertion or deletion that changes the reading frame
 - e.g. CAA (Gln) -----> C-A (frameshift)

Quiz

1. Which of the following mutations could occur by a single nucleotide substitution?
 - 1.1 Phe to Leu
 - 1.2 Lys to Ala
 - 1.3 Ala to Thr
2. A codon for Lys can be converted by a single nucleotide substitution to a codon for Ile. What is the sequence of the original codon for Lys?