

DNA is a very stable and compact storage medium. The language of DNA consists of four bases, each of which may be thought of as a character in the DNA alphabet. For our purposes, these characters (in order) are: A, T, G, and C. We would like to store digital data in DNA.

Over the next 24 hours, please construct a program in the language of your choice that satisfies as many of the following objectives as possible. Please tell us where we can clone your solution, or email your source only in a tarball or zip file.

Objective 1: Provide a mechanism to encode arbitrary standard ASCII text strings as DNA. Recall that ASCII characters have numeric values of 0 to 255.

Examples:

Input: a, Output: TGAT

Input: cat, Output: TGA~~CT~~GATTCTA

Objective 2: RNA is very similar to DNA, but it replaces T with U. Make your program capable of encoding as either DNA or RNA.

Objective 3: Provide an interface for identifying standard ASCII text substrings encoded as DNA; the interface should provide the zero-based index where the substring begins or -1 if it does not exist.

Objective 4: DNA actually consists of two, complementary strands that are attached in such a way that every A matches T and G matches C (and the inverses). Given a complementary strand of DNA, output the ASCII equivalent of the primary strand.

Examples:

Input: ACTGACTAAGAT, Output: cat

Objective 5: Given two (single) strands of DNA, provide an interface that allows finding the longest common subsequence of the two. Recall that common subsequences need not be consecutive.