

Biology Challenge #3

Sometimes, biologists find a molecule of DNA from an unknown species. In order to identify the species it belongs to, they first calculate the *GC content* of the DNA molecule, which is the percentage of guanine (G) and cytosine (C) nucleotides inside of it. Why is this useful? Because any two organisms of the same species share 99.9% of their DNA; since we know that cytosine and guanine always appear in equal amounts in a DNA molecule, we can calculate the GC content for an unknown strand and compare it to the GC content of other species' strands. If we find a match, we've identified the species of the unknown strand; otherwise, we may have found an entirely new species.

For example, the GC content of this DNA strand:

AGCTATAG

is 37.5%. There are 2 guanine nucleotides, and 1 cytosine nucleotide, for a total of 3. Dividing 3 by 8, which is the total number of nucleotides in this DNA strand, gives us 0.375. We finally multiply by 100 to get the DNA strand's GC content of 37.5%.

Write a Python program that asks the user for a DNA sequence, then outputs its GC content. For instance, if the user gives your program the following DNA strand:

TAGCCTAATACGCC

it should output the following to the screen:

GC content: 50.0%

As in the previous challenges, you should write your program such that someone can give it *any* DNA strand they want. One thing to be careful of: the GC content *is always a percentage and it is always a decimal number*. When you divide two integers in Python, Python will give you back an integer: for example, `3 / 8` in Python gives us 0, not 0.375. To tell Python to use decimal division, use the float function: for example, `float(3) / float(8)` gives us 0.375.