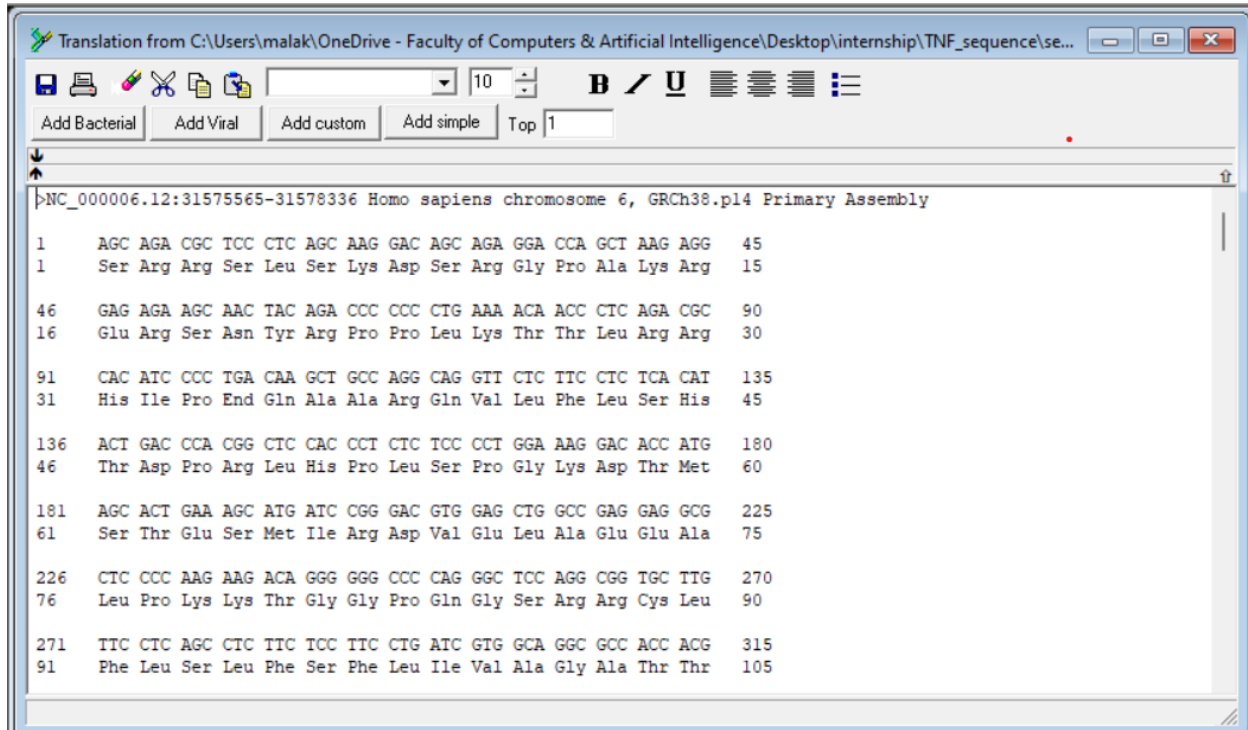




## Task 2: Generate a Translation of a DNA or RNA Sequence into Amino Acids

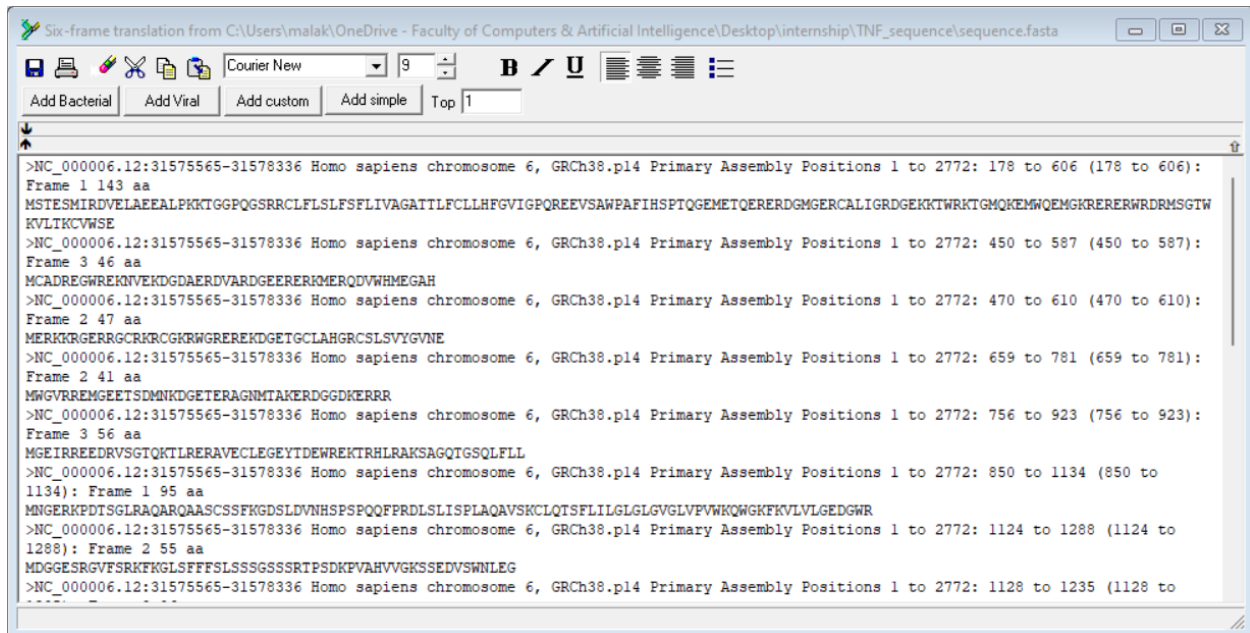
### Output:

In this task, I translated the DNA sequence of the TNF gene into its corresponding amino acid sequence using BioEdit. The screenshot captures the output generated by the 'Translate' feature in BioEdit, showcasing the amino acid sequence derived from the TNF gene. This sequence is crucial for understanding the protein structure and function of TNF, which plays a significant role in inflammation and immune responses.



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

**Output:** In analysis of the TNF gene sequence, I identified several potential Open Reading Frames (ORFs), each characterized by specific start and stop positions. These ORFs represent candidate protein-coding regions within the gene.



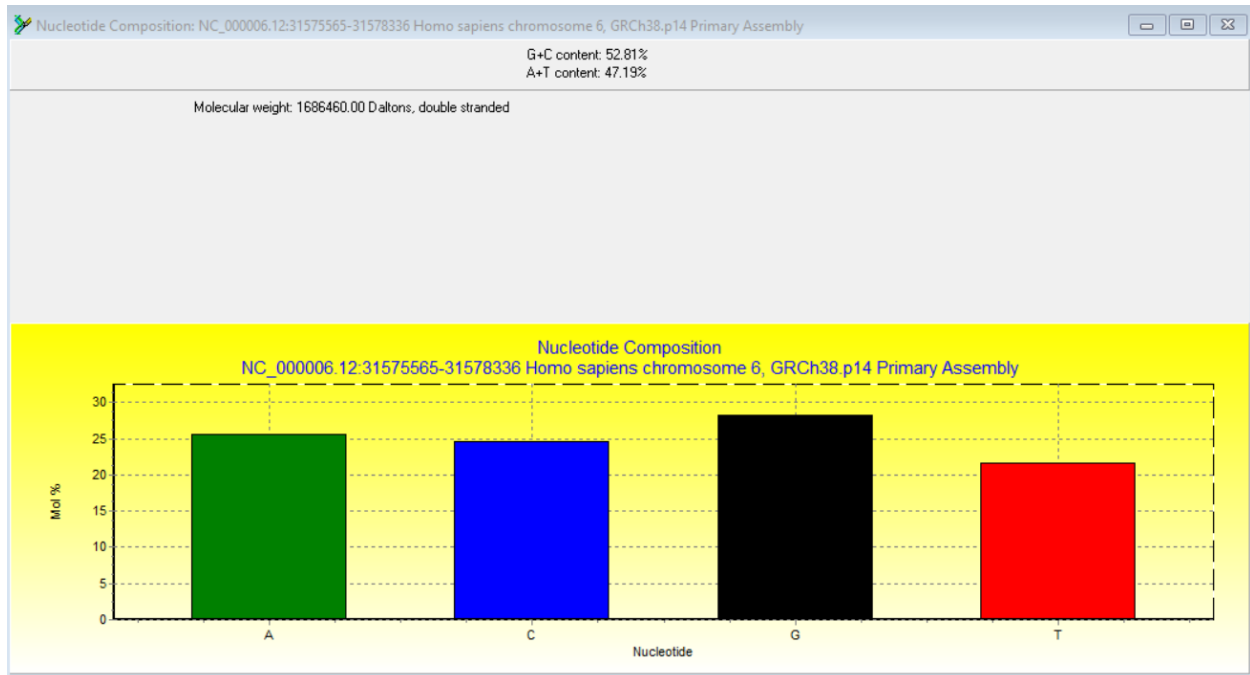
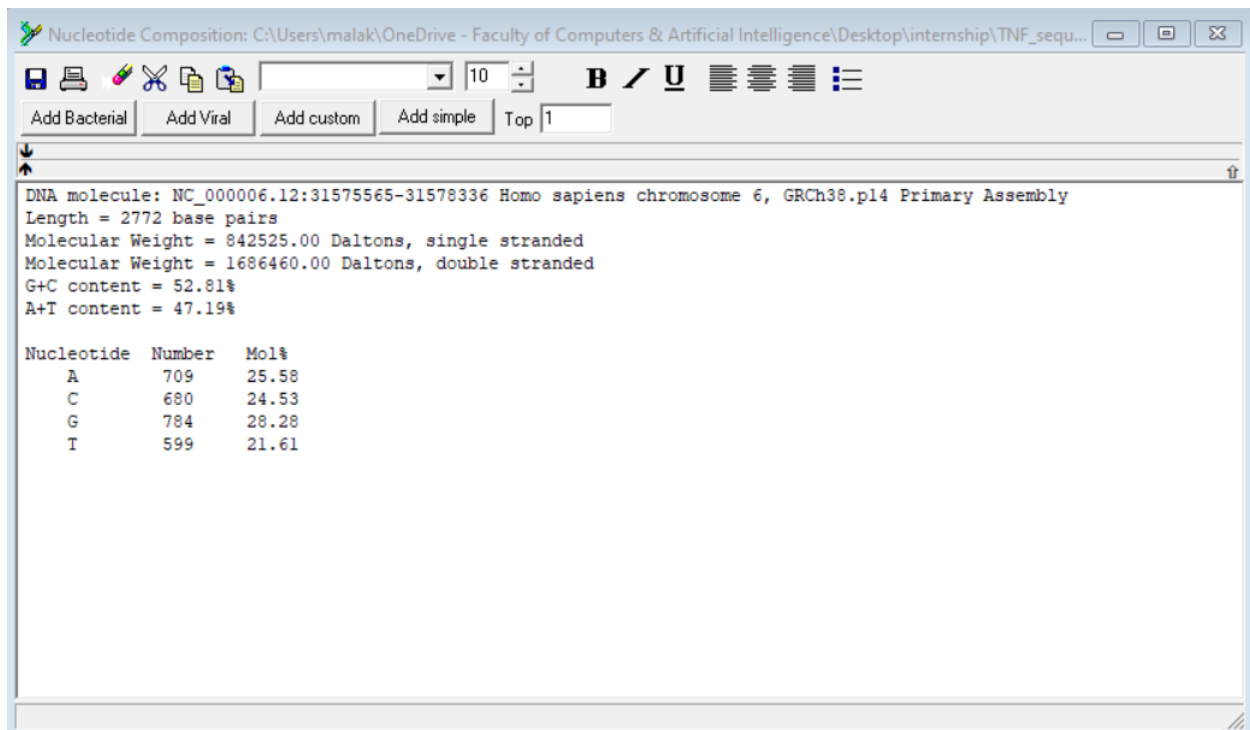
The screenshot shows a web-based tool titled "Six-frame translation from C:\Users\malak\OneDrive - Faculty of Computers & Artificial Intelligence\Desktop\internship\TNF\_sequence\sequence.fasta". The tool displays the results of a six-frame translation of a DNA sequence. The results are organized into several sections, each representing a different reading frame. Each section includes the frame number, the amino acid sequence, and the corresponding nucleotide positions. The sequences are shown in a monospaced font, and the tool includes a search bar and a "Top" button for navigation.

```
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 178 to 606 (178 to 606):
Frame 1 143 aa
MSTESMIRDVLAEEALPKKTGGPQGSRRCLFLSLFSLIVAGATTLFCLLHFGVIGPQREEVSAWPAFIHSPQTQEMETQERERDGMGERCALIGRDGEKKTWRKTGMQKEMWQEMGKRERERWRDRMSGTW
KVLTKCVWSE
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 450 to 587 (450 to 587):
Frame 3 46 aa
MCADREGWREIQVVEKDGDAERDVARDGEERERKMERQDVWHMEGAH
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 470 to 610 (470 to 610):
Frame 2 47 aa
MERKKRGERRGCRKRCGKRWGREKDKGTGCLAHGRCSLSVYGVNE
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 659 to 781 (659 to 781):
Frame 2 41 aa
MWGVRRREMGEETSDMNKDGETERAGNMTAKERDGGDKERRR
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 756 to 923 (756 to 923):
Frame 3 56 aa
MGEIRREEDRVSGTQKTLRERAVECLEGEYTDENREKIRHLRAKSAGQTGSOLFLL
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 850 to 1134 (850 to
1134): Frame 1 95 aa
MNGERKPDITSLRAQARQAASCSFSGKDSLDVNHSPSPQFPRLDLSLISPLAQAVSKCLQTSFLILGLGLGVLPVWVKQWQKFKVLVLGEDGWR
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 1124 to 1288 (1124 to
1288): Frame 2 55 aa
MDGGESRGVFSRKFGLSFFFLSSSSGSSSRTPSDKPFVAHVVGKSSSDVSWNLEG
>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly Positions 1 to 2772: 1128 to 1235 (1128 to
1235): Frame 1 44 aa
MSTESMIRDVLAEEALPKKTGGPQGSRRCLFLSLFSLIVAGATTLFCLLHFGVIGPQREEVSAWPAFIHSPQTQEMETQERERDGMGERCALIGRDGEKKTWRKTGMQKEMWQEMGKRERERWRDRMSGTW
KVLTKCVWSE
```

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

**Output:**

- High GC content often indicates stability and functional importance.
- Low GC content might suggest a different function or evolutionary history.
- Compare these results to known patterns in other genes or species.
- Consider any potential implications for gene expression, regulation, or protein structure.



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

## Output:

The analysis of the TNF gene sequence, i identified several potential Open Reading Frames (ORFs), each characterized by specific start and stop positions. These ORFs represent candidate protein-coding regions within the gene.

[Save link (only for 7 days): [https://algggen.lsi.upc.es/cgi-bin/promo\\_v3/promo/promo.cgi?dirDB=TF\\_8.3&idCon=172435770900&getFile=resumSearchRes.html](https://algggen.lsi.upc.es/cgi-bin/promo_v3/promo/promo.cgi?dirDB=TF_8.3&idCon=172435770900&getFile=resumSearchRes.html)]

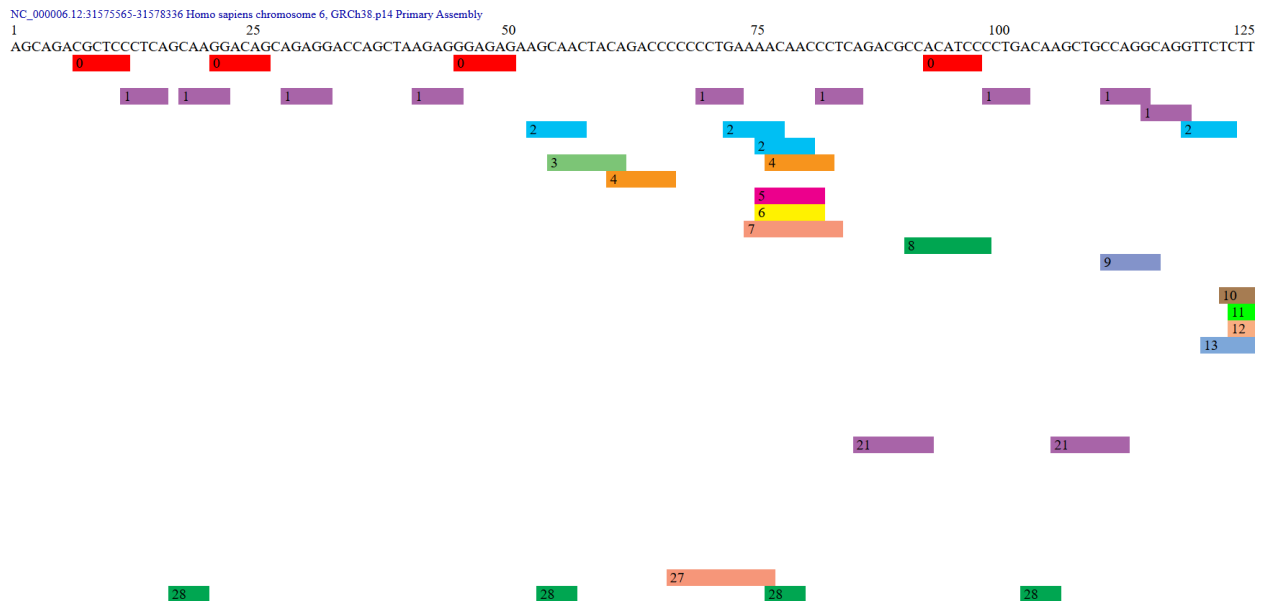
Factors predicted within a dissimilarity margin less or equal than 15 % :

1	TFII-I [T00824]	1	GR-alpha [T00337]	2	FOXp3 [T04280]	3	c-Myb [T00137]	4	RXR-alpha [T01345]	5	PR B [T00696]	6	PRA [T01661]	7	RAR-beta [T00721]
8	PEA3 [T00685]	9	AP-2alphaA [T00035]	10	STAT4 [T01577]	11	c-Ets-1 [T00112]	12	Elk-1 [T00250]	13	PU.1 [T02068]	14	ER-alpha [T00261]	15	RAR-alpha1 [T00719]
16	Egr-3 [T00243]	17	IRF-1 [T00423]	18	NF-AT1 [T00550]	19	YY1 [T00915]	20	XBP-1 [T00902]	21	ENKTF-1 [T00255]	22	NF-kappaB1 [T00593]	23	NFICTF [T00094]
24	Pax-5 [T00070]	25	p53 [T00671]	26	PPAR-alpha:RXR-alpha [T05221]	27	EBF [T05427]	28	C/EBPbeta [T00581]	29	HNF-4alpha [T03828]	30	LEF-1 [T02905]	31	NF-1 [T00539]
32	GR-beta [T01920]	33	VDR [T00885]	34	PXR-1:RXR-alpha [T05671]	35	TBP [T00794]	36	MEF-2A [T01005]	37	HNF-3alpha [T02512]	38	HOXD9 [T01424]	39	HOXD10 [T01425]
40	IRF-2 [T01491]	41	GATA-3 [T00311]	42	HNF-1A [T00368]	43	ETf [T00270]	44	AP-1 [T00029]	45	c-Jun [T00133]	46	GR [T05076]	47	NF-AT2 [T01945]
48	STAT1beta [T01573]	49	c-Ets-2 [T00113]	50	C/EBPalpha [T00105]	51	CREB [T00163]	52	ATF-2 [T00167]	53	ATF [T00051]	54	MAZ [T00490]	55	E2F-1 [T01542]
56	Ik-i [T02702]	57	T3R-beta1 [T00851]	58	GATA-1 [T00306]	59	GCF [T00320]	60	TFIID [T00820]	61	NF-Y [T00150]	62	TCF-4 [T02918]	63	ATF3 [T01313]
64	NF-AT1 [T01948]	65	ELF-1 [T01113]	66	NF-kappaB [T00590]	67	RelA [T00594]	68	RBP-Jkappa [T01616]	69	SKY [T00997]	70	TCF-4E [T02878]	71	AhR:Arnt [T05394]
72	CTF [T00174]	73	c-Fos [T00123]	74	AR [T00040]	75	USF2 [T00878]	76	STAT5A [T04683]	77	HNF-1C [T01951]	78	HNF-1B [T01950]	79	GATA-2 [T00308]
80	E2F [T00221]														

Zoom Data (txt)

	1	10	20	30	40	50	60	70	80	90	100	110	120	130	140	150																									
NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly	0	1	28	0	1	74	0	1	2	3	4	28	1	27	2	4	6	1	8	21	80	0	1	45	1	9	21	28	1	2	13	1	10	11	12	4	14	15	45	4	16

Distribution of the nucleotides over the given chain:



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

### Output:

In the exploration of the TNF gene sequence using the MEME Suite, uncovered intriguing motifs—those elusive snippets of genetic choreography. These motifs, akin to musical motifs in a symphony, might hold the key to TNF's regulatory dance.

#### 1. Motif Discovery:

- MEME Suite diligently scanned the sequence, spotlighting recurring patterns.
- Each motif—like a soloist on stage—could signify a transcription factor binding site or a functional element.
- We eagerly awaited their unveiling.

#### 2. Functional Significance:

- Consider the context:
  - Are these motifs conserved across species?
  - Do they align with known regulatory elements?
- Perhaps they're the secret cues that orchestrate TNF's expression.

#### 3. Potential Roles:

- Some motifs might be backstage passcodes—allowing specific factors to access TNF's promoter.
- Others could be enhancer invitations—beckoning regulatory proteins to join the performance.
- We speculate on their roles: Are they conductors, soloists, or harmonizers?

#### 4. Saving the Encore:

- Document these motifs—their positions, sequences, and significance.
- Perhaps they'll inspire future studies, unraveling TNF's genetic symphony further.

	Logo	E-value	Sites	Width	More	Submit/Download
1.		2.6e+000	5	38	<a href="#">↓</a>	<a href="#">→</a>
2.		1.0e+003	5	32	<a href="#">↓</a>	<a href="#">→</a>
3.		1.0e+003	4	11	<a href="#">↓</a>	<a href="#">→</a>

Stopped because requested number of motifs (3) found.

☒ Only Motif Sites [?](#)
☐ Motif Sites+Scanned Sites [?](#)
☐ All Sequences [?](#)
[Download PDF](#) [?](#)
[Download SVG](#) [?](#)

Name <a href="#">?</a>	p-value	Motif Locations <a href="#">?</a>
1. NC_000006.12:31575565-31578336	9.73e-23	

Sequences

Role	Source	Alphabet	Sequence Count	Total Size
Primary Sequences	sequence.fasta	DNA	1	2772

Background Model

Source: built from the (primary) sequences

Order: 0

Name	Freq.	Bg.	A	T	Bg.	Freq.	Name
Adenine	0.236	0.236	A	T	0.236	0.236	Thymine
Cytosine	0.264	0.264	C	G	0.264	0.264	Guanine

Other Settings

Motif Site Distribution

Objective Function

Starting Point Function

Site Strand Handling

Maximum Number of Motifs

Motif E-value Threshold

Minimum Motif Width

Maximum Motif Width

Minimum Sites per Motif

Maximum Sites per Motif

ANK: Any number of sites per sequence

E-value of product of p-values

E-value of product of p-values

Sites may be on either strand

3

no limit

6

50

2

5

Show Advanced Settings

1. **E-values:** 2.6e+000 **Site Count:** 5 **Width:** 38

Standard Reverse Complement

**Log Likelihood Ratio:** 154 **Information Content:** 43.9 **Relative Entropy:** 44.4 **Bayes Threshold:** 9.09276

Name	Strand	Start	p-value	Sites
1. NC_000006.12:31575565-31578336	+	413	2.69e-17	GGGGAAATGG AAGACCAAGATGGGAAATATGGATGGGTAAAG TGTGCGCTGA
1. NC_000006.12:31575565-31578336	+	638	2.68e-16	AAGCAGATAT ATAAATAATATGAGACATATGGGGTTTAAAG AGAGATGGGG
1. NC_000006.12:31575565-31578336	+	735	1.66e-14	AGAGCGGGAA ATATACCTAAGAGACATATGGGGATAAAG AGAAGAGAGG
1. NC_000006.12:31575565-31578336	+	690	4.56e-14	ATGGGGGAG AAACCAATATATTAATAAAATGGTAAACAAAGAG GCGGGAAATA
1. NC_000006.12:31575565-31578336	+	524	1.28e-13	ATGTGGCAAG AATGGGGAACAGAGAGAGAAATGAGACAGG ATGTCTGGCA

2. **Logo** **E-value** 1.0e+003 **Sites** 5 **Width** 32 **More** **Submit/Download**

3. **Logo** **E-value** 1.0e+003 **Sites** 4 **Width** 11 **More** **Submit/Download**

Stopped because requested number of motifs (3) found.

## Task 7: Predict Coding/Non-Coding Regions in a Genome Using GENSCAN

### Output:

GENSCAN analysis of the TNF gene sequence, we identified coding exons—segments encoding proteins—and non-coding introns. These regions together compose the genetic symphony of TNF.

```
GENSCAN Output

View gene model output: PS | PDF

GENSCAN 1.0   Date run: 22-Aug-124   Time: 16:28:06

Sequence /tmp/08_22_24-16:28:05.fasta : 2815 bp : 52.82% C+G : Isochore 3 (51 - 57 C+G%)

Parameter matrix: HumanIso.smat

Predicted genes/exons:

Gn.Ex Type S .Begin ...End .Len Fr Ph I/Ac Do/T CodRg P.... Tscr...
-----
1.01 Init + 221 406 186 1 0 94 105 207 0.703 22.03
1.02 Intr + 1013 1058 46 1 1 106 89 4 0.929 0.77
1.03 Intr + 1246 1293 48 2 0 140 82 25 0.987 6.24
1.04 Term + 1595 2016 422 0 2 132 55 518 0.985 48.73
1.05 PolyA + 2792 2797 6 1.05

Suboptimal exons with probability > 1.000
```

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

- fasta format

```
File Edit View

>NC_000006.12:31575565-31578336 Homo sapiens chromosome 6, GRCh38.p14 Primary Assembly
AGCAGACGCTCCCTCAGCAAGGACAGCAGAGGACAGCTAAGAGGGAGAGAAGCAACTAC
AGACCCCTGAAACACACCTCAGACGCCACATCCCTGACAAGCTGCCAGGCAGGTT
CTCTTCTCTCACATACTGACCCACGGCTCCACCTCTCTCCCTGGAAGGACACCATG
AGCACTGAAAGCATGATCCGGGACGTGGAGCTGGCCGAGGAGGCGCTCCCAAGAAGACA
GGGGGGCCCCAGGGCTCCAGGCGGTGCTTGTCTCAGCCTTCTCTCTCTGATCGTG
GCAGGCGCCACCACGCTCTTCTGCTGCTGCACTTTGGAGTGATCGGCCCCAGAGGGAA
GAGGTGAGTGCTGCCAGCCTTCATCCACTCTCCACCCAAGGGGAAATGGAGCGCAA
GAGAGGGAGAGAGATGGGATGGGTGAAAGATGTGCGTGATAGGAGGGATGGAGAGAAA
AAAACGTGGAGAAAGACGGGGATGCAGAAAGAGATGTGGCAAGAGATGGGAAGAGAGAG
AGAGAAAGATGGAGAGACAGGATGTCTGGCAGATGGAAGGTGCTCACTAAGTGTGTATGG
AGTGAATGAATGAATGAATGAATGAACAAGCAGATATATAAATAAGATATGGAGACAGAT
GTGGGTGTGAGAAGAGAGATGGGGGAAGAAACAAGTGATATGAATAAGATGGTGAGAC
AGAAAGAGCGGGAAATATGACAGCTAAGGAGAGAGATGGGGGAGATAAGGAGAGAAGAG
ATAGGTGTCTGGCACACAGAAGACACTCAGGGAAGAGCTGTTGAATGCTGGAAGGTG
AATACACAGATGAATGGAGAGAGAAACAGACACCTCAGGGCTAAGAGCGCAGGCGAGA
CAGGCAGCCAGCTGTTCTCTTTAAGGGTGACTCCCTCGATGTTAACCATTTCTCTTCT
CCCCAACAGTTCCCAGGGACCTCTCTAATCAGCCCTCTGGCCAGGCAGTCAGTAAG
TGTTCCAAACCTCTTTCTAATCTGGGTTTGGGTTTGGGGTAGGGTAGTACCGGTA
TGGAAAGCAGTGGGGGAAATTTAAAGTTTGGTCTTGGGGAGGATGGATGGAGGTGAAG
TAGGGGGGTATTTCTAGGAAGTTAAGGGTCTCAGCTTTTCTTTCTCTCTCTCTC
AGGATCATCTTCTCGAACCCGAGTGACAAGCCTGTAGCCCATGTTGTAGTAAGAGCTC
TGAGGATGTGCTTGGAACTTGGAGGGCTAGGATTTGGGGATTGAAGCCCGGTGATGGT
AGGCAGAACTTGGAGACAATGTGAGAAGGACTCGCTGAGCTCAAGGGAAGGGTGGAGAA
CAGCACAGGCCTTAGTGGGATACTCAGAACGTCATGGCCAGGTGGGATGTGGGATGACAG
ACAGAGAGGACAGGAACCGGATGTGGGTGGGCAGAGCTCGAGGGCAGGATGTGGAGAG
-----
Ln 1, Col 1      2,906 characters      100%      Windows (CRLF)      UTF-8
```



- phylip format :

```
File Edit View
1 2772 I
NC_000006. AGCAGACGCT CCCTCAGCAA GGACAGCAGA GGACCAGCTA AGAGGGAGAG
AAGCAACTAC AGACCCCCC TGAAAACAAC CCTCAGACGC CACATCCCCT
GACAAGCTGC CAGGCAGGTT CTCTTCTCT CACATACTGA CCCACGGCTC
CACCTCTCT CCCCTGAAA GGACACCATG AGCACTGAAA GCATGATCCG
GGACGTGGAG CTGGCCGAGG AGGCGCTCCC CAAGAAGACA GGGGGGCCCC
AGGGCTCCAG GCGGTGCTTG TTCCTCAGCC TCTTCTCCTT CCTGATCGTG
GCAGGCGCCA CCACGCTCTT CTGCCTGCTG CACTTTGGAG TGATCGGCCC
CCAGAGGGA GAGGTGAGTG CCTGGCCAGC CTTATCCAC TCTCCACCC
AAGGGGAAAT GGAGACGCAA GAGAGGGAGA GAGATGGGAT GGGTGAAAGA
TGTGCGCTGA TAGGGAGGGA TGGAGAGAAA AAAACGTGGA GAAAGACGGG
GATGCAGAAA GAGATGTGGC AAGAGATGGG GAAGAGAGAG AGAGAAAGAT
GGAGAGACAG GATGTCTGGC ACATGGAAGG TGCTCACTAA GTGTGTATGG
AGTGAATGAA TGAATGAATG AATGAACAAG CAGATATATA AATAAGATAT
GGAGACAGAT GTGGGGTGTG AGAAGAGAGA TGGGGGAAGA AACAAGTGAT
ATGAATAAAG ATGGTGAGAC AGAAAGAGCG GGAATATGA CAGCTAAGGA
GAGAGATGGG GGAGATAAGG AGAGAAGAAG ATAGGGTGTC TGGCACACAG
AAGACACTCA GGGAAAGAGC TGTGAATGC CTGGAAGGTG AATACACAGA
TGAATGGAGA GAGAAAACCA GACACCTCAG GGCTAAGAGC GCAGGCCAGA
CAGGCAGCCA GCTGTTCTC CTTTAAGGGT GACTCCCTCG ATGTTAACCA
TTCTCTTCT CCCCAACAGT TCCCCAGGGA CCTCTCTCTA ATCAGCCTC
TGGCCAGGC AGTCAGTAAG TGTCTCCAAA CCTTTTCTT AATTCTGGGT
TTGGGTTTGG GGGTAGGGTT AGTACCGTA TGGAAAGCAGT GGGGGAAATT
TAAAGTTTGG GTCTTGGGGG AGGATGGATG GAGGTGAAAG TAGGGGGGTA
TTTTTAGGA AGTTTAAGGG TCTCAGCTTT TTCTTTTCTC TCTCTCTTC
AGGATCATCT TCTCGAACC CGAGTGACAA GCCTGTAGCC CATGTTGTAG
-----
Ln 1, Col 1 3,788 characters 100% Windows (CRLF)
```