

# A Tutorial on Viewing and Checking the Local GC Content of an Optimized Gene Sequence

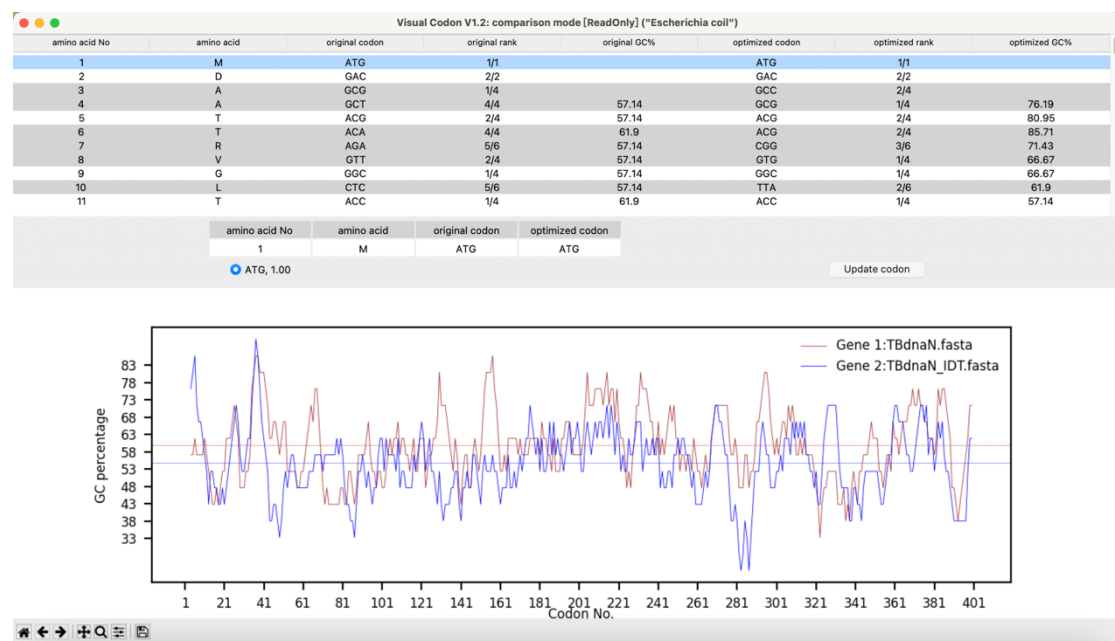
## Check local GC content of the optimized gene sequence

Here, we use the sequence optimized by the IDT tool as an example to show how to check the local GC content of the optimized gene sequence. Please note that we show the operation steps using macOS. The operation steps in Windows and Linux are similar.

- (1) Enter the Web interface of the IDT codon optimization tool at <https://sg.idtdna.com/pages/tools/codon-optimization-tool>, import the original sequence TBdnaN.fasta, select the expression host, import the codon frequency table, and export the optimized fasta sequence. Please note that this step is conducted in the Web page of IDT (<https://sg.idtdna.com/pages/tools/codon-optimization-tool>), not with our program;
- (2) Create a new folder on the local computer desktop, customize the folder name, and copy the visual\_codon.py, TBdnaN.fasta, and TBdnaN\_IDT.fasta to the folder;
- (3) Open the terminal, cd to the above folder, enter the following command, and click Enter to run.

```
python3.12 visual_codon.py
```

After selecting a host organism, a GUI interface will appear, with the title visual codon. Click the menu in the upper left corner, File-Open to open the original gene sequence, TBdnaN.fasta, File-Import Gene 2 to compare to import the optimized gene sequence TBdnaN\_IDT.fasta. At this time, you can see that each column of the Treeview table at the top of the GUI interface corresponds to the relevant information on amino acids at each site before and after optimization. The middle of the interface is the modifiable codon and the frequency of codon usage. The bottom of the interface is a comparison chart of GC content before and after optimization, as shown in Figure 1.

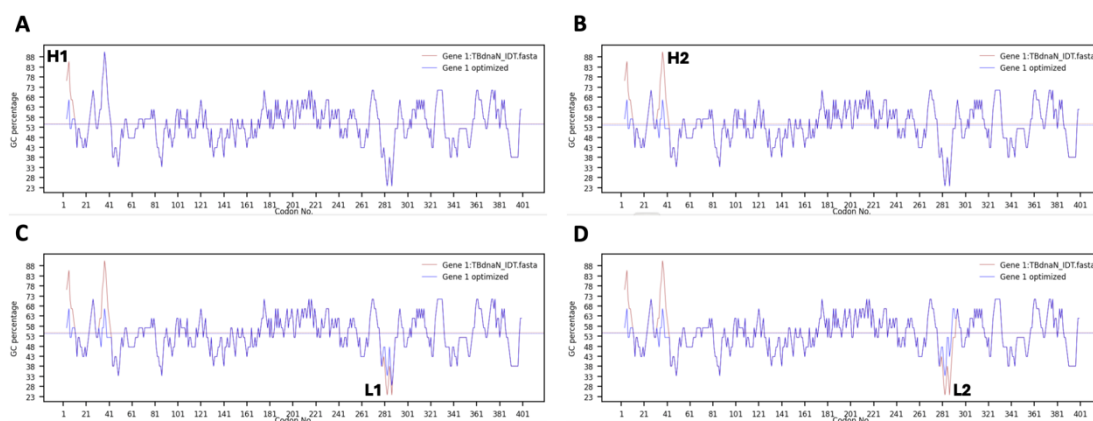


**Figure 1. The program GUI.** The brown and blue lines of the GC content chart at the bottom of the interface represent the GC content of each locus of the Gene 1 (original gene sequence) and Gene 2 (gene sequence after IDT tool optimization), respectively. The brown and blue horizontal lines record the average GC contents of Gene 1 and Gene 2, respectively.

### Adjust local GC content of gene sequence

If there are local highs and lows in the GC content of the optimized sequence, further optimization is required. Here we take IDT's optimized sequence TBdnaN\_IDT.fasta as an example to describe the operation of our software.

(1) Click the menu File-Open Gene 1 to edit to open the TBdnaN\_IDT.fasta. Find the local high and low points according to the GC content graph at the bottom of the GUI interface. Mouse left-clicking the points in the graph will select the corresponding lines in the Treeview table. Or user can use the scrollbar on the left to navigate to the points in the graph. By referring to the codon frequency shown in the middle of the graphical interface, the user can modify the codon, and the GC content map at the bottom of the interface will display the change of GC content in real-time, as shown in Figure 2.



**Figure 2. The process of adjusting the local high and low points of GC content.** (A) Adjusting the GC content of the first peak (H1). (B) Adjusting the GC content of the second peak (H2). (C) Adjusting the GC content of the first valley (L1). (D) Adjusting the GC content of the second valley (L2).

(2) After the user finishes adjusting the codon, click the menu File-Save optimized gene to save the adjusted gene sequence (TBdnaN\_IDT\_visual\_codon.fasta), click the menu File-'Export table to txt' to export the entire Treeview table (TBdnaN\_IDT\_visual\_codon.txt), click the menu File-'Export changed codons' to export the changed codons (changed\_codons.txt).