Mini Project: Comprehensive Sequence Analysis of the Human TNF Gene

Project Title:Comprehensive Sequence Analysis of the Human TNF Gene

Objective:

To apply bioinformatics skills to download, analyze, and interpret the sequence of the human TNF gene,which encode a proinflammatory cytokine called TNF.

Project Overview:

In this mini project, you will perform a series of bioinformatics tasks using the human TNF gene as your sequence of interest. The project will guide you through downloading the sequence, translating it, finding ORFs, analyzing sequence composition, identifying transcription factor binding sites, searching for functional motifs, predicting coding/noncoding regions, and converting sequence file formats.

Task 1: Download a Biological Sequence from NCBI and View/Edit It

Objective: Download the human TNF gene sequence and view it using BioEdit.

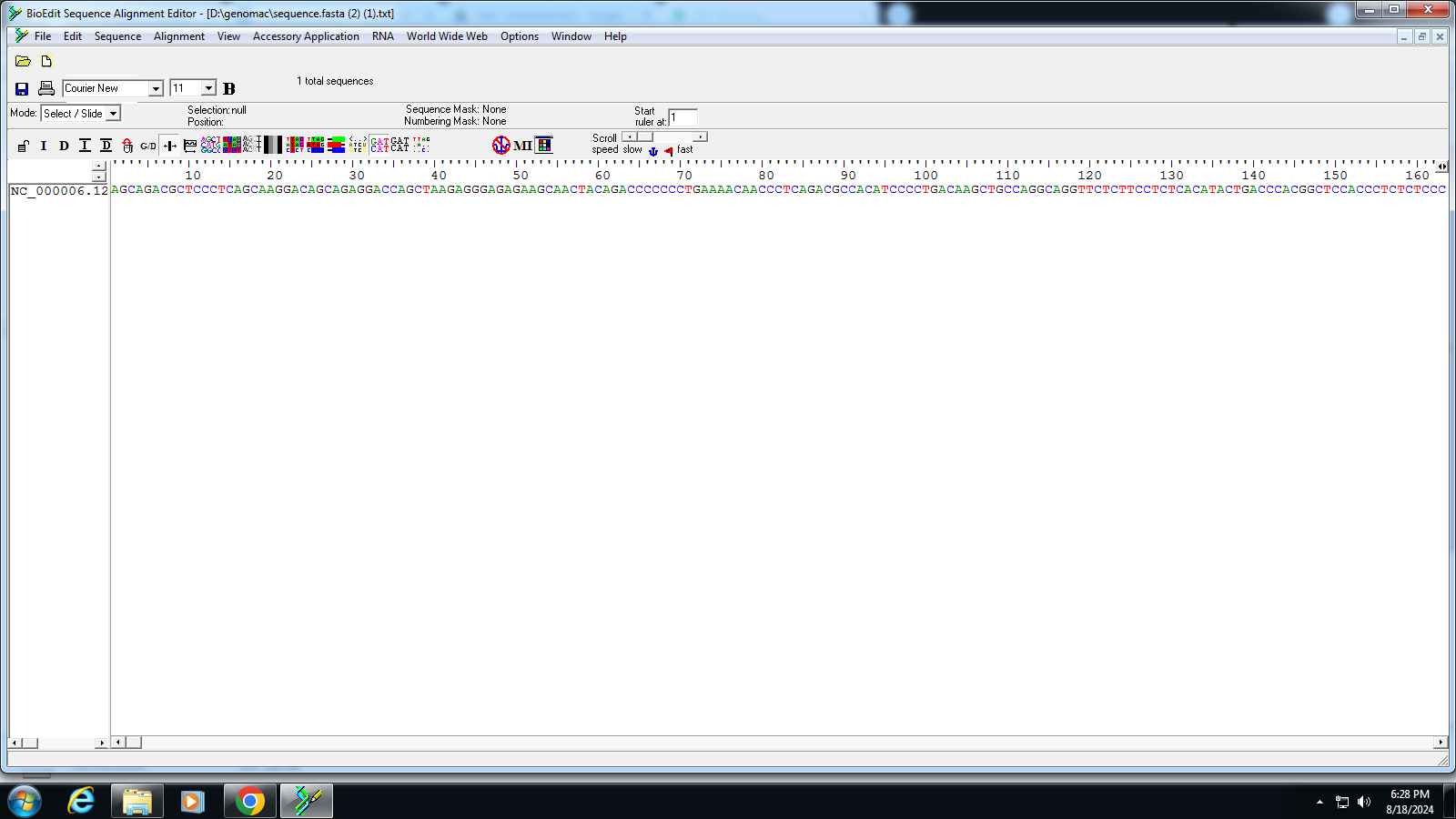
Access the NCBI homepage at NCBI:

tumor necrosis factor [Homo sapiens (human)]

Chromosome 6, NC\_000006.12 (31575565..31578336)

This gene encodes a multifunctional proinflammatory cytokine that belongs to the tumor necrosis factor (TNF) superfamily. This cytokine is mainly secreted by macrophages. It can bind to, and thus functions through its receptors TNFRSF1A/TNFR1 and TNFRSF1B/TNFBR. This cytokine is involved in the regulation of a wide spectrum of biological processes including cell proliferation, differentiation, apoptosis, lipid metabolism, and coagulation. This cytokine has been implicated in a variety of diseases, including autoimmune diseases, insulin resistance, psoriasis, rheumatoid arthritis ankylosing spondylitis, tuberculosis, autosomal dominant polycystic kidney disease, and cancer. Mutations in this gene affect susceptibility to cerebral malaria, septic shock, and Alzheimer disease. Knockout studies in mice also suggested the neuroprotective function of this cytokine.

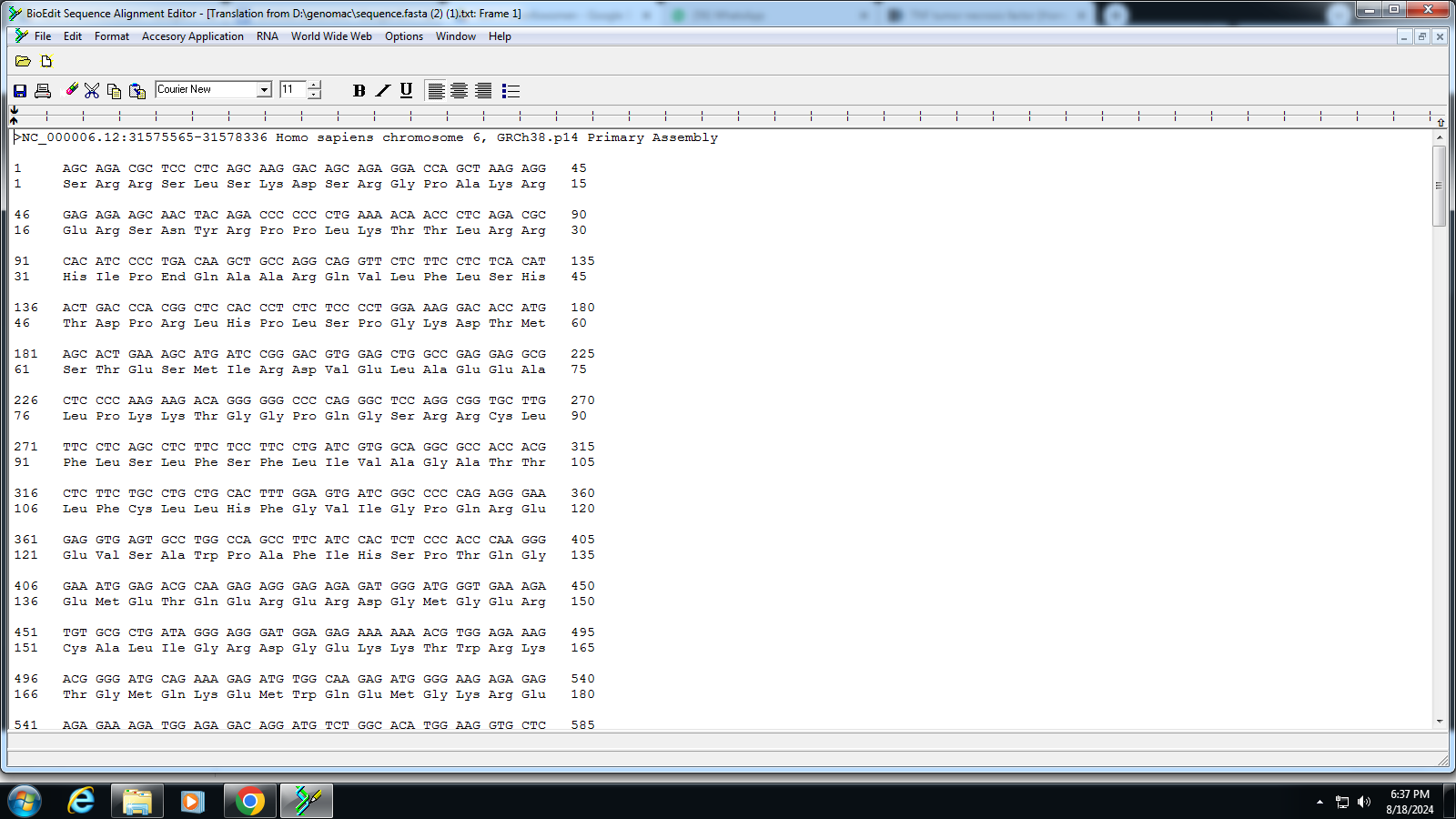
Open the sequence in BioEdit and view/edit it.



Task 2: Generate a Translation of a DNA or RNA Sequence into Amino Acids

Objective:Translate the DNA sequence of the TNF gene into an amino acid sequence.

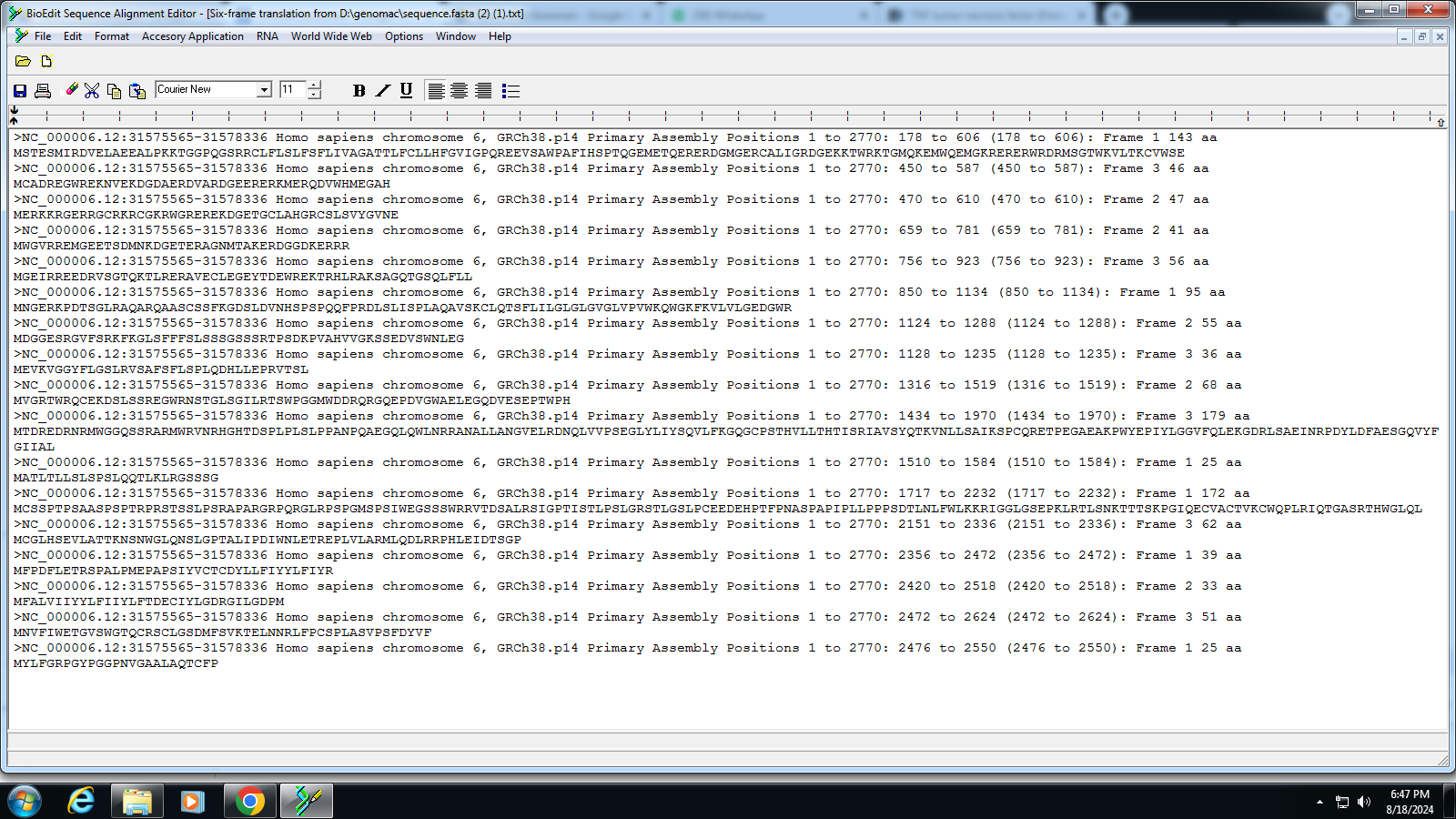
'Translate' feature in BioEdit



INTERPRETATION: Three codons are needed to specify three amino acids. Codons can be described as messengers that are located on the messenger RNA (mRNA). It is a sequence of three nucleotides that code for one specific amino acid; therefore, every three nucleotides represent one codon.the bioedit is used to successfully translate the sequence to protein

Task 3: Find ORFs (Open Reading Frames) in a DNA or RNA Sequence

Objective: Identify the ORFs within the TNF gene sequence.



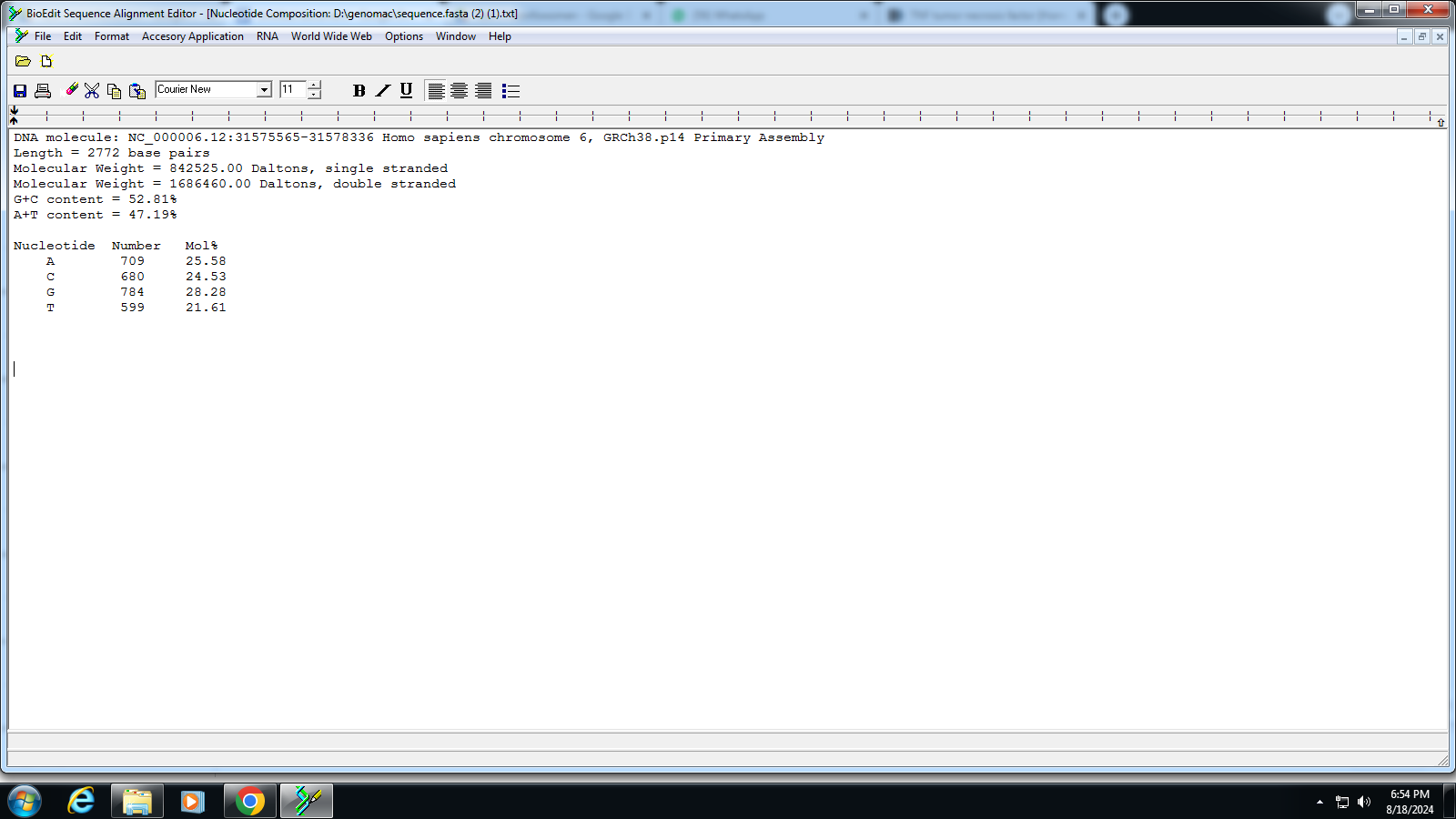
Interpret : An ORF is a potential gene identified by a potential protein-coding segment in DNA bordered by start and stop codons;

The opening reading frame of our sequence from position 1-2770 was found to be in Frame 1 143 aminoacid were found between position 178 to 606

Frame 2 47 aminoacid were found between position 470 to 610

Task 4: Analyze Sequence Composition (Nucleotide or Amino Acid Frequencies)

Objective:Analyze the nucleotide composition of the TNF gene sequence.



DNA molecule: NC\_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly

Length = 2772 base pairs

Molecular Weight = 842525.00 Daltons, single stranded

Molecular Weight = 1686460.00 Daltons, double stranded

G+C content = 52.81%

A+T content = 47.19%

Nucleotide Number Mol%

A 709 25.58

C 680 24.53

G 784 28.28

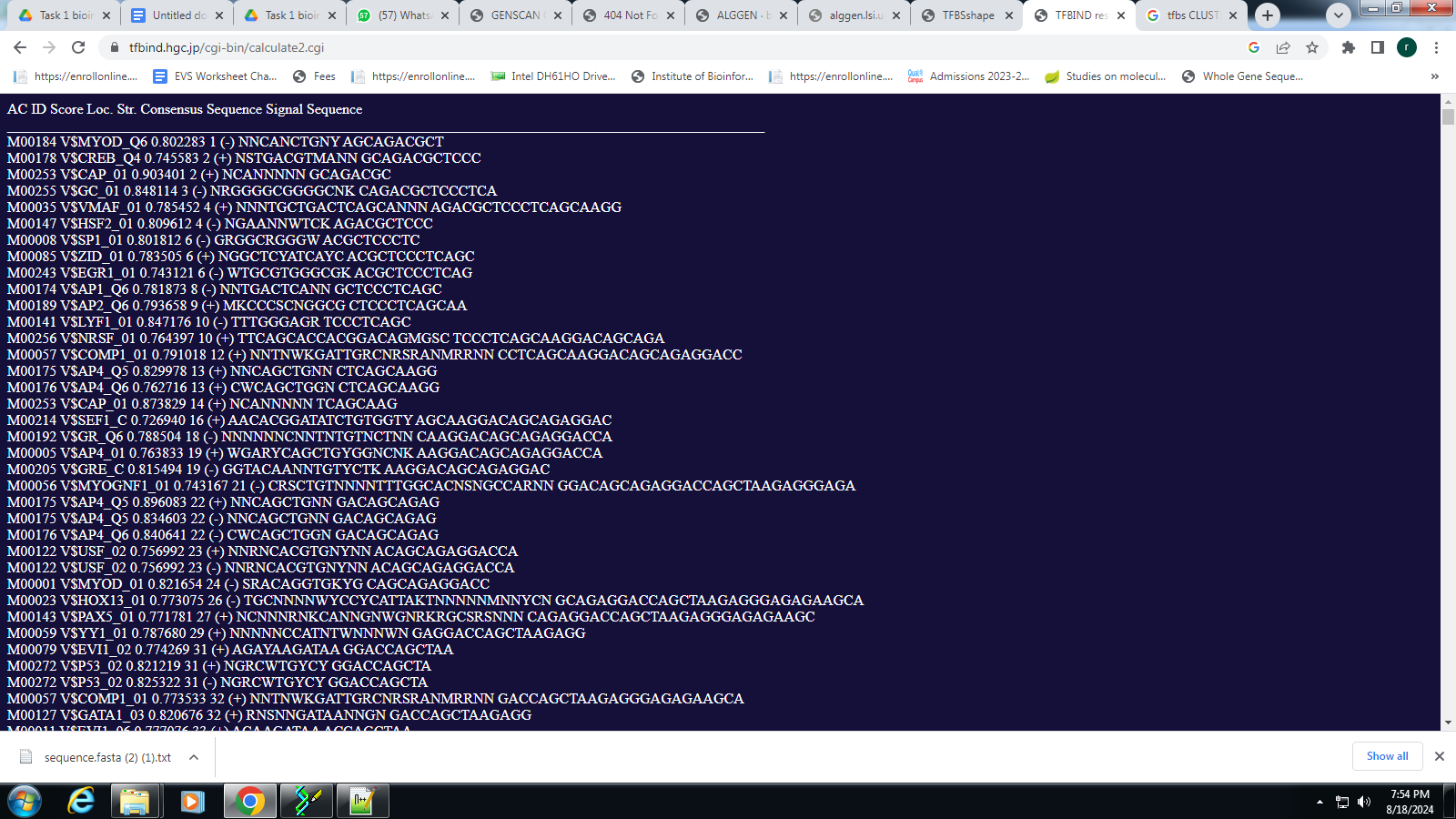
T 599 21.61

INTERPRETATION : the composition was analyzed using bioedit software G+C was found to be 52.81% and A+T was found to be 47.19%

Task 5: Identify Transcription Factor Binding Sites Using the PROMO Tool

Objective:

Identify potential transcription factor binding sites in the TNF gene promoter region.



Interpretation : i was not able to access promo tool so i used TFBIND tool

TFBIND tool software for searching transcription factor binding sites (including TATA boxes, GC boxes, CCAAT boxes, transcription start sites (TSS)). This tool uses weight matrix in transcription factor database TRANSFAC R.3.4 developed by Dr. Wingender et al, and the cut-offs originally estimated by our research

Column1: Transcription factor matrix ID (from TRANSFAC R.3.4).

Column2: Transcription factor label (from TRANSFAC R.3.4). V means vertebrate.

Column3: Similarity (0.0-1.0) between a registered sequence for the transcription factor binding sites and the input sequence (at the position shown in the next column).

Column4: Position on the input sequence.

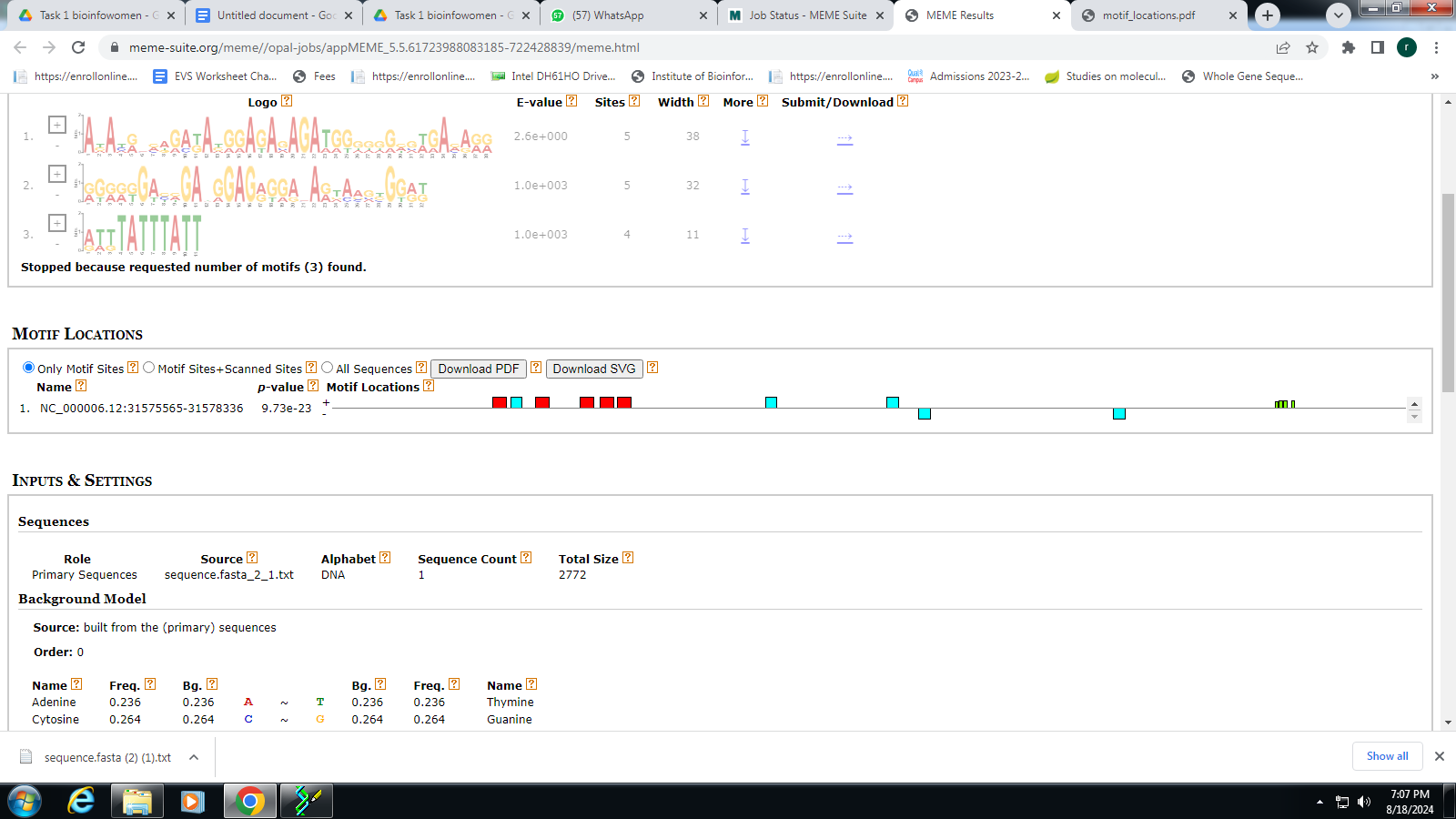
Column5: Strandness. + and - means forward and reverse strands that the transcription factor binds, respectively.

Column6: Consensus sequence (fixed) of the transcription factor binding sites. S = C or G, W = A or T, R = A or G, Y = C or T, K = G or T, M = A or C, N = any base pair.

Column7: Subsequence from the input sequence at the position - corresponding to the consensus sequence.

Task 6: Search for Functional Motifs in a Genome or Transcriptome Using MEME Suite

Objective:Search for functional motifs in the TNF gene sequence using MEME Suite.

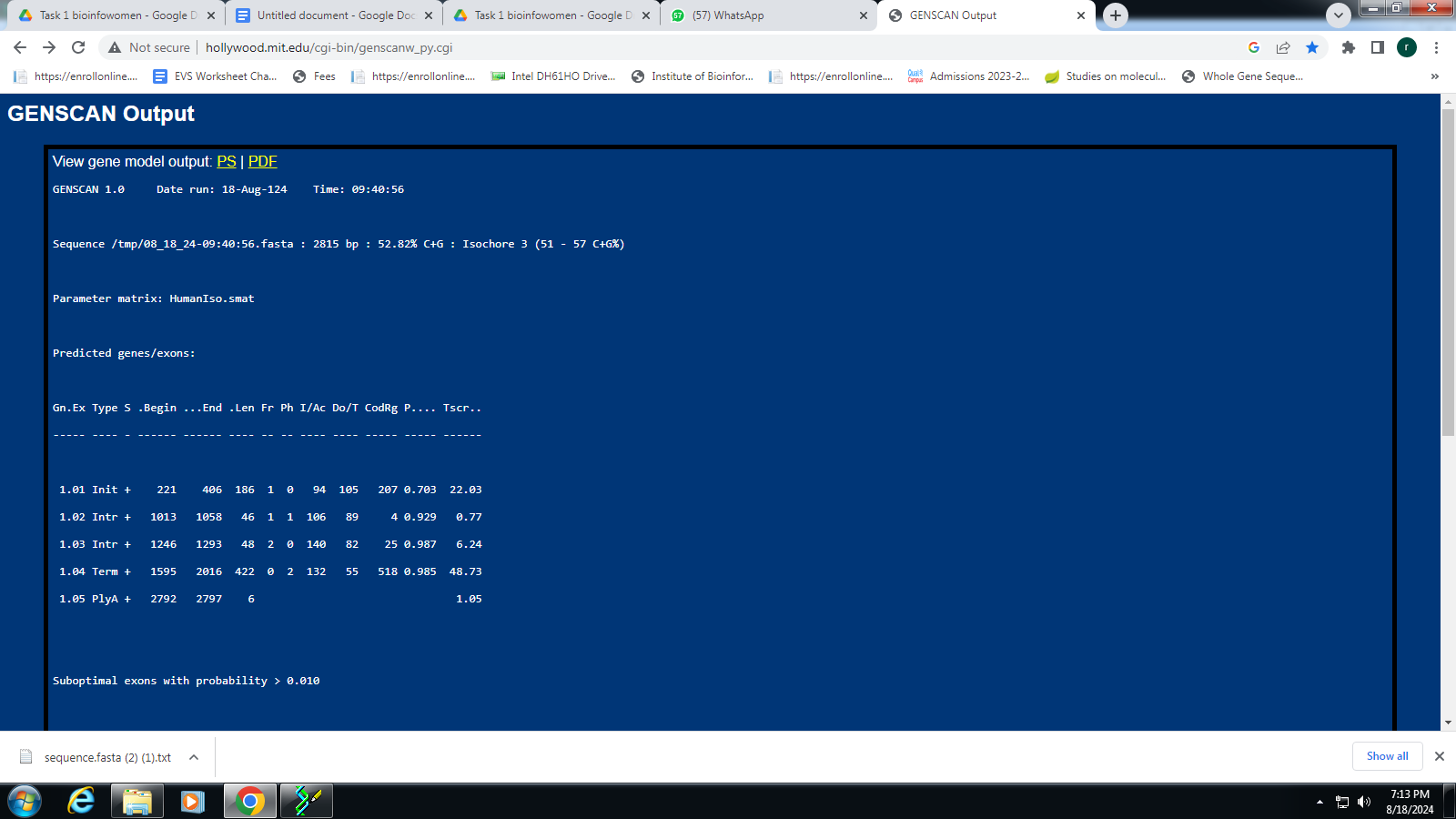


INTERPRET : amino acid residues that are conserved across species and many of which have been assigned functional roles based on experimental evidence

The MEME Suite allows you to discover novel motifs in collections of unaligned nucleotide or protein sequences, and to perform a wide variety of other motif-based analyses motifs were identified , E value of each motifs were detected. In motif position p valve and motif position was provide Motifs were depicted as boxes to known position of motifs

Task 7: Predict Coding/Non-Coding Regions in a Genome UsingGENSCAN

Objective: Predict the coding and non-coding regions within the TNF gene sequence.



INTERPRET :Genes are composed of regions that go on to form codons (and hence are considered protein coding), and regions that are only there for regulatory reasons (meaning they do not code for proteins). Coding regions are referred to as exons, whereas non-coding regions are known as introns.

GENSCAN is a program to identify complete gene structures in genomic DNA. It is a GHMM-based program that can be used to predict the location of genes and their exon-intron boundaries in genomic sequences from a variety of organisms.

genscan is used to identify coding and noncoding regions.

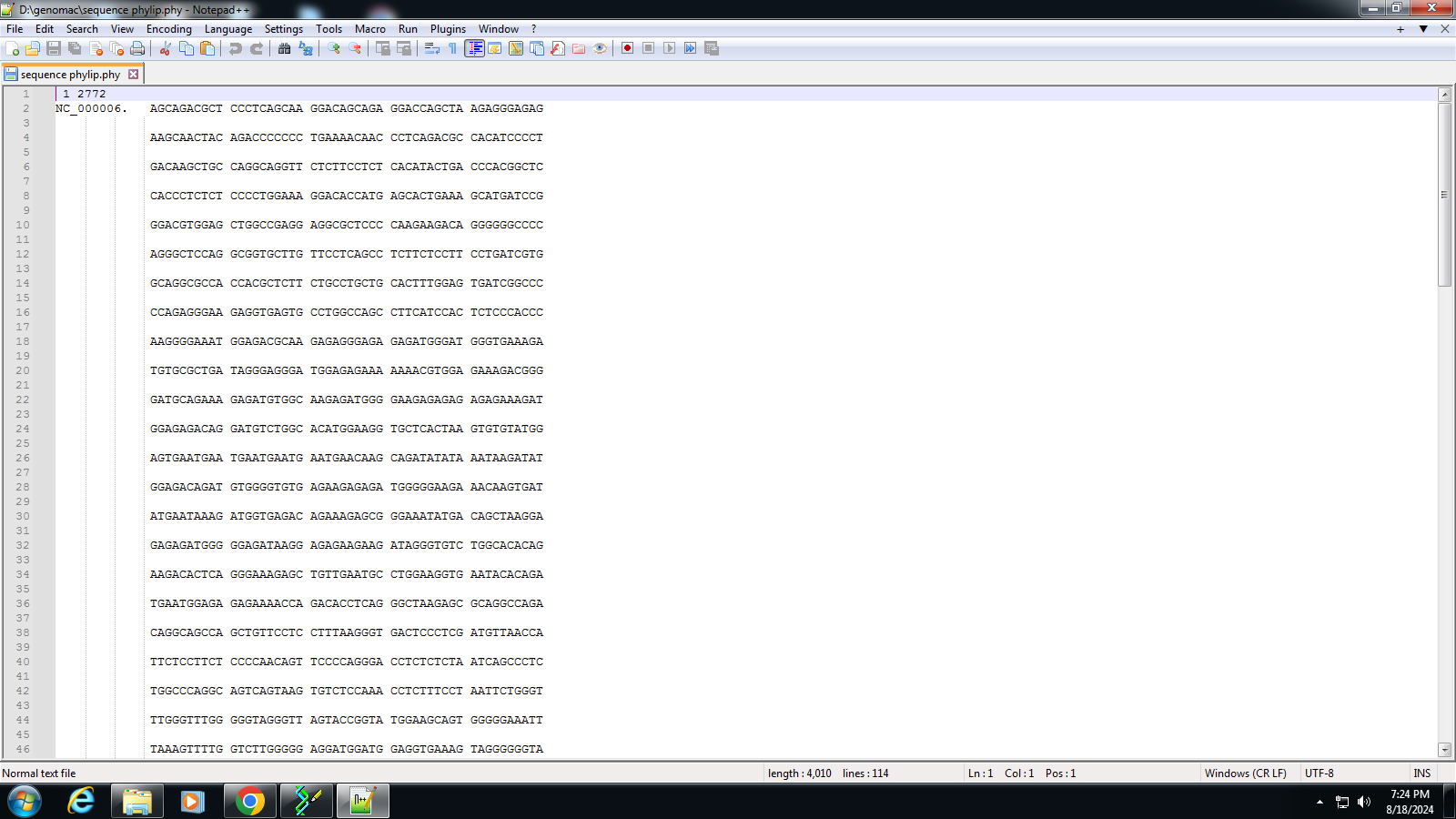
This provided us with length 2815 bp .

In predicted genes : gene no ,strand type +or- ,start and stop position of exon ,frame initiator signal score ,terminal score ,probability ,quality of exon , exon score were shown

Task 8: Convert Between Sequence File Formats Using BioEdit (FASTA to

PHYLIP)

Objective: Convert the TNF gene sequence from FASTA format to PHYLIP format.



INTERPRET :PHYLIP format is a plain text format containing exactly two sections: a header describing the dimensions of the alignment, followed by the multiple sequence alignment itself. The PHYLIP file format stores a multiple sequence alignment.

Bioedit tool was used to convert fasta into phylip format.