Secret Codon

Write a message in DNA

Give new meaning to DNA as beads on a string.

Introduction

Proteins are long chains of individual amino acid subunits. The order of the amino acids in the chain is determined by the DNA sequence of the gene that encodes for it. DNA is a long chain of four different nucleotides (adenine, thymine, cytosine, and guanine), often abbreviated A, G, C, and T. These 4 nucleotides (sometimes referred to as bases) need to give instructions for the 20 different amino acids that compose proteins. Each amino acid is encoded by a sequence of three DNA bases, called a **codon**. Since it takes three DNA bases to designate an amino acid, there are enough combinations of the 4 different bases to represent all of the amino acids, as well as three stop codons that indicate when the protein ends. Each base can be in any position, which yields 4^3, or 64, possible combinations, so there is some redundancy between the 20 amino acids. This just means that a given amino acid can be encoded by more than one DNA codon sequence. This is commonly referred to as the **genetic code**. Individual amino acids are often abbreviated using one or three letter abbreviations. For example, the amino acid arginine can be abbreviated Arg or R. Since there are only 20 different amino acids, there are 6 letters of the alphabet that don't encode a specific amino acid. With the 20 letters that do, however, you can write a secret message and use the genetic code to determine the DNA sequence that corresponds to your amino acid code.

Materials and Preparation

Pony beads – 4 different colors
Cotton string
Amino acid codon table

To do and notice

1. Think of a word or short phrase that you want to encode into your DNA strand. Make sure it can be spelled or sounded out without using the letters B, J, O, U, X, or Z. These letters are not abbreviations for any of the amino acids.

2. You will need to determine what amino acids the letters in your phrase correspond to. Then, you will need to know the DNA sequence that encodes for those amino acids. Use the amino acid codon table to determine what DNA sequences correspond to the one-letter amino acid abbreviations that make up your word.

3. All proteins start with a methionine residue that is encoded by the DNA sequence ATG. They end when the DNA encodes one of the three stop codons. Make your DNA strand by stringing the beads in the order of your phrase. Be sure to include the proper start and stop codons in your sequence.

Use the following color key for the bases:

T: yellow
C: blue
A: red
G: green

4. Trade strands with a friend and see if you can decode their secret message!

What's going on?

The single-letter amino acid abbreviations provide a fun way to write secret messages using the genetic code. The message can be written with four different colors that represent the four different bases that make up DNA. Using the key above, I made the chain in the picture at the top of the page. The sequence is:

RYGGBBRGRRBGGBYRRBGRYYBYYGYRYYGRGRRYYGYGRGYRR

The sequence starts with RYG, which corresponds to the bases ATG. This triplet is the code of the methionine start codon that begins every protein sequence. The next three beads are GBB, which represent GCC. This codes for alanine, which has the single letter abbreviation "A". Uncode the remaining sequence and you should get the phrase "art and science" before ending in a stop codon.

In January 2008, researchers at the J. Craig Venter Institute announced that they had constructed the entire genome of a small bacterium from scratch, thus creating the first example of synthetic life. To distinguish the man-made genome from the natural one, the scientists inserted "watermarks" into the DNA sequence. These sequences were decoded to their one letter amino acid abbreviations and revealed five watermarks commemorating those who had worked on the project: VENTERINSTITVTE, CRAIGVENTER, HAMSMITH, CINDIANDCLYDE, GLASSANDCLYDE

1st base	2nd base			
	T	C	A	G
T	TTT Phenylalanine/Phe/F	TCT Serine/Ser/S	TAT Tyrosine/Tyr/Y	TGT Cysteine/Cys/C
	TTC Phenylalanine/Phe/F	TCC Serine/Ser/S	TAC Tyrosine/Tyr/Y	TGC Cysteine/Cys/C
	TTA Leucine/Leu/L	TCA Serine/Ser/S	TAA Ochre (Stop)	TGA Opal (Stop)
	TTG Leucine/Leu/L	TCG Serine/Ser/S	TAG Amber (Stop)	TGG Tryptophan/Trp/W
C	CTT Leucine/Leu/L	CCT Proline/Pro/P	CAT Histidine/His/H	CGT Arginine/Arg/R
	CTC Leucine/Leu/L	CCC Proline/Pro/P	CAC Histidine/His/H	CGC Arginine/Arg/R
	CTA Leucine/Leu/L	CCA Proline/Pro/P	CAA Glutamine/Gln/Q	CGA Arginine/Arg/R
	CTG Leucine/Leu/L	CCG Proline/Pro/P	CAG Glutamine/Gln/Q	CGG Arginine/Arg/R
A	ATT Isoleucine/Ile/I	ACT Threonine/Thr/T	AAT Asparagine/Asn/N	AGT Serine/Ser/S
	ATC Isoleucine/Ile/I	ACC Threonine/Thr/T	AAC Asparagine/Asn/N	AGC Serine/Ser/S
	ATA Isoleucine/Ile/I	ACA Threonine/Thr/T	AAA Lysine/Lys/K	AGA Arginine/Arg/R
	ATG Methionine/Met/M	ACG Threonine/Thr/T	AAG Lysine/Lys/K	AGG Arginine/Arg/R
G	GTT Valine/Val/V	GCT Alanine/Ala/A	GAT Aspartic acid/Asp/D	GGT Glycine/Gly/G
	GTC Valine/Val/V	GCC Alanine/Ala/A	GAC Aspartic acid/Asp/D	GGC Glycine/Gly/G
	GTA Valine/Val/V	GCA Alanine/Ala/A	GAA Glutamic acid/Glu/E	GGA Glycine/Gly/G
	GTG Valine/Val/V	GCG Alanine/Ala/A	GAG Glutamic acid/Glu/E	GGG Glycine/Gly/G

letters that do not correspond with a specific amino acid: B,J,O,U,X,Z