12.5: The Genetic Code

Introduction

You can think of the sequences of bases in the coding strand of DNA or in messenger RNA as coded instructions for building protein chains out of amino acids. There are 20 amino acids used in making proteins, but only four different bases to be used to code for them.

Obviously one base can't code for one amino acid. That would leave 16 amino acids with no codes.

If you took two bases to code for each amino acid, that would still only give you 16 possible codes (TT, TC, TA, TG, CT, CC, CA and so on) - still not enough.

However, if you took three bases per amino acid, that gives you 64 codes (TTT, TTC, TTA, TTG, TCT, TCC and so on). That's enough to code for everything with lots to spare. You will find a full table of these below.

A three base sequence in DNA or RNA is known as a codon.

The code in DNA

The codes in the coding strand of DNA and in messenger RNA aren't, of course, identical, because in RNA the base uracil (U) is used instead of thymine (T).

The table shows how the various combinations of three bases in the coding strand of DNA are used to code for individual amino acids - shown by their three letter abbreviation.

first base in codon	second base in codon: T	second base in codon: C	second base in codon: A	second base in codon: G	third base in codon
T	TTT Phe	TCT Ser	TAT Tyr	TGT Cys	T
	TTC Phe	TCC Ser	TAC Tyr	TGC Cys	C
	TTA Leu	TCA Ser	TAA stop	TGA stop	A
	TTG Leu	TCG Ser	TAG stop	TGG Trp	G
C	CTT Leu	CCT Pro	CAT His	CGT Arg	T
	CTC Leu	CCC Pro	CAC His	CGC Arg	C
	CTA Leu	CCA Pro	CAA Gln	CGA Arg	A
	CTG Leu	CCG Pro	CAG Gln	CGG Arg	G
A	ATT Ile	ACT Thr	AAT Asn	AGT Ser	T
	ATC Ile	ACC Thr	AAC Asn	AGC Ser	C
	ATA Ile	ACA Thr	AAA Lys	AGA Arg	A
	ATG Met	ACG Thr	AAG Lys	AGG Arg	G
G	GTT Val	GCT Ala	GAT Asp	GGT Gly	T
	GTC Val	GCC Ala	GAC Asp	GGC Gly	C
	GTA Val	GCA Ala	GAA Glu	GGA Gly	A
	GTG Val	GCG Ala	GAG Glu	GGG Gly	G

The table is arranged in such a way that it is easy to find any particular combination you want. It is fairly obvious how it works and, in any case, it doesn't take very long just to scan through the table to find what you want.

The colors are to stress the fact that most of the amino acids have more than one code. Look, for example, at leucine in the first column. There are six different codons all of which will eventually produce a leucine (Leu) in the protein chain. There are also six for serine (Ser).

In fact there are only two amino acids which have only one sequence of bases to code for them - methionine (Met) and tryptophan (Trp). You have probably noticed that three codons don't have an amino acid written beside them, but say "stop" instead. For obvious reasons these are known as stop codons. We'll leave talking about those until we have looked at the way the code works in messenger RNA.

The code in messenger RNA

You will remember that when DNA is transcribed into messenger RNA, the sequence of bases remains exactly the same, except that each thymine (T) is replaced by uracil (U). That gives you the table:

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

In many ways, this is the more useful table. Messenger RNA is directly involved in the production of the protein chains (see the next page in this sequence). The DNA coding chain is one stage removed from this because it must first be transcribed into a messenger RNA chain.

Start and stop codons

The **stop codons** in the RNA table (UAA, UAG and UGA) serve as a signal that the end of the chain has been reached during protein synthesis.

The codon that marks the start of a protein chain is AUG. If you check the table, that's the amino acid, methionine (Met). That ought to mean that every protein chain must start with methionine. That's not quite true because in some cases the methionine can get chopped off the chain after synthesis is complete.

Contributors

Jim Clark (Chemguide.co.uk)