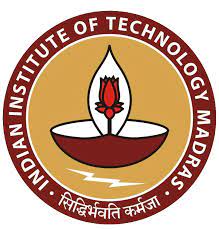
..

**BT 3040: Bioinformatics**

**Assignment 1**



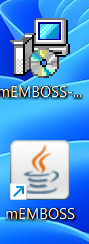
Indian Institute of Technology Madras

Atharva Mandar Phatak | BE21B009

Department of Biotechnology

**Q1) Download the EMBOSS package (http://emboss.sourceforge.net/download/) and copy to your Windows system. In case of Linux use the command: sudo apt-get install jemboss. For Mac users, use the online tool links given in the instructions doc.**

Installed Emboss following the instructions, on Windows 11 PC. Created Shortcuts in Desktop



## **Q2) Using EMBOSS, find the complementary strand for the sequence: CTCGGATTTGTAAAGATCATGATCTCATACATAGTACCTAGCCATTG**

Used Reverse SQ to find the complementary strand. We can get the non-reversed output by deselecting the respective box under ‘Advance Options’

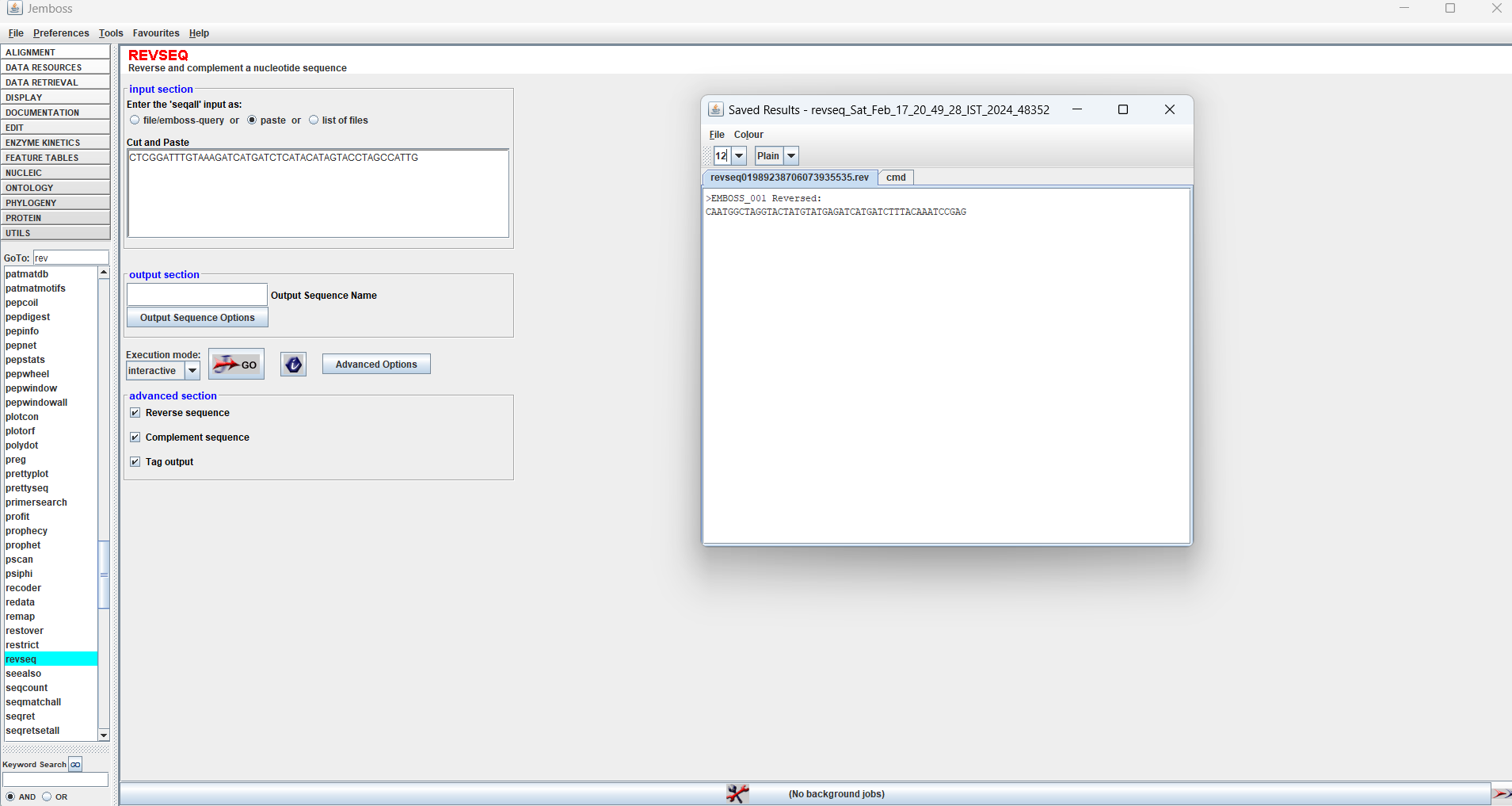
Outputs received:

>EMBOSS\_001

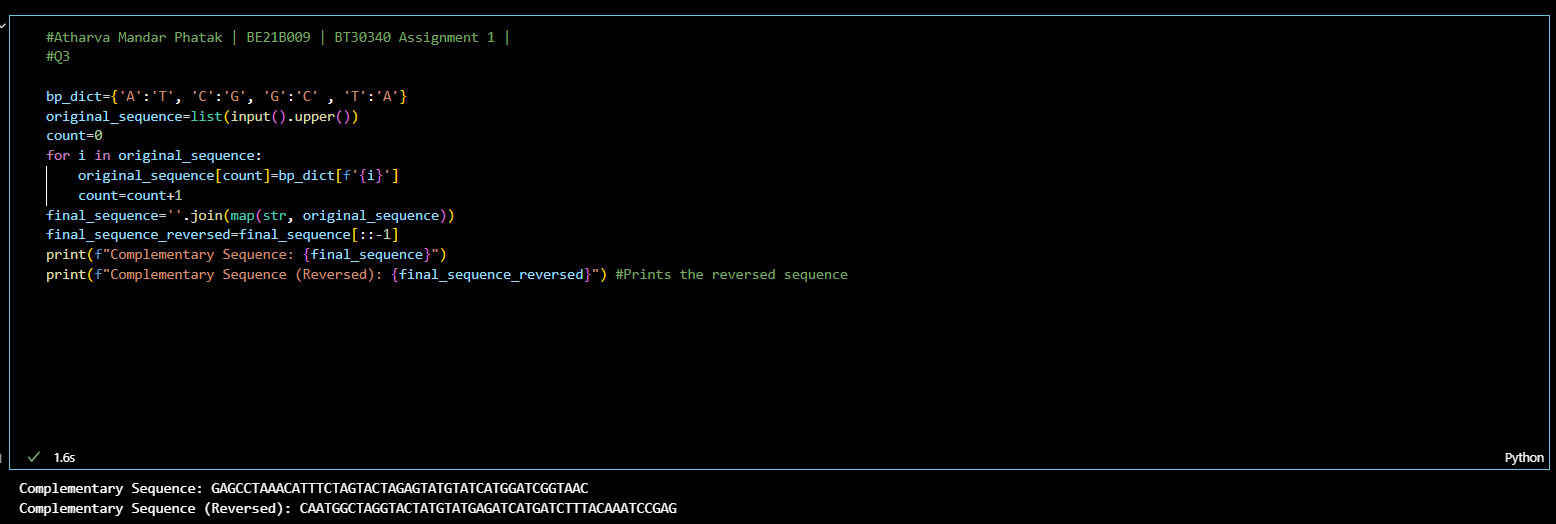
GAGCCTAAACATTTCTAGTACTAGAGTATGTATCATGGATCGGTAAC

>EMBOSS\_001 Reversed:

CAATGGCTAGGTACTATGTATGAGATCATGATCTTTACAAATCCGAG



## **Q3) Write a program to find the complementary strand for the sequence given in Q2.**



## **Q4) (i)Using EMBOSS, find the protein sequence for GACATTGTGAACAGTAAAAAAGTCCATGCAATGCGCAAGGAGCAGAAGAGGAAGCAGGGCAAGCAGCGCTCCATGGGCTCTCCCATGGACTACTCTCCTCTGCCCATCGACAAGCATGAGCCTGAATTTGGT CCATGCAGAAGAAAACTGGATGGG**

## **(ii) Identify the reading frame equivalent to the following sequence. PIQFSSAWTKFRLMLVDGQRRVVHGRAHGALLALLPLLLLAHCMDFFTVHNV**

1. Protein sequence Using Emboss. In Advance options select ‘All Six’ to get all the frames

>\_1

DIVNSKKVHAMRKEQKRKQGKQRSMGSPMDYSPLPIDKHEPEFGPCRRKLDG

>\_2

TL\*TVKKSMQCARSRRGSRASSAPWALPWTTLLCPSTSMSLNLVHAEENWMG

>\_3

HCEQ\*KSPCNAQGAEEEAGQAALHGLSHGLLSSAHRQA\*A\*IWSMQKKTGWX

>\_4

PIQFSSAWTKFRLMLVDGQRRVVHGRAHGALLALLPLLLLAHCMDFFTVHNV

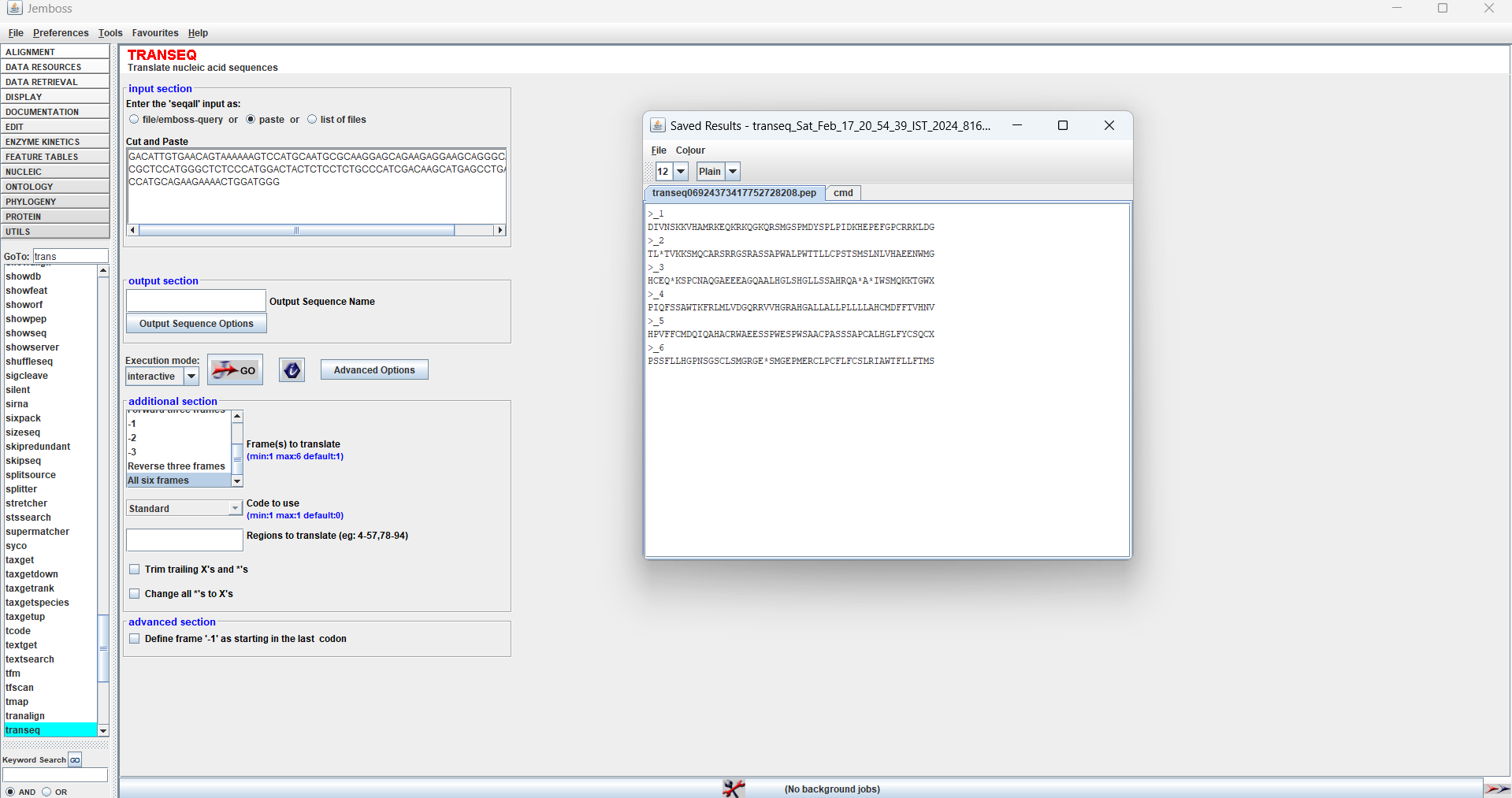
>\_5

HPVFFCMDQIQAHACRWAEESSPWESPWSAACPASSSAPCALHGLFYCSQCX

>\_6

PSSFLLHGPNSGSCLSMGRGE\*SMGEPMERCLPCFLFCSLRIAWTFLLFTMS

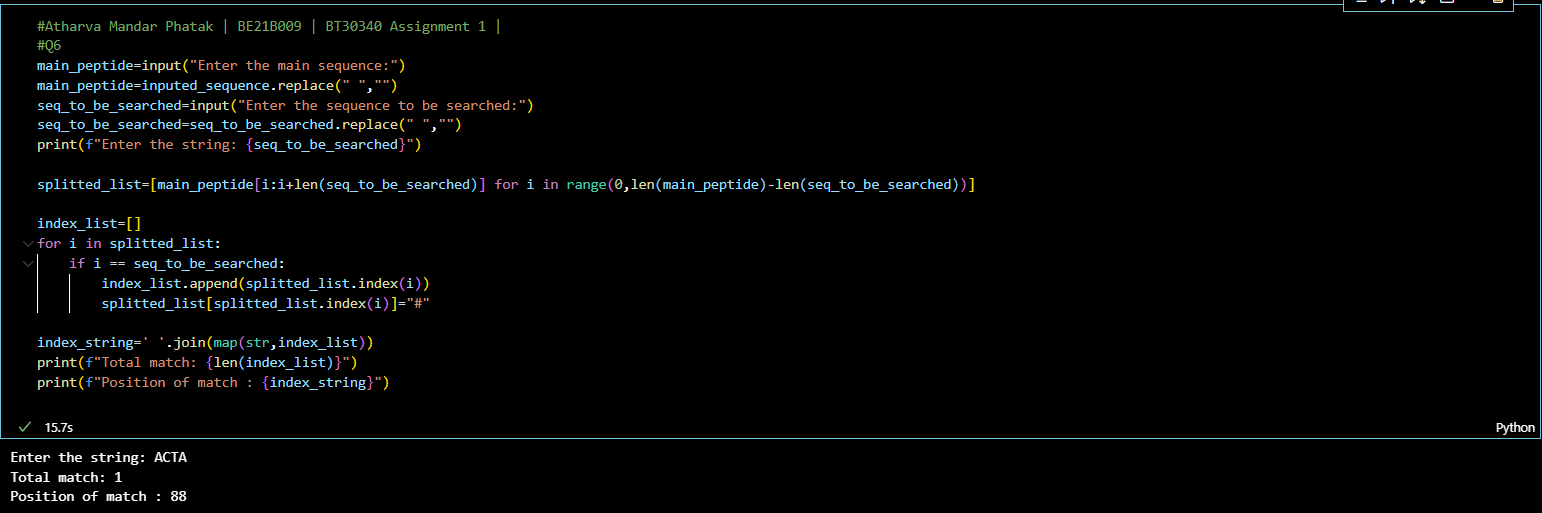
1. The reading frame equivalent for the given sequence was found to be the **fourth sequence**. Highlighted in yellow above.



## **Q5) Write a program to translate the given DNA sequence (refer Q4) to protein sequence.**



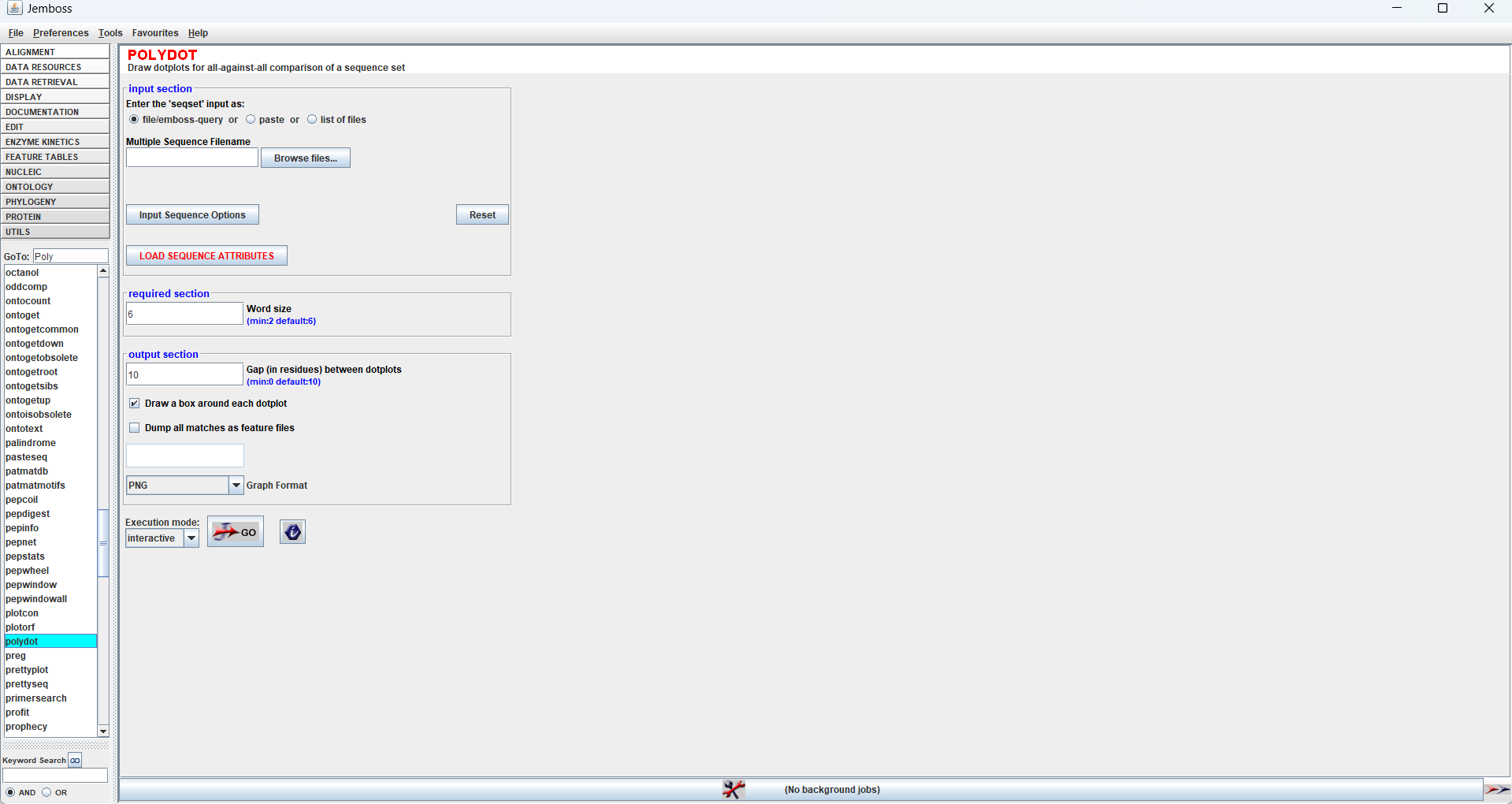
## **Q6) Write a code to search for the following strings ‘AAG’, ‘GTC’, ‘GAG, ‘ACTA’, and ‘ATAT’ in the DNA sequence provided in Q4. The program should print the total number of matches for the each of the given strings and the start positions of the matches.**



## **Q7) Familiarize with other applications in EMBOSS. For example, melting temperature, bending, curvature etc.**

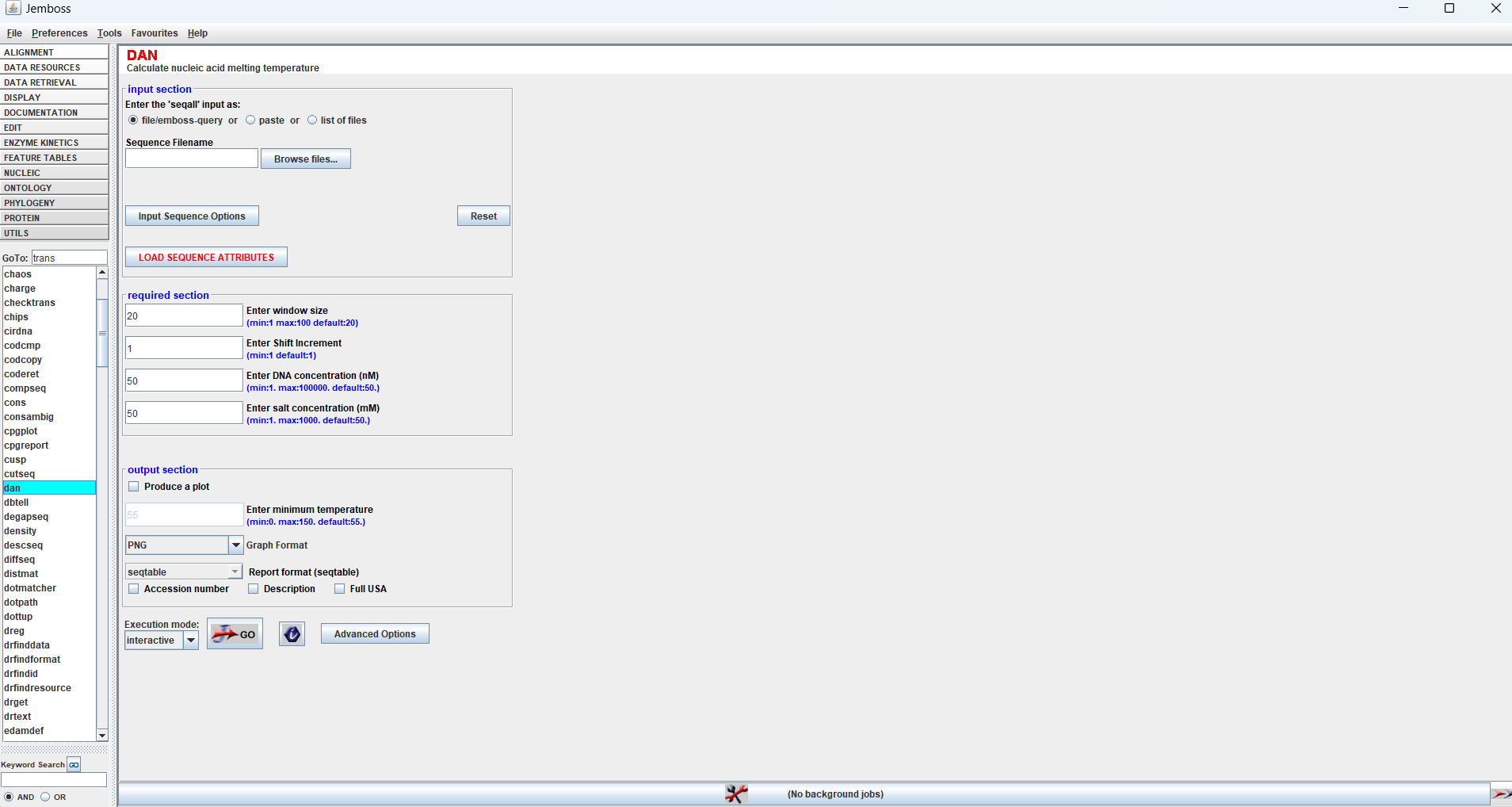
### **POLYDOT**

Draws dot plot for all-against-all comparison of sequence set



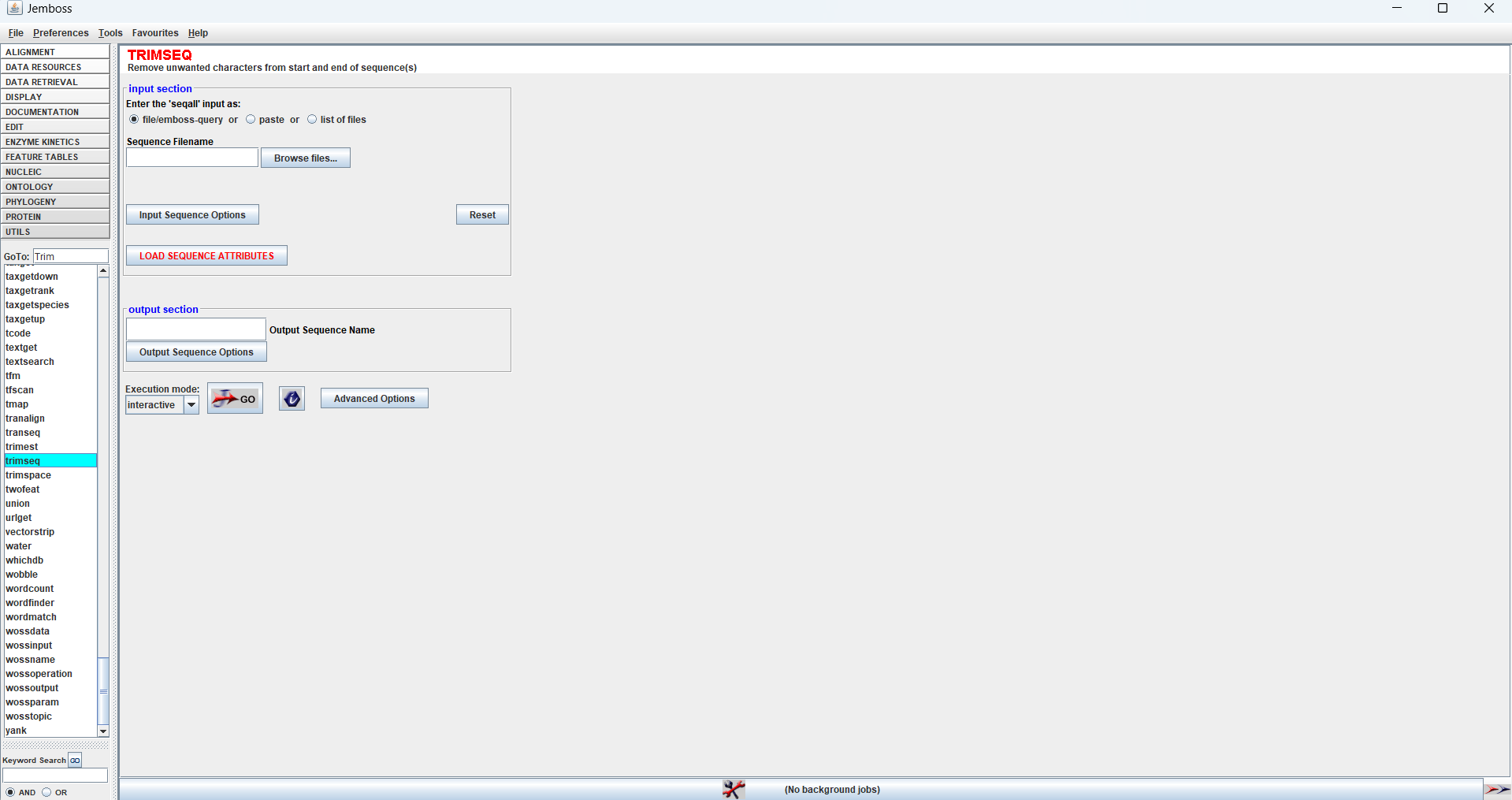
### **DAN**

Calculates nucleic acid melting temperature



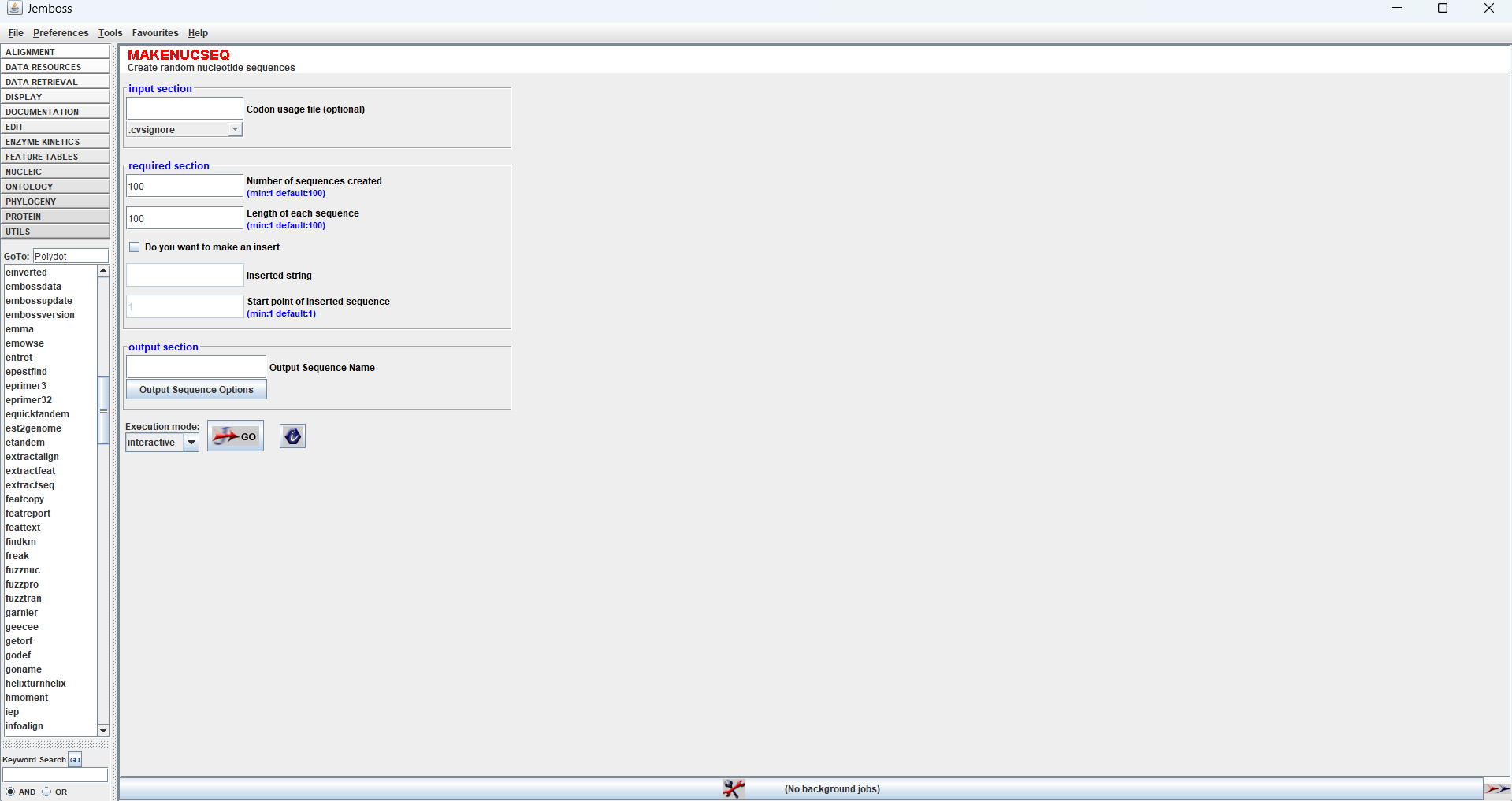
### **TRIMSEQ**

Removes unwanted characters from start to end.



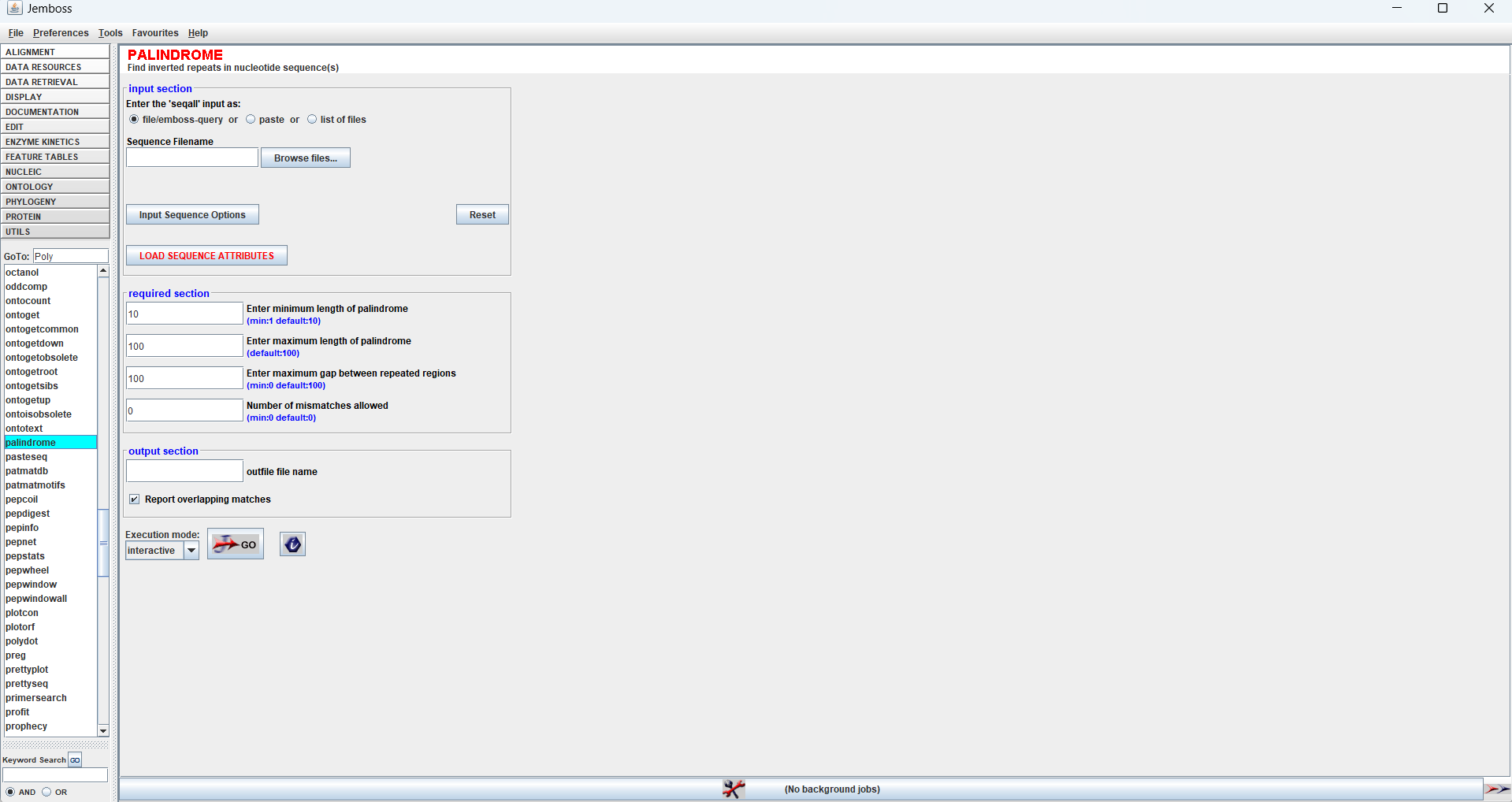
### **MAKEUCESQ**

Creates random nucleotide sequences.

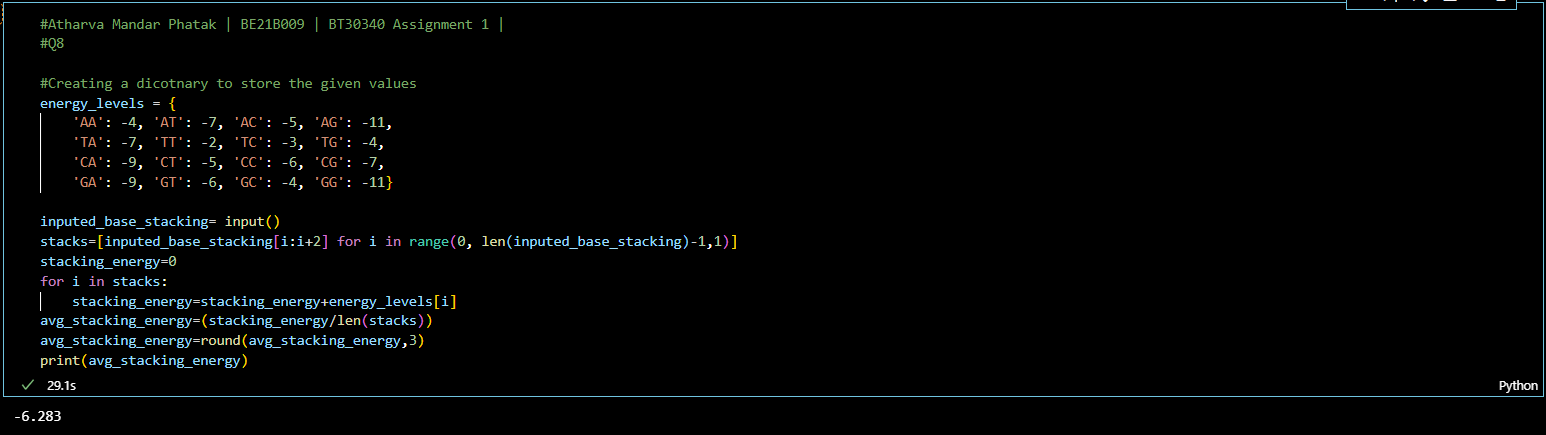


### **PALINDROME**

Finds Palindromes, inverted repeats in the nucleotide sequence(s)

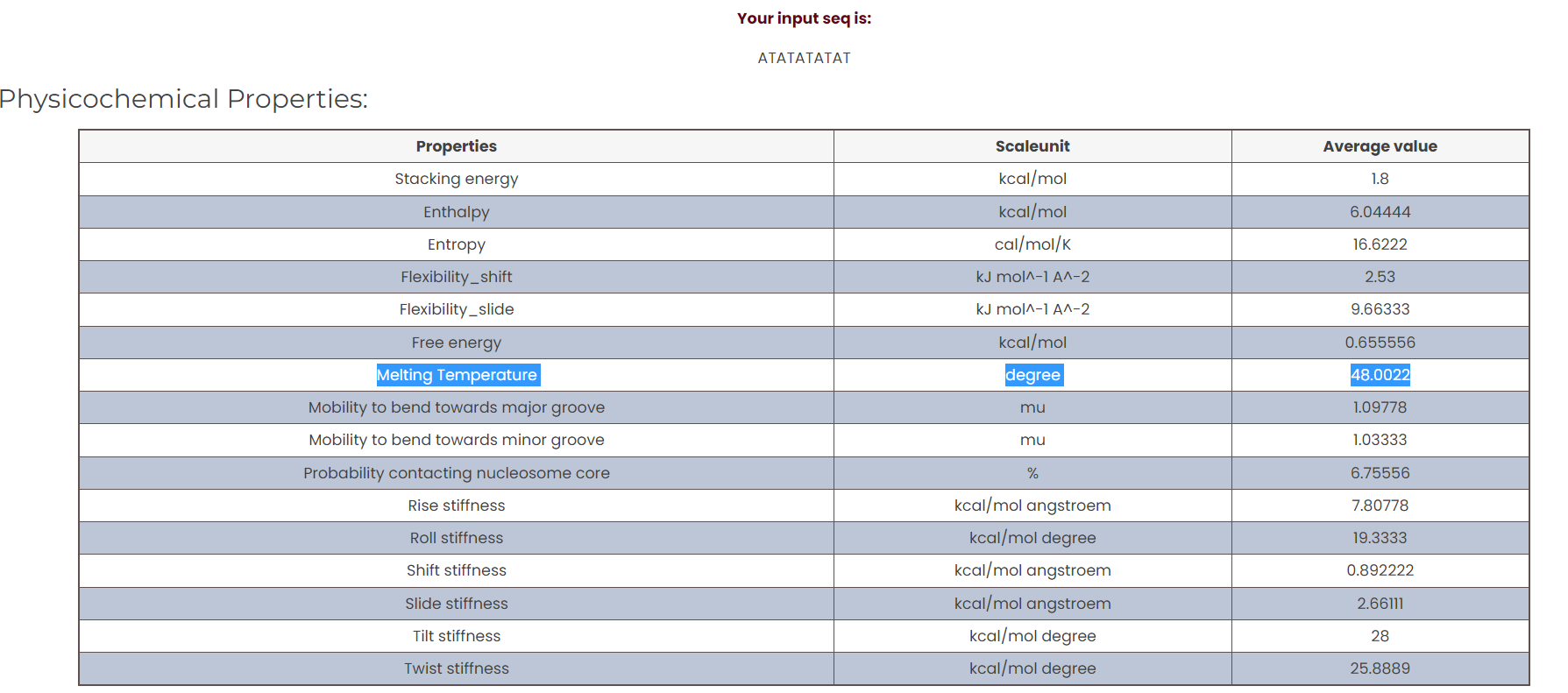


## **Q8) Write a program to compute the average base stacking energy for the sequence given in Q2 (AA: - 4; AT: -7; AC: -5; AG: -11; TA: -7; TT: -2; TC: -3; TG: -4; CA: -9; CT: -5; CC: -6; CG: -7; GA: - 9; GT: -6; GC: -4; GG: -11).**

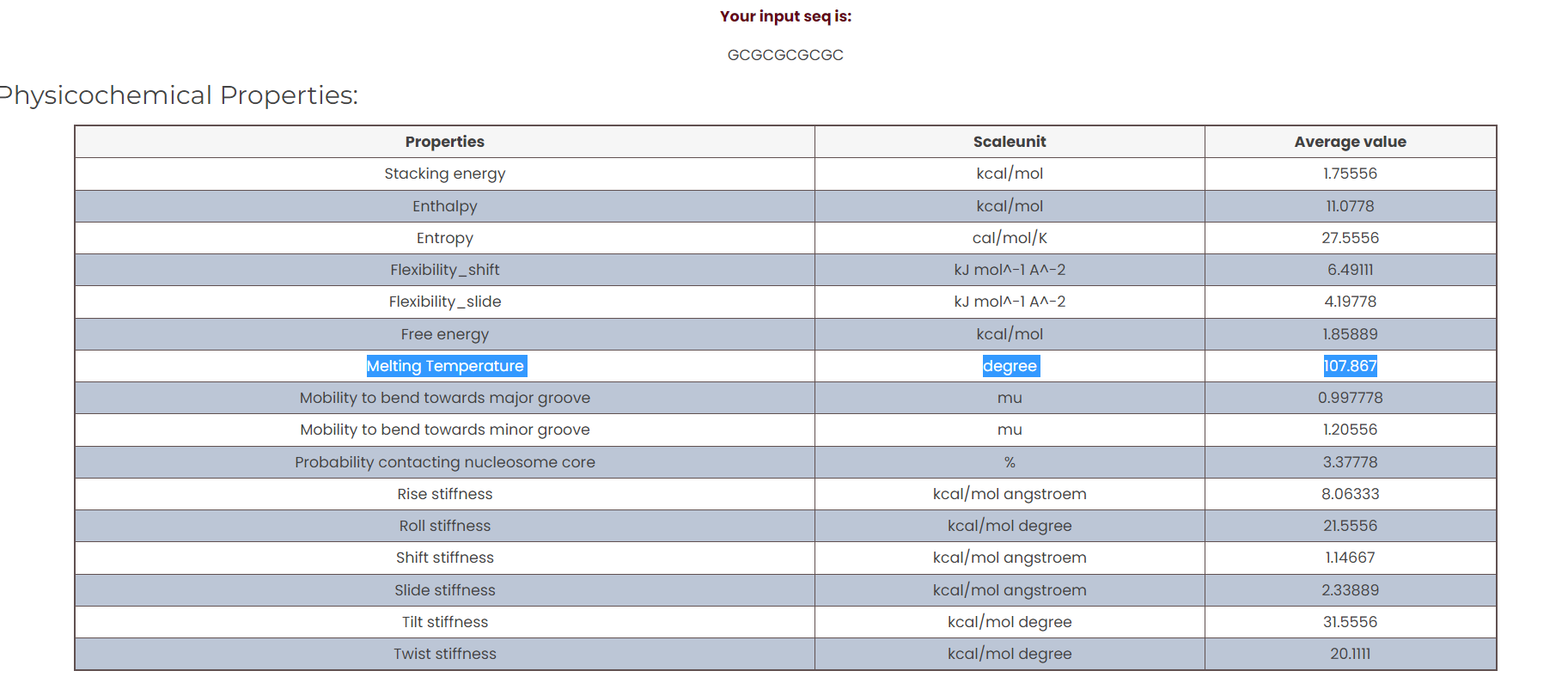


## **Q9) Compute the average melting temperature of the following sequences using Seq2Feature tool ( https://www.iitm.ac.in/bioinfo/SBFE/index.html ) and comment on the results (Enter one sequence at a time in fasta format) (i) ATATATATAT ii) GCGCGCGCGC**

i) Average Melting Temperature: 48.0022o C   
ii) Average Melting Temperature: 107.867o C

(i) 

ii)



Thus i) sequence shows lower melting point as compared to the second sequence.   
**Reason** : G-C pairs have a higher stacking energy than A-T pairs because they make three hydrogen bonds in water whereas AT only forms two hydrogen bonds resulting into GC pair have high melting point. Thus, the sequence with more GC content will have a higher melting point.

## **Q10) Calculate the AT and GC content of the sequence AAATGGCCCTAA using Seq2Feature too**

AT Content = 5.3334 %

GC Content = 41.6667 %

### 

