

TRABAJO DE PYTHON

Este trabajo cuenta el 30% de la nota.

Resuelve los siguientes ejercicios en Python

Utiliza tantas celdas de código como necesites

1.- Busca la secuencia del gen CNTNAP2 (cromosma 7) del genoma humano. Descarga la secuencia en formato FASTA e introduce la en este Notebook . (1 pto)

```
In [1]: archivo = open("sequence.fasta")
```

```
In [2]: archivo
```

```
Out[2]: <_io.TextIOWrapper name='sequence.fasta' mode='r' encoding='UTF-8'>
```

2.- Muestre el encabezado y la secuencia y guarde cada uno en variables separadas. Elimine el elemento "\n" del encabezado con replace() si es que lo tiene. ¿Cuál es la longitud de la secuencia?. (1 pto)

```
In [3]: contenido = archivo.readlines()
```

```
In [4]: contenido
```

Out[4]:

```
'>NM_014141.6 Homo sapiens contactin associated protein 2 (CNTNAP2), mRNA\n',
'CTTCAAGAACCTACGGAGAGTCGGACTGCATCTCCGAGCGAGCTCTGGAGCGCCGCCGGGAGG\n',
'CGAAGGATGCAGGCCGCTCCGGCGCCGGCTGCAGCAGCTCTGCTGTGGATTGTAGCAGCTGCC\n',
'TCTGCAGAGCCTGGACGGCTCCCTCACGTCCAAAATGTGATGAGCCACTGTCTGGACTCCCCA\n',
'TGTGGTTTCAGCAGCTCCTCCATCTGGTAGCTATTCTCCGGCTATGCCAAGATAAAACAAGAGA\n',
'GGAGGTGCTGGGGATGGTCTCCATCAGACAGCAGCATTATCAATGGCTCAGGTTGACTTGCAATC\n',
'GGAAGCAGATCAGTGCATTGCAACCCAAGGAAGGTAGCAGCTCAGATTGGGTGACCCAATACCGGAT\n',
'GCTCTACAGCAGCACAGGGAGAACTGGAAACCTATCATCAAGATGGGATATCTGGCATTCCCAGG\n',
'AACATTAACCTGACGGTGTGGTCCGGCACGAATTACAGCATCCGATTATTGCCGCTATGTGCGCATAG\n',
'TGCCTCTGGATTGGAATGGAGAAGGTGCATTGGACTCAGAATTGAAAGTTATGGCTGTTACTGGG\n',
'TGATGTTATCAACTTGTGGCATGGTATTACCATATAGATTAGAAACAAGAAGATGAAAACACTG\n',
'AAAGATGTCAATTGCTTGAACCTTAAGACGTCTGAAAGTGAAGGAGTAATCTGCACGGAGAAGGACAGC\n',
'AAGGAGATTACATTACCTTGGAACTGAAAAAAAGCCAAGCTGGTCTCAGTTAAACTTAGGAAGCAACCA\n',
'GCTTGGCCCCATATATGCCACACATCAGTGATGACAGGAAGTTGCTGGATGACCACACTGGCACTCT\n',
'GTGGTCATTGAGCAGGCCAGGGGGAGCATTAAACCTCACTCTGGACAGGAGCATGCAGCACTCCGTACCA\n',
'ATGGAGAGTTTGAACCTGGACTTGGACTATGAGATAACCTTGGAGGCATCCCTTCTGGCAAGCC\n',
'CAGCTCCAGCAGTAGAAAGAATTCAAAGGCTGCATGGAAAGCATCAACTACAATGGCGTCAACATTACT\n',
'GATCTTGGCAGAAGAAGAAATTAGAGCCTCAAATGTGGAAATTGAGCTTTCTGTGGAACCCCT\n',
'ATACGGTGCCTGTCTTTCAACGCTACAAGTTACCTGGAGGTGCCGGACGGCTAACCAGGACCTGTT\n',
'CTCAGTCAGTTCCAGTTAGGACATGGAAACCCATGGTCTCTGGTCTCAGTCACCTTGCGGATAAT\n',
'TTGGGCAATGTGGAGATTGACCTCACTGAAAGCAAAGTGGGTGTTCACATCAACATCACACAGACCAAGA\n',
'TGAGCCAATCGATATTCCTCAGGTTCTGGGTTGAATGATGGACAGTGGCACGAGGTTGCTTCCCTAGC\n',
'CAAGGAAAATTTGCTATTCTACCACATCGATGGAGATGAAGCATCAGCAGTTGAACTAATAGTCCCCT\n',
'CAAGTTAAAATGGCGAGAAGTACTTTTGAGGTTTCTGAACCAAGATGAATAACTCAAGTCACTCTG\n',
'TCCTTCAGCCTTCATTCAAGGATGCATGCAGCTATTCAAGTGGACGATCAACTGTAAATTATACGA\n',
'AGTGGCACAAAGGAAGCCGGGAAGTTTGCAGATGTGAGCATTGACATGTGTGCGATCATAGACAGATGT\n',
'GTGCCCAATCACTGTGAGCATGGTGGAAAGTGTCTGCAGAACATGGGACAGCTCAAATGCACTTGTGATG\n',
'AGACAGGATAACAGTGGGCCACCTGCCACAACCTATCTACGAGCCTCTGTGAAGCCTACAAACACCT\n',
'AGGACAGACATCAAATTATTACTGGATAGATCCTGATGGCAGCGGACCTCTGGGGCCTCTGAAAGTTAC\n',
'TGCAACATGACAGAGGACAAAGTGTGGACCATAGTGTCTCATGACTTGCAGATGCAGACGCCGTGGTCG\n',
'GCTACAACCCAGAAAAACTCAGTGACACAGCTCGTTACAGCGCCATGGACAGATAAGTGCCT\n',
'CACTGACAGTGCCTGAGTACTGCAGCATGAGCTATGTCCTATTCTGCAAGATGTCAAGATTGTTAACACC\n',
'CCAGATGGAAGCCCTACACTGGTGGGTTGGCAAAGCCAACGAGAAGCAGTACTACTGGGAGGCTCTG\n',
'GGCCTGGAATCCAGAAATGTGCCTGCCATCGAACGCAACTGCACAGATCCAAAGTACTACTGTAAC\n',
'CGACGCGGACTACAAGCAATGGAGGAAGGATGCTGGTTCTTATCATAAAAGATCACCTGCCAGTGAGC\n',
'CAAGTGGTGGTTGGAGATACTGACCGTCAAGGCTCAGAACGCAAATTGAGCGTAGGTCTCTGCCTGCC\n',
'AAGGAGACAGGAATTGGAAATGCCCTCTTCCAAACCCATCCTCACCTGCACTCTACTTT\n',
'CCAAGGGAAACTAGCGCTGACATTCTTCACTTCAAAACATTAACCCCTGGGGAGTGTGAACTT\n',
'AATATGGGAAAGGAAGATTTCATCAAGCTGGAGCTGAAGTCTGCCACAGAAGTGTCTTTCTTGTGAA\n',
'TGGGAAATGGGCCAGTAGAGATTGAGGTGACCAACCCCTCAACGATGACCGAGTGGCACCGG\n',
'CACTGCAGAGAGGAATGTCAAGCAGGCCAGCCTACAGGTGGACCGCTACCGCAGCAGATCCGCAAGGCC\n',
'CCAACAGAAGGCCACACCCGCTGGAGCTACAGCCAGTTATTGTGGGTGCTGGGGCCAGCAGG\n',
```

'GCTTCCTGGCTGCATCCGCTCTTGAGGATGAATGGGTGACACTTGACCTGGAGGAAAGAGCAAAGGT\n',
'CACATCTGGGTTCATATCCGGATGCTCGGCCATTGCACCAGCTATGGAACAAACTGTGAAAATGGAGGC\n',
'AAATGCCTAGAGAGATAACACGGTTACTCCTGCGATTGCTCTAATACTGCATATGATGGAACATTGCA\n',
'ACAAAGATGTTGGTGCATTTTGAAAGAAGGGATGTGGCTACGATATAACTTCAGGCACCAGCAACAAA\n',
'TGCCAGAGACTCCACGAGCAGAGTAGACAACGCTCCGACCAGCAGAACTCCCACCCGACCTGGCACAG\n',
'GAGGAGATCCGCTTCAGCTCAGCACCAAGGCACCTGCAATTCTCCTACATCAGCTCTTCACCA\n',
'CAGACTTCTGGCAGTCCTCGTCAAACCCACTGGAAGCTTACAGATTGATACAACCTGGTGGCACCCG\n',
'AGAGCCATACAATATTGACGTAGACCACAGGAACATGGCCAATGGACAGCCCCACAGTGTCAACATCACC\n',
'CGCCACGAGAAGACCATCTTCTCAAGCTCGATCATTCTCTGTGAGTTACCATGCCAAGTTCAT\n',
'CCGACACCTCTTCATTCTCCAAGTCGCTCTTCTGGAAAAGTTATAGAACAGGGAAAATTGACCA\n',
'AGAGATTACAACACACACCCAGGATTCACTGGTGCCTCCAGACTCCAGGGCGAGCTGGTGGAGT\n',
'CCTCTCAAGGCCGCTTGAGGCAGACAAACGCCCTCGCTCACGTCACATCCAGGGCACCTGGCACCTGGATCA\n',
'CCAACTGCCGGGCGCTCGCCGCTGACCCCTCCCCATGTCGTCGCCACCGACCCCTGGCACCTGGATCA\n',
'CCTGGATTCAAGCCAGTGCAGTTTCCATATAATCCAGGACAAGGCCAAGCTATAAGAAATGGAGTCAAC\n',
'AGAAAATCGGCTATCATTGGAGGCGTCATTGCTGTGATTTCACCATCTGTGACCCCTGGTCTTCC\n',
'TGATCCGGTACATGTTCCGCCACAAGGGCACCTACCATACCAACGAAGCAAAGGGGGGGAGTCGGCAGA\n',
'GAGCGCGGACGCCGCATCATGAACAACGACCCCAACTTCACAGAGACCATTGATGAAAGCAAAAGGAA\n',
'TGGCTATTGAGGGGTGGCTACTTGGCTATGGATAGGGAGGGAGGAATTACTAGGGAGGAGAGAAAGG\n',
'GACAAAAGCACCTGCTTCATCTTGGACATCCTTAAAATATCAGCACAAGTTGGGGGAGGCAGGC\n',
'AATGGAATATAATGAAATATTCTTGAGACTGATCACAAAAAAACCTTTTAATATTCTTATAGC\n',
'TGAGTTTCCCTCTGTATCAAAACAAAATAACAAAAATGCTTTAGAGTTAACGAAATGGTTGAAA\n',
'TTTGTTAGGTACTATCTGCTTATTTGTGTGTTAGAGGTGTTCAAAGACCGTGGTAACAGGGCAA\n',
'GTTTCTACGTTTAAGAGCCTTAGAACGTGGTATTTCTTGAGAAAAGCTAATGCACCTACA\n',
'GATGGCCCCAACATTCTCTTCTGTTCTAGTCACCTTAATGGGCTGTTACAGAAACTAGTTGT\n',
'GTTTATATACTATTCCCTTGATGTCCTATAAGTCGGAAAAGAAAGGGCAAAGAGAACCTATTATTG\n',
'CAGTTTTAAGCAGAGCTCAATCTATGCCAGCTCTGGCATCTGGGTTCTGACTGATACCAGCAGT\n',
'GAAGGAAGAGAGTGCATGGCACCTGGTGTGAAACGACACAATCAGCACAACTGGAGAGAGGCATTAAAGA\n',
'ACCAGGGAAGGTAGTTGATTTTCAATTGAATTCTACAAGCTAATATTGTCACGTATGATGTTAGA\n',
'CCAATAGCTGTAACTATCAGCTGCAATACCATGGTGACCGAGCTGTTACAAAGATTCTGTTTAT\n',
'CTGAAACATACTGGATTATATGTATAAGCGCCTCAATGGGAATTAGAGCCAGATGTTATGATTG\n',
'TTGCTTTTCTTTATAGTTAGTAAACAAATGGATAATTCTAGTGAATGCATAAATTAGG\n',
'TTGCGTTCTTATTTGCTTAAATCTGGTAGTTTCCACCCCTGTGACACAATCTTAATAGACAGT\n',
'GTCCTGAAATGGACACAACACAATAAGTCAGTTATTCTGGTACTCTGGATGATATGGAAAACA\n',
'CTGCCATATTTAAATCAACTACTCCACGTGTTTCCATCCAATCACACTGCTGTGATTGGATCTT\n',
'TCTTCTAAGACGGACACATTGAACCTCAGGTTCATCACAAACCTGGTACCTGTTGCTCCAGAGGATG\n',
'GAGAAGTGTAGTTAATCACACCTTAGTTAATCTGAAATCTGACCCAGTTATTTAACAAATAAC\n',
'CTCATTGATTATATTAAAAGTAATACACTTCCTGAAACAAATGGGACAATGCATCCAAAATCTT\n',
'TTAAACAGATTACACAAAATTATTCCAGAAAGGCTACCATTTATCATCATTATATTCAAGCCTTTA\n',
'TACTTAATAAGCACTTCTAAAAGTCTTGAGATCCACATTCTGAGGAATTCAATATGACTTTT\n',
'CCTTCTTGCTGGAGAGGTTAAGAGGAGGTTGAAGGTAGATGCTATTGTTGATGGCCGGCT\n',
'GAATAAAATGGAAATTCTAGTTGTTAGAATTATGCATTTTCAAGATTCTCAGTGTGCCTAACTTA\n',
'TTGGAGCACATCAGTTCTGGTAATGGAAAACATTACCTAGAGTTGCCAGTGGCACATTACACCAGTA\n',

'CAGAGCACATTCAAAGGAGACATTGGACCAGTTAATTCCCATAACA GTCAAGGTAA CAGAACAAAAGGG\n',
'AATCCTGATGCCCTTTACCATGCTGGTGAGCTAGGC ACTGT CATGGACACCCTTAATTTAAAAGG\n',
'TTTAATCATTCTTCTATAAAATACATTAAAATGGAAAAACTTAATATCACTAAATATCAGAACAT\n',
'GTAACATTACAAATGACATATTGAAAGCAAAGGCTGTTTATTAGCCAAGATGATT ACCATTAGGAGT\n',
'TAC TTTATGTATTGTTGAAAGCAAATT TAAACATGATGTTAGAAGTGTCTGATTTAAACCTGG\n',
'TTTACAGGTATTACTCTGCACTTACAAATAATGCCAGATGGAATTTATTATTCCTGCAATTCCAT\n',
'GATAGCTCTGTTCTTATGCATTGTCACACTTCCCTTTCCAAAATGAGTAGAGAACATGAGC\n',
'CACCCAAAACAGCTCTGCTACTAAATGTTCTCATCCTTCTCCCTCCTTCCACAAAAA\n',
'GGTAAAAATGAGATCCAATCCTCACC AAAATTCAAA CCTAGGACACTGGAATGACTGCAGGGATCA\n',
'GTGGTTCTCCATATCACCATCAATTAAAGACATATAGGACACTGTCTCCTCAAGAGGGTTACAATGTG\n',
'GCCATCAGACAGGAACCAACGGTGGATAAGTATTAGTAACAGT GCAAATAATGCTGGAAATC\n',
'TTGACCTCTCCTGGGATTATGGGTGTAACAAAATCCCTACATCTGTTTATGAAGGCCATATTCACTGAC\n',
'ATTTAAATGGTAATAATCTGTTTATGTGAAAGAAAAGAATTAAGTCAACTCTCTCCTGGGTT\n',
'ATAGCCTAGCACAGTGCAGCCTCCATAACC ATGACATTCCCGCCAA GCTCTCAGTGCCTAATCCTGCTT\n',
'TGTATTACATCTCACAAATCTTGACATCTTCAATACATTATCAAGCAAGCACAAGTATGCT\n',
'GGTAGTAGCCTCTTAAATAATGTATAGACAACAACAGCACAAAATAGACTGTTAAAGTTCA\n',
'GGGAAAGTTGGTGGCTGATTTAAAGTTGTCAGGAAACATCTCTGTGATGAAAGCAAATGTCGATGTT\n',
'TGAAAAAAGCTAGGAGATGACTTGAATGAATGCAAGGTTAGT GAGATCCTAAGCTCTCAAATAGCATA\n',
'TTCCCTAGAGCTCAAGAAAGCTGGTCCAGGAGGTTGAAAAAGCTATTTGTTGTTAAATTATTTCTGGC\n',
'CCTTCTTAATATTTAAAATGTATTTCCCTGTGGCTTCAACCACCTGCTCAAAAAAAGAGACTTGT\n',
'ACATGAAAGTTTCATTAAGAGCTGAAAACAAGAAATTAGAGAGCCATTCTAGAAAATGCTCTACTGC\n',
'CCTGCATTGACAACAAGCATCCTTACTAACAGAGCAGGAATT CAGAGGCACAAGAAAAGCATTGG\n',
'CATGAGCCAAGAGTCTGTCTTAATGTTACTTTGAAAATCTGCTGAGCGGCCACCATATGCAGGCTGAG\n',
'AGCTGGGCACAGGGGAAGCCATTGGAAAGCACTCAGGAACAAGCACACAGCTGTGGGACTTGAAACATGCA\n',
'AGTGTTCAGGTTGTCAAGAAGCTTTCTTCTATGATGGAATCTGTTTTCTATCCTACTTT\n',
'TTTCTCTCTCCTCCTCACCACTTACCCCTGCTTACGCACTAAACGTTTAATGGCCCGTTTAT\n',
'GTCTCATGCCTCCAAACAACACTGAATTGAAACCCCCCATT TTCTTACCCCTGTTGAGCAAT\n',
'TTTCCCAAAAAGGGCAGCAATTAAATTGAATTCAAGTAAGCCAGC CAAAGATAGGTCTAAATTG\n',
'CTAGTCCCACTGAGAACCCACTGATCCTAAACCAGT GCGAAACAAACAGTAACATGTCCCAGCTGACTT\n',
'CAGCTAAGAACCAATGGCTCTACCCCCGCCCGCTTTTTGTTGTTTGTGAGACGG\n',
'AGTCTTGCCTGTCCCCCAGGCTGGAGTGC ACTGGCGCAATCTGGGCTCACTGCAACCTCCTCCTC\n',
'CCACATTGAGGC GATTCTCTGCCTCAGCCTCCAAAGTAGCTGGGATTACAGGCACCCCATCACACCC\n',
'AGCTAATT TTTTTTTTTGTATTAGTAGAACGCCAGGTTTCA CCATGTTGGCCAGGGTGGTCTCG\n',
'AACTCCTGACCTCAAGT GATCCGTCCACCTCGGCCTGCAAATGCTGGGATTACAGGTGTGAGCCACCG\n',
'TGCCGAGCCAGCCCCATT TTAATGATGTTGGTAAAGAGTGGACCATGAGAATTAGCTGACAGCAT\n',
'CCCCTTCTCTCCCTGCCTGGTGGACCCCTCCCTGTGACCTTGGTCAAGTCCTCGAACTTTGTC\n',
'CCGTATTAAAGATGGAGCTGTTTACCTACTTCATAAGACAGTTGCAGGTGCCATTGATTCTGACTGC\n',
'AAAATACCTGAAACCC TTATAAAAGACTGAAAGTCAACGGAGCCTAGT GAAAGACTTACTTGTGGCTT\n',
'GTGGTTGAAAGTCACATCAAAGACAAATGTGGCCAGTT CAGGAATTGGAGACTTACTGGCATGGCTCT\n',
'ACAGCTGCTCAGTTAATCATGCAGACTAACCTGCAACACTGGGAGATGCAACATAGCAAAGGACA\n',
'GAGAAATTAGAATT TTTGTGCAAGAACGCCCTAAATCCCACCTGAATGTAAC TTACAGCTCCCTACCT\n',
'ACTCTCACACATGCCCTCAAACATGCTAGATTGGCTTACATAGGCCAACACAAAATACAAACGTGACG\n',

```
'TGTTCATGTAGCCTAGTGGCTATATGCCTATTCTCCATGTACCCCTGCATGGTAGTGCTGCAAACCTTAAA\n',
'GTACATTCTTTCACAGCAGTATTTTTTCATAAGTGGCATATAAATGTCATTCAATGAAATGGGGAAA\n',
'TCACGTTGAGAAGTTGGTCTGTCATCTCCATTGAGCAAAGACTGGCAGGAGATAATAAAATAATG\n',
'GGCACACATGTATTAATATACAGCACGCATTACAAGTTACAGATAAAATTGTGCTATAAGAAC\n',
'AGCTCTACCAAGACAGTCTGCACCATTCCAAGTCTCAGTTAATTACAGCAACTGCTGCTTCGGAGAT\n',
'GGCTGTGAAAATATGGAAGTCCCTCTAAGTAGGCCAAGAACAGTTCTAGATTACTAAGTTTATT\n',
'TGTCAGGTTTTAAATTTTCAGTGAGCGTGGTAGTCAGAGGTTAGTGCTGTGAAAAGCTGGGCTA\n',
'AATATTCTTCTGTAAGTCAAACAGGATTCCATCCCTGTGAAATAACACAAAATTCACTCTCTAAA\n',
'GCAACAGCATGTAAACTAGAATGAAAGAAGGAAATTATGTACGTATGCCTAATATTCTTGTGAATGTCT\n',
'TTCATTTAACTAAAATTATATTAGAAACCAAGATTGATAAATAAAAATTCAAAGTAGTTTAATTATCCT\n',
'AAAAA\n',
'\n']
```

In [5]: contenido[0]

Out[5]: '>NM_014141.6 Homo sapiens contactin associated protein 2 (CNTNAP2), mRNA\n'

In [6]: contenido[1:]

Out[6]:

```
'CTTCAAGAACCTACGGAGAGTCGGACTGCATCTCCGAGCGAGCTTTGGAGCGCCGCCGGAGG\n',
'CGAAGGATGCAGGCCGCTCCGGCGCCGGCTGCGGGCAGCGCTCTGCTGTTGAGCTCAGCAGCTGCC\n',
'TCTCAGAGCCTGGACGGCTCCCTCACGTCCAAAATGTATGAGCCACTTGTCTGGACTCCCCCA\n',
'TGTGGCTTCAGCAGCTCCTCCATCTGGTAGCTATTCTCCGGCTATGCCAAGATAAACAAAGAGA\n',
'GGAGGTGCTGGGGATGGTCTCCATCAGACAGCGACCATTATCAATGGCTCAGGTTGACTTGGCAATC\n',
'GGAAGCAGATCAGTGCATTGCAACCCAAGGAAGGTATAGCAGCTCAGATTGGGTGACCCAATACCGGAT\n',
'GCTCTACAGCGACACAGGGAGAACTGGAAACCCATCATCAAGATGGGAATATCTGGGCATTTCCCGGA\n',
'AACATTAACCTGACGGTGTGGTCCGGCACGAATTACAGCATCCGATTATTGCCGCTATGTGCGCATAG\n',
'TGCCTCTGGATTGGAATGGAGAAGGTGCGATTGGACTCAGAATTGAAGTTATGGCTTCTTACTGGC\n',
'TGATGTTATCAACTTGATGGCCATGTTGATTACCATATAGATTAGAAACAAGAAGATGAAACACTG\n',
'AAAGATGTCATTGCCCTGAACTTAAGACGTCTGAAAGTGAAGGAGTAATCCTGCACGGAGAAGGACAGC\n',
'AAGGAGATTACATTACCTTGGAACTGAAAAAAAGCCAAGCTGGCCTCAGTTAAACTTAGGAAGCAACCA\n',
'GCTTGGCCCCATATATGCCACACATCAGTGTGACAGGAAGTTGCTGGATGACCCACTGGCACTCT\n',
'GTGGTCATTGAGGCCAGGGCGGAGCATTAACCTCACTCTGGACAGGAGCATGCAGCACTCCGTACCA\n',
'ATGGAGAGTTTGAECTACCTGGACTTGGACTATGAGATAACCTTGGAGGCATCCCTTCTGGCAAGCC\n',
'CAGCTCAGCAGTAGAAAGAATTCAAAGGCTCATGGAAAGCATCAACTACAATGGCGTCAACATTACT\n',
'GATCTTGGCAGAAGGAAGAAATTAGGCCCTCAAATGTGGGAATTGAGCTTTCTTGTGTGGAAACCC\n',
'ATACGGTGCCTGTTTCAACGCTACAAGTTACCTGGAGGTGCCGGACGGCTAACCGGACCTGTT\n',
'CTCAGTCAGTTCCAGTTAGGACATGGAAACCCAAATGGTCTCTGGTCTCAGTCACTTGCGGATAAT\n',
'TTGGGCAATGTGGAGATTGACCTCACTGAAAGCAAAGTGGGTGTTCACATCAACATCACAGACCAAGA\n',
'TGAGCCAAATCGATATTCCTCAGGTTCTGGGTGAATGATGGACAGTGGCACGAGGTTGCTTCCTAGC\n',
'CAAGGAAAATTTGCTATTCTCACCATCGATGGAGATGAAGCATCAGCAGTTGAACTAATAGTCCCCTT\n',
'CAAGTTAAAATGGCGAGAAGTACTTTTGGAGGTTCTGAACCAGATGAATAACTCAAGTCAGTCACT\n',
'TCCTTCAGCCTCATTCAAGGATGCATGCAGCTCATTCAAGTGGACGATCAACTGTAAATTATACGA\n',
'AGTGGCACAAAGGAAGCCGGGAAGTTTGCAGATGTCAGCATTGACATGTTGCGATCATAGACAGATG\n',
'GTGCCCAATCACTGTGAGCATGGTGGAAAGTGCTCGCAACATGGGACAGCTCAAATGCACTTGTGATG\n',
'AGACAGGATAACAGTGGGCCACCTGCCACAACCTATCTACAGGCCCTCTGTGAAGCCTACAAACACCT\n',
'AGGACAGACATCAAATTATTACTGGATAGATCCTGATGGCAGCGGACCTCTGGGGCCTCTGAAAGTT\n',
'TGCAACATGACAGAGGACAAGTGTGGACCATAGTGTCTCATGACTTGCAAGATGCAGACGCCGTGG\n',
'GCTACAACCCAGAAAAACTCAGTGACACAGCTGTTACAGGCCCTCATGGACAGATAAGTGCAT\n',
'CACTGACAGTGGCGAGTACTGCGAGCAGTATGTCCTATTCTGCAAGATGTCAAGATTGTTAACACC\n',
'CCAGATGGAAGCCCTTACACTTGGTGGTTGCAAGCCAACGAGAAGCACTACTTGAGGCTCTG\n',
'GGCCTGGAATCCAGAAATGTGCCCTGGCATCGAACGCAACTGCACAGATCCAAGTACTACTGTA\n',
'CGACGCCGACTACAAGCAATGGAGGAAGGATGCTGGTTCTTACATAACAGATCACCTGCCAGTGAGC\n',
'CAAGTGGTGGTTGGAGATACTGACCGTCAAGGCTCAGAAGCCAATTGAGCGTAGGTCTCTGCCTGCC\n',
'AAGGAGACAGGAATTATTGGAATGCCCTTCCAAACCCATCCTCACCTGCACTTCTACTTT\n',
'CCAAGGGAAACTAGCGCTGACATTCTTCACTTCAAAACATTAACCCCTGGGGAGTGTGAA\n',
'AATATGGGAAAGGAAGATTTCATCAAGCTGGAGCTGAAGTCTGCCACAGAAGTGTCTTTCTT\n',
'TGGGAAATGGGCCAGTAGAGATTGAGGTGACCAACCCCTCTCAACGATGACCAGTGGCACCGGGT\n',
'CACTGCAAGAGAGGAATGTCAAGCAGGCCAGCCTACAGGTGGACGGCTACCGCAGCAGATCCGCA\n',
'CCAACAGAAGGCCACACCCGCTGGAGCTCTACAGCCAGTTATTGTGGGTGGCTGGGGCCAGCAGG\n',
'GCTTCCCTGGCTGCATCGCTCTTGAGGATGAATGGGTGACACTTGACCTGGAGGAAAGAGCAAAGG
```

'CACATCTGGTTCATATCCGGATGCTCGGCCATTGCACCAAGCTATGGAACAAACTGTGAAAATGGAGGC\n',
'AAATGCCTAGAGAGATAACCACGGTTACTCCTGCGATTGCTTAATACTGCATATGATGGAACATTTCGA\n',
'ACAAAGATGTTGGTCATTTTGAGAAGGGATGTGGCTACGATATAACTTCAGGCCAGCAACAAA\n',
'TGCCAGAGACTCCAGCAGCAGAGTAGACAACGCTCCGACCAGCAGAACCTCCACCCGACCTGGCACAG\n',
'GAGGAGATCCGCTTCAGCTCAGCACCAAGGCAGCCCTGCATTCTCCTACATCAGCTCCTCACCA\n',
'CAGACTTCTGGCAGTCTCGTCAAACCCACTGGAAGCTTACAGATTGATACAACCTGGTGGCACCCG\n',
'AGAGCCATACAATATTGACGTAGACCACAGGAACATGGCCAATGGACAGCCCCACAGTGTCAACATCACC\n',
'CGCCACGAGAAGACCATCTTCTCAAGCTGATCATTACCTCTGTGAGTTACCATCTCCAAGTTCAT\n',
'CCGACACCCCTTCATTCTCCAAGTCGCTCTTCTGGGAAAAGTTAGAAACAGGGAAAATTGACCA\n',
'AGAGATTCAAAATACAACACCCCAGGATTCACTGGTTCGCTCCAGACTTCAGATGCCAACAGATGCC\n',
'CCTCTCAAGGCCCTTGAGGCAGACAAACGCCTCGGTCACGCCACATCCAGGGCAGCTGGTGGAGT\n',
'CCAAC TGCGGGGCTCGCCGCTGACCCCTCCCCCATGTCGTCGCCACCGACCCCTGGCACCTGGATCA\n',
'CCTGGATTCAAGCCAGTGGATTTCATATAATCCAGGACAAGGCAAGCTATAAGAAATGGAGTCAAC\n',
'AGAAAATCAGGCTATCATTGGAGGCGTATTGCTGTGGTATTTCACCATCTGTGCAACCTGGTCTTCC\n',
'TGATCCGTACATGTTCCGCCACAAGGGCACCTACCATACCAACGAAGCAAGGGGGCGAGTCGGCAGA\n',
'GAGCGCGGACGCCCATCATGAAACAACGACCCCAACTTCACAGAGACCAATTGATGAAAGCAAAAGGAA\n',
'TGGCTATTGAGGGGTGGCTACTTGGCTATGGGATAGGGAGGAGGAATTACTAGGGAGGAGAGAAAGG\n',
'GACAAAAGCACCTGCTTCATACTCTTGAGCACATCCTAAAATATCAGCACAAGTTGGGGAGGCAGGC\n',
'AATGGAATATAATGGAATATTCTTGAGACTGATCACAAAAAAACCTTTTAATATTCTTATAGC\n',
'TGAGTTTCCCTCTGTATCAAAACAAAATAACAAAAATGTTAGAGTTAACGAAATGGTTGAA\n',
'TTTGTAGGTACTATCTGCTTATTGTGTGTTAGAGGTGTTAAAGACCCGTGGTAACAGGGCAA\n',
'GTTTCTACGTTTAAGAGCCTTAGAACGTGGTATTCTTGAGAAAAGCTAATGCACCTACA\n',
'GATGGCCCCAACATTCTCTTGTCTAGTCACCTTAATGGGCTTACAGAAACTAGTTCGT\n',
'GTTTATATACTATTCTTGTCTATAAGTCGAAAAGAAAGGGGCAAAGAGAACCTATTATTGC\n',
'CAGTTTTAAGCAGAGCTCAATCTATGCCAGCTCTGGCATCTGGGTTCTGACTGATACCAGCAGT\n',
'GAAGGAAGAGAGTGCATGGCACCTGGTGTAAACGACACAATCAGCACAACTGGAGAGAGGCATTAAAGA\n',
'ACCAGGGAAAGGTAGTTGATTTCATTGAATTCTACAAGCTAATATTGTTCCACGTATGAGTCTTAGA\n',
'CCAATAGCTGAACTATCAGCTGCAATACCATGGTGACCAGCTGTTACAAAAGATTCTGTTTAT\n',
'CTGAAACATACTGGATTATATATGTATAAGCGCCTCAATGGGAATTAGAGCCAGATGTTATGATTG\n',
'TTGCTTTTCTTTATAGTTAGTTAGCAAAATATGGATAATTCTAGTGAATGCATAAATTAGG\n',
'TTGCGTTCTTATTGCTTAAATCTGAGTTTCCACCCCTGTGACACAATCCTAATAGACAGT\n',
'GTCCTGTAATGGACACAACACAATAAGTCAGTTATTCTGCTGTTACTCTGGATGATATGGAAAACA\n',
'CTGCCATATTAAATCAACTACTCCACGTGTTCCATCCAATCACACTGCTGTGATTCAAGGGATCTT\n',
'TCTTCTAAGACGGACACATTGAAACCTCAGGTTCATCACAAACCTGGTACCTGTTGCTTCCAGAGGATG\n',
'GAGAAGTGTAGTTAATCACACCTTTAGTTAATCTGAAATCTGACCCAGTTATTAAACAATAAAC\n',
'CTCATTGATTATATTAAAAGTAATACACTTCTGTAAACAAATGGGGACAATGCATCCAAAATCTT\n',
'TTAAACAGATTACACAAAATTATTCCAGAAAGGCTACCATTTATCATCATTATATTCAAGCCTTTA\n',
'TACTTAATAAGCACTTCTAAAAAGTCTGAGATCCACCATCTGAGGAATTCAATATGTCACCTTTT\n',
'CCTTCTTGCCTGGAGAGGTTAAGAGGAGGTTGAGGTATAGATGCTATTGTTGATGGCCCGCT\n',
'GAATAAAATGGAAATTCTAGTTGTTAGAATTATGCATTCTTCAAGATTCTCAGTGTGCTTAACCTT\n',
'TTGGAGCACATCAGTTCTGGTAATGGAAAACATTACCTAGAGTTGCCAGTGGCACATTACACCAGTA\n',
'CAGAGCACATTCCAAAGGAGACATTGGACCAGTTAATTCCATACAAGTAACAGAACAAAAGGG\n',

'AATCCTGATGCCCTTTACCATTGCTGGTGAGCTAGGCAGTCAGTCATGGACACCCCTAATTTAAAAGG\n',
'TTTAATCATTCTTCTATAAAATACATTAAAATGGAAAAACTTAATATCACTAAATATCAGAACAT\n',
'GTAACATTACAATGACATATTGAAAGCAAAGGCTGTTTATTAGCCAAGATGATTACCTAGGAGT\n',
'TACTTTATGTATTGTTGAAAGCAAATTAAACATGATGTTTAGAAGTGGTCTGATTTTAAACCTGG\n',
'TTTACAGGTATTACTTCTGCACTTACCAAATAATGCCAGATGGAATTATTATTCTGCAATTCCCAT\n',
'GATAGCTCTGTTCTTATGCATTGTCACACTTCCCTTCCAAATGAGTAGAGAATTAAAGC\n',
'CACCCAAAACAGCTCTGCTACTAAATGTTCTACCTTCTCCCTCCTTCCACAAAAA\n',
'GGTAAAAATGAGATCCAATCCTCACCACAAATTCAACCTAGGACACTGGAATGACTGCAGGGATCA\n',
'GTGGTTCTCCATATCACCATCAATTAAAGACATATAGGACACTGTCCTCAAGAGGGTTACAATGTG\n',
'GCCATCAGACAGGAAACCAAACGGTGATAAGTATTAGTAACTAAGTGCCAATAATGCTGGAAATC\n',
'TTGACCTCTCCTGGGATTATGGGTGTAACAAAATCCCTACATCTGTTATGAAGGCCATATTCAAGTAC\n',
'ATTTAAATGGTAAATAATCTGTTATGTGAAAGAAAAAGAATTAAAGTCTTCTTCAACTCTCCTGG\n',
'ATAGCCTAGCACAGTGCAGCCTCCATAACCATGACATTCCGCCAAGCTCAGTGCCTAATCCTGCTT\n',
'TGTCAATTACATCTCACAAATCTTGACATCTTACATTCCAATACATTCAAGCAAGCACAAGTATGCT\n',
'GGTAGTAGCCTTTAAATAATGTATAGACAACAAACGACAAAAATAGACTGTTAAAGTTCA\n',
'GGGAAAGTTGGTGGCTGATTTAAAGTTGTCAGGAAACATCTCTGTATGAAGCAAATGCGATGTT\n',
'TGAAAAAAGCTAGGAGATGACTTGAATGAATGCAAGGTTAGTGAGATCCTAAGCTCTCAAATAGCATA\n',
'TCCCTAGAGCTCAAGAAAGCTGGCCAGGAGGTTGAAAAAGCTATTGTTGTTAAATTATTTCTGGC\n',
'CCTCTTAATATTTAAAATGTATTCCTTGCGCTTCAACCACCTGCTCAAAAAAAAGAGACTTGT\n',
'ACATGAAAGTTTCATTAAGAGCTGAAAACAAGAATTAGAGAGCCATTCTAGAAAATGCTCTACTGC\n',
'CCTGCATTGACAAACAAGCATCTTACTAACAAAGAGCAGGAATTAGGAGGACAAGAAAAAGCATTGG\n',
'CATGAGCCAAGAGTCTGCTTAATGTTACTTTGAAAATCTGCTGAGCGGCCACCATATGCAGGCTGAG\n',
'AGCTGGGCACAGGCGAAGCCATTGGAAAGCACTTCAGGAACAAGCACACAGCTGTTGGACTTGAACATGCA\n',
'AGTGTTCAGGTTGTCAAGAAGCTTTCTTCTATGATGGAATGTTCTTCTATCCTACTTT\n',
'TTTCTCTTCTCCTCCTCACCACTTACCTGCTTACGCACTAAACGTTTAATGGCCCGTTTAT\n',
'GTCTCATGCCTCAAACAACACTGAATTGAAACCCCCATTTCCTCACCACTGTTGAGCAAT\n',
'TTTCCCAAAAAAAGGGCAGCAATTATTGAATTCAAGTAAGCCAGCAGCAAGATAGGTCTAAATTG\n',
'CTAGTCCCAGTAGAACCAACCTGATCTAAACCACTGCGAAACAAACAGTAACAATGTCCTCAGCTGACTT\n',
'CAGCTAAGAACCAATGGCTCTACCCCCGCCCGCTTTTGTGTTTTGTTGAGACGG\n',
'AGTCTGCTCTGCCCCAGGCTGGAGTCAGTGGCGCAATCTGGGCTCACTGCAACCTCCTCCTCCTC\n',
'CCACATTGAGGCGATTCTCCTGCCTCAGCCTCCAAAGTAGCTGGATTACAGGCACCCGCATCACACCC\n',
'AGCTAATTTTTTTTTTGTATTAGTAGAAGCCAGGTTCACCATGTTGCCAGGGTGGCTCG\n',
'AACTCCTGACCTCAAGTGATCCGTCCACCTCGGCCCTGCAAATTGCTGGGATTACAGGTGTGAGCCACCG\n',
'TGCCGAGGCCAGCCCCATTAAATGATGTTGGTTAAGAGTGGACCATGAGAATTAGCTGACAGCAT\n',
'CCCCTTCTCTCCTGCCTGGGGACCCCTCCCTGTCAGCTCGAACCTTGTGAACTTGTCAACTTTGTC\n',
'CCGTATTAAAGATGGAGCTGTTTACCTACTTCATAAGACAGTTGCGAGGTGCCATTGATTCTGACTGC\n',
'AAAATACCTGAAACCCCTATATAAGACTGAAGTCACGGAGCCTAGTGAAGAGACTTACTTGTGGCTT\n',
'GTGGTTGAAAGTCACATCAAAGACAAATGTGCCACGTTCAGGAATTGGAGACTTACTGGCATGGCTCT\n',
'ACAGCTGCTCAGTTATTAAATCATGCAGACTAACCTGTCACACTGGGAGATGCAACATAGCAAAGGACA\n',
'GAGAAATTAGAATTGGCTGAGAAAGCCCTAAATTCCACCTGAATGAACTTACAGCTCCCTACCT\n',
'ACTCTCACACATGCCCTCAAACATGCTAGATTGGCTTACATAGGCCAACACAAATACAAACGTGACG\n',
'TGTTCATGTAGCCTAGTGGCTATATGCCTATTCTCCATGTACCCCTGCATGGTAGTGCTGCAAACTTAAA\n',

```
'GTACATTCTTCACAGCACTTTTCTAAGTGGCATATAATGTCATTCAATGAAATGGGAAA\n',
'TCACGTTGAGAAGTTGGTCTGTCATCTCCATTGAGCAAAGACTGGCAGGAGATAATAAAATAATG\n',
'GGCACACATGTATTAATATACAGCACGCATTACAAGTTATTTCCAGATAAAATTGTGCTATAAGAAC\n',
'AGCTCTACCAAGACAGTCTGCAACCATTCCAAGTCTCAGTTAATTCAGCAACTGCTGCTTCGGAGAT\n',
'GGCTGTGAAAATATGGAAGTCCCTCTCAAGTAGGCCAAGAACAGTTCTAGATTTACTAAGTTTATTT\n',
'TGTCAGGTTTTAAATTTTCAGTGAGCGTGGTACTGCAGAGGTTAGTGCTGTGAAAAGCTGGGCTA\n',
'AATATTCTTCTGAAAGTCAAACAGGATTCCATCCCTGTGAAATAACACAAAATTCACTCTCTAAAA\n',
'GCAACAGCATGTAAACTAGAATGAAAGAAGGAAATTATGTACGTATGCCTAATATTCTTGTGAATGTCT\n',
'TTCATTAACTAAAATTATATTAGAAACCAGATTGATAAATAAAAATTCAAAGTAGTTTAATTATCCT\n',
'AAAA\n',
'\n']
```

```
In [7]: contenido[0].replace("\n", "")
```

```
Out[7]: '>NM_014141.6 Homo sapiens contactin associated protein 2 (CNTNAP2), mRNA'
```

```
In [8]: "" .join(contenido [1:])
```

Out[8]:

'CTTCAAGAACCCCTACGGAGAGTCGGACTGCATCTCCGCAGCGAGCTTGGAGCGCCGCCGGGAGG\nnCGAAGGATGCAGGCGCTCCGCAGCCGGCTGC
GGGGCAGCGCTCTGCTGTGGATTGTCAGCAGCTGCC\nnTCTGCAGAGCCTGGACGGCTCCACGTCCAAAATGTGATGAGCCACTTGTCTCTGGACTCCC
CCA\nnTGTGGCTTTAGCAGCTCCTCCATCTGGTAGCTATTCTCCCGCTATGCCAAGATAAAAGAGA\nnGGAGGTGCTGGGGATGGTCTCCATCAGA
CAGCGACCATTATCAATGGCTCAGGGTGAATTGGCAATC\nnGGAAGCAGATCAGTGCATTGCCAACCAAGGAAGGTAGCAGCTCAGATTGGGTGACCCAAAT
ACCGGAT\nnGCTCTACAGCGACACAGGGAGAACTGGAAACCCATCATCAAGATGGGAATATCTGGCATTCCCGGA\nnAACATTAACTCTGACGGTGTTGCT
GGCACGAATTACAGCATCCGATTATTGCCGCTATGTGGCATAG\nnTGCCTCTGGATTGGAATGGAGAAGGTGCAATTGGACTCAGAATTGAAGTTATGGCT
TCTTACTGGC\nnTGATGTTATCAACTTGATGGCATGGTATTACCATATAGATTAGAAACAAGAAGATGAAAACACTG\nnAAAGATGTCATTGCCCTGAAC
TTAAGACGCTGAAAGTGAAGGAGTAATCTGCACGGAGAAGGACAGC\nnAAGGAGATTACATTACCTGAACTGAAAAAGCCAAGCTGGTCTCAGTTAAA
CTTAGGAAGCAACCA\nnGCTTGGCCCCATATATGCCACACATCAGTGATGACAGGAAGTTGCTGGATGACCACACTGGCACTCT\nnGTGGTCATTGAGGCCA
GGGGCGGAGCATTAAACCTACTCTGGACAGGAGCATGCAGCACTCCGTACCA\nnATGGAGAGTTGACTACCTGGACTATGAGATAACCTTGGAGGCA
TCCCTTCTGGCAAGCC\nnCAGCTCCAGCAGTAGAAAGAATTCAAAGGCTGCATGGAAAGCATCAACTACAATGGCGTCAACATTACT\nnGATCTTGCCAGAA
GGAAGAAATTAGAGCCCTAAATGTGGAAATTGAGCTTTCTTGTTGGAACCC\nnATACGGTGCCTGCTTTTCAACGCTACAAGTTACCTGGAGGTGCC
GGACGGCTTAACCAGGACCTGTT\nnCTCAGTCAGTTCCAGTTAGGACATGGAACCCAAATGGTCTCCTGGTCTCAGTCATTGCGGATAAT\nnTTGGCAAT
GTGGAGATTGACCTACTGAAAGCAAAGTGGGTGTTACATCAACATCACAGACCAAGA\nnTGAGCCAATCGATATTCCTCAGGTTCTGGGTTGAATGATGG
ACAGTGGCACGAGGTTGCTCCTAGC\nnCAAGGAAATTGCTATTCTCACCACATCGATGGAGATGAAGCATCAGCAGTCAACTAATAGTCCCTT\nnCAAGT
TAAAACGGGAGAAGTACTTTTGAGGTTCTGAAACCAGATGAATAACTCAAGTCAGTCTG\nnTCCTCAGCCTCATTCCAAGGATGCATGCAGCTCATT
AAGTGGACGATCAACTGTAAATTATACGA\nnAGTGGCACAAAGGAAGCCGGGAAGTTGCGAATGTCAGCATTGACATGTGCGATCATAGACAGATGT\nnG
TGCCCAATCACTGTGAGCATGGGAAAGTGCTGCACACATGGGACAGCTCAAATGCACTTGATG\nnAGACAGGATACAGTGGGCCACCTGCCACA
ATCTACGAGCCTCCTGTGAAGCCTACAAACACCT\nnAGGACAGACATCAAATTACTGGATAGATCCTGATGGCAGCGGACCTCTGGGCTCTGAAAGTTA
C\nnTGCAACATGACAGAGGACAAAGTGTGGACCATAGTGTCTCATGACTTGCAAGATGCAGACGCCTGGTC\nnGCTACAACCCAGAAAATCTCAGTGCACACA
GCTGTTACAGCCTCCATGGACCAAGATAAGTGCAT\nnCACTGACAGTGGCAGACTGCGAGCAGTATGTCCTATTCTGAAAGATGTCAGATTGTA
ACACC\nnCCAGATGGAAGCCCTAACACTTGGGGTTGGCAAAGCCAACGAGAAGCACTACTACTGGGGAGGCTCTG\nnGGCCTGGAATCCAGAAATGTCCTGCG
GCATCGAACGCAACTGCACAGATCCAAGTACTACTGTAACG\nnCGACGCGACTACAAGCAATGGAGGAAGGATGCTGGTTCTATCATAAAAGATCACCTG
CCAGTGAGC\nnCAAGTGGGGTTGGAGATACTGACCGTCAAGGCTCAGAAGCCAATTGAGCGTAGGTCCTCTGCGCTGCC\nnAGGAGACAGGAATTATTGAAT
GCCGCTCTTCCAAACCCATCCTCCTACCTGCACTTCTACTTT\nnCCAAGGGAAACTAGCGCTGACATTCTTCTACTTCAAACATTAACCCCTGGGG
AGTGTGTTCTGAA\nnAATATGGGAAAGGAAGATTTCATCAAGCTGGAGCTGAAGTCTGCCACAGAAGTGTCTTCTATTGATG\nnTGGGAAATGGGCCAGTAGA
GATTGTAGTGAGGTACCAACCCCTCTAACGATGACCGTGGCAGCGGAGT\nnACTGCAGAGAGGAATGTCAGCAGGCCAGCCTACAGTGGACCGGCTACCGC
AGCAGATCCGCAAGGCC\nnCCAACAGAAGGCCACACCCGCTGGAGCTACAGCCAGTTATTGAGGTGGCTGGGGGCCAGCAGG\nnGCTTCTGGCTGCA
TCGCTCCTGAGGATGAATGGGGTACACTGACCTGGAGGAAAGGCAAAGGT\nnCACATCTGGGTTCATATCCGGATGCTGGGCCATTGCACAGCTATGGA
ACAAACTGTAAAATGGAGGC\nnAAATGCCATAGAGAGATACCACGTTACTCTGCGATTGCTTAATACTGCATATGATGGAACATTGCA\nnACAAAGATGTT
GGTCATTGGAAGAAGGGATGTGGCTACGATATAACTTCAGGCACCAAGCAAAN\nnTGCCAGAGACTCCAGCAGCAGAGTAGACAACGCTCCGACCA
GAACCTCCACCCGGACCTGGCACAG\nnGAGGAGATCCGCTCAGCTCAGCACCACCAAGGCGCCCTGCATTCTCTACATCAGCTCCTCAGCA\nnCAGACTT
CTTGGCAGTCTCGTCAAACCCACTGGAAGCTTACAGATTGATACAACCTGGTGGCACCCG\nnAGAGCCATACAATATTGACGTAGACCACAGAACATGGCA
ATGGACAGCCCCACAGTGTCAACATCACC\nnCGCCACGAGAAGACCATCTTCTCAAGCTGATCATTCTCTGTGAGTTACCATGTCAGTTCA
nCCG
ACACCCCTTCAATTCTCCAAGTCGCTTTCTGGAAAAGTTATGAAACAGGGAAAATTGACCA\nnAGAGATTCAAACACACCCAGGATTCACTGGT
TGCTCTCCAGAGTCCAGTTCAACCAGATGCC\nnCCTCTCAAGGGCGCCTTGAGGAGACAAACGCCCTGGCTCACGTCCACATCCAGGGCAGCTGGTGGAGT
\nCCAACGCGGGGCTCGCCGCTGACCCCTCTCCCCATGTCGTCGCCACCGACCCCTGGCACCTGGATCA\nnCCTGGATTGAGCCAGTGCAGGATTTCCATATA
ATCCAGGACAAGGCCAGCTATAAGAAATGGAGTCAC\nnAGAAACTCGGCTATCATTGGAGGGCGTATTGCTGTGGTATTTCACCACCTGTGACCCCTGGTC
TTCC\nnTGATCCGGTACATGTTCCGCCACAAGGGCACCTACCATACCAACGAGCAAAGGGGGCGGAGTCGGCAGA\nnGAGCGCGACGCCCATGAAACAC
GACCCCAACTTCACAGAGACCAATTGATGAAAGCAAAGGAA\nnTGGCTCATTGAGGGGTGGCTACTTGGCTATGGGATAGGGAGGAGGAAATTACTAGGGAGGA
GAGAAAGG\nnGACAAAAGCACCCCTGCTCATACTCTGAGCACATCCTAAATATCAGCACAAGTTGGGGAGGAGGCAGGC\nnAATGGAATATAATGGAATATTCTT
GAGACTGATCACAAAAAAACCTTTAATATTCTTATAGC\nnTGAGTTTCCCTGTATCAAACAAATAATACAAAAATGCTTTAGAGTTAAG

CAATGGTTGAAA\nTTTAGGTACTATCTGCTTATTGAGGGTTCTAAAGACCCGTTAACAGGGCAA\nGTTTCTACGTTTAAGAG
CCCTAGAACGTGGGTATTTCTTGAGAAAAGCTAATGCACCTACA\nGATGCCCAACATTCTTCCTTGCCTAGTCACCTTAATGGCTGTT
ACAGAAACTAGTCGT\nGTTATATACTATTCCTTGATGTCCTATAAGTCGGAAAAGAAAGGGCAAAGAGAACCTATTGCT\nCAGTTTAAGCAGAG
CTCAATCTATGCCAGCTCTGGCATCTGGGTTCTGACTGATACCAGCAGTT\nGAAGGAAGAGAGTGCATGGCACCTGGTGTAAACGACACAATCAGCACAA
CTGGAGAGAGGCATTAAGA\nACCAGGGAAAGTAGTTGATTTCATTGAATTCTACAAGCTAATATTGTCACGTATGTTAGA\nCCAATAGCTGTA
ACTATCAGCTGCAATACCATGGTACCAGCTGTTACAAAAGATTTCTGTTTAT\nCTGAAACATACTGGATTATATGTATAAGCGCTCAATGGGAA
TTAGAGCCAGATGTTATGATTG\nTTGCTCTTTCTTTATAGTTAGTAAAGATAATTCTAGTGAATGCATAAAATTAGG\nNTGCGTT
CTTATTTGCTTAAATCTCTGGTAGTTTCCACCCCTGTACACAATCCTAATAGACAGT\nGTCCTGTAATGGACACAACAAATAAGCTAAGTTATTATT
GCTGTTACTCTGGATGATGGAAAACA\nCTGCCATATTTAAATCAACTACTCCACGTGTTCCAGGGATG\nGAGAAGTGTAGTTAATCACACCTCTAGTTAATCTGA
AACTTGACCCAGTTAACAAATAATAC\nCTCATTGATTATTTAAAGTAATACACTCCTGAAACAAATGGGACAATGCATCCAAAAATCTT\nTTAACAGATTACACAAAATTATTCAGAAAGGCTACCATTTATCATCATTATTTCAAGCCTCTTA\nTACTTAATAAGCACTTCTAAAGTCTTGAGAT
CCCACCATCTGAGGAATTCAATATGACTCTTTT\nCTTGGCTGGAGAGGTTAAGAGGAGTTGAAGGTATAGATGCTATTGTTCTGATGGCCGG
CT\nGAATAAAATGGAAATTCTAGTTGTTAGAATTATGCATTCTTTCAAGATTCTCAGTGTGCTTAACCTTA\nTTGGAGCACATCAGTTCTGGTAATGGA
AAACATTACCTAGAGTGCAGGGCACATTACACAGTA\nCAGAGCACATTCAAAGGAGACATTGGACCAGTTAACCCATACAAGTCAAGTAACAGAAC
AAAGGG\nAATCTGATGCCCTTACATTGCTGGTAGCTCAGGCACTGTACGGACACCCTAATTTAAAGG\nNTTAATCATTCTATAAAATACA
TTAAAATGGAAAATACTTAATACACTAACAT\nGTAACATTACAAATGACATATTGAAAGCAAAGGCTTTATTAGCCAAGATGATTAC
CATTAGGAGT\nTACTTATGATTGAAAGCAAATTAAACATGATGTTAGAAGTGTCTGATTTAAACCTGG\nNTTACAGGTATTACTCTGCAC
TTACCAAATAATGCCAGATGGAAATTATTATTCCTGCAATTCCAT\nGATGCTCTTATGCATTGCTCAACACTTCCCTTTCCAAAATGAG
TAGAGATTAAAGC\nCACCCAAAACAGCTCTGCTACTAAAATGTTCTCATCTTCTCCTCCTCTGCAACAAA\nGGTAAAATGAGATCCA
ATCCTCTACCAAAATTCAAACCTAGGACACTGGAATGACTGCAGGGATCA\nGTGTTCTCCATATCACCATAATTAGACATATAGGACACTGCTTCTT
CAAGAGGGTACAATGTG\nGCCATCAGACAGGAAACCAACGGTGGATAAGTATTAGTAACAGTGCACAAATAATGCTGAAATC\nTTGACCTCTCCTT
GGATTATGGGTGTAACAAAATCCTACATGTTTATGAAGGCCATTAGTAC\nNTTTAAATGGTAAATACTGTTATGTGAAAGAAAAGATTAGTC
TTCTTCAACTCTCCTTGG\nATAGCCTAGCACAGTCAGCCTCCATAACCATGACATTCCGCCAGCTCTCAGTGCCTAATCTGCTT\nHTGTCATTAC
ATCTCACAAATCTGACATCTTCAATACATTCAAGCAAGCACAAGTATGCT\nGGTAGCTCTTAAATAATGTATAGACAACAACAGAC
AAAAATAGACTGTTAAAGTTCA\nGGAAAGTTGGCTGATTAAAGTTGAGACAGCTCAAAATAGCATA\nTTCCCTAGAGCTCAAGAAAGCTGGCCAGGAGGTTGAAA
AGCTATTGTTGTTAAATTATTCCTGG\nCCTTCTTAATATTAAAATGATTCCCTGGCTTCAACCACCTGCTAAAAAGAGACTGTT\nAC
ATGAAAGTTCTTCAAAAGAGCTGAAAACAAGAACAGATTAGAGAGCATTCTAGAAAATGCTACTGC\nCCTGCATTGACAACACAAGCATCTTACTAACAG
AGCAGGAATTAGGGACAAGAAAAGCATTGG\nCATGAGCCAAAGAGTCTGCTTAATGTTACTTTGAAAATGCTGAGCGGCCACCATATGCAGGCTGAG
AGCTGGCACAGGCGAAGCCATTGGAGCACTTCAGAACAGCACACAGCTGGACTTGAACATGCA\nAGTGTTCAGGTTGTCAAGAAGCTTCTT
CCTCTATGATGAACTGTTCTTCTACTT\nCTTCTTCTCCTCACCACATTACCTGCTTACGAGTAAACGTTAATGGCCGT
TTAT\nGTCATGCCTCAAACAAACTGAATTGAAACCCCCATTCTTCTCACCACATTACCTGCTTACGAGTAAACGTTAATGGCCGT
AATTGAATTCAAGTAAGCCAGCAAAGATAGGTCTAAATTG\nCTAGTCCAGTAGAACCCACCTGATCCTAAACCAAGTGCAGAACAAACAGTAACAATGTCCCCA
GCTGACT\nCAGCTAAGAACCAATGGCTCTACCCCGCCGCTTTTGTGTTGAGACGG\nAGTCTGCTCTGCCCCAGGCTG
GAGTGCAGTGGCGCAATCTGGGCTACTGCAACCTCCCTCCTC\nCCACATTGAGGCAGTCTGCTCAGCCTCCAAGTAGCTGGATTACAGGCAACCC
GCCATCACACCC\nAGCTAATTGTTGTTATTAGTAGAACAGCAGGTTACCATGTTGGCCAGGGTGGCTCG\nAACTCCTGACCTCAAGTGAT
CCGTCACCTCGGCCTGCAAATTGCTGGATTACAGGTGTGAGCCACCG\nTGCGAGGCCAGCCCCATTGTTAAATGATGTTGGTAAAGAGTGGACCATGAG
ATTAGCTGACAGCAT\nCCCCTTCTCTCCCTGCTTGGAGCCATTGATTCTGACTGC\nAAAATACCTGAAACCCATTATAAGACTGAAGTCAACGGAGCCTAGTG
AAAGACTTACTTGTGGCTT\nGTGGTGAAGTCACATCAAAGACAATGTGGCCACGTTAGGAATTGGAGACTTACTGGCATGGCTC\nACAGCTGCTCAG
TTATTAAATCATGCAGACTAACCTGTAACACTGGGAGATGCAACATAGCAAAGGACA\nGAGAAATTAGAATTGTCAGAAAGCCCTAAATCCACCTGA

```
ATGTAACCTACAGCTCCCTACCT\nACTCTCACACATGCCCTAAACATGCTAGATTGGCTTACATAGGCCAACACAAAATACAACGTGACG\nTGTTCATG  
TAGCCTAGTGGCTATATGCCTATTCTCCATGTACCCCTGCATGGTAGTGCTGCAAACCTTAAA\nGTACATTCTTCACAGCAGTATTTTTTCTATAAGTGGCAT  
ATAAATGTCATTCAATGAAATGGGGAA\nTCACGTTGAGAAGTTGGTCTGCTATCTCCATTGAGCAAAGACTGGCAGGAGATAATAAAATAATG\nGGCA  
CACATGTATTAAATACAGCACGACATTACAAGTTATTTCCAGATAAAATTGTGCTATAAGAAC\nAGCTTACCAAGACAGTCTGCACCATTCCAAGTCTCA  
GTTAATTACAGCAACTGCTGCTTCGGAGAT\nGGCTGTGAAATATGGAAGTTCTCAAGTAGGCCAGAACAGTTCTAGATTTACTAAGTTTATT\nTGTCAGGTTTTAAATTTTCAGTGAGCGTGGTACTGCAGAGGTTAGTGCTGTGAAAGCTGGCTA\nAATATTCTTGTAAAGTCAAACAGGATTCCAT  
CCCCTGTGAAATAACACAAAATTCACTCTCTAAAA\nGCAACAGCATGTAAACTAGAATGAAAGAAGGAAATTATGTACGTATGCCATTCTTGTGAATGT  
CT\nTTCATTAACAAATTATATTAGAAACCAGATTGATAAATAAAAAATTCAAAGTAGTTAATTATCCT\nAAAA\n\n'
```

In [9]: `len(contenido[1:])`

Out[9]: 137

3.- Para el encabezado, extrae empleando expresiones regulares el identificador. (1pto)

In [10]: `import re`

In []: `contenido[0]='>NM_014141.6 Homo sapiens contactin associated protein 2 (CNTNAP2), mRNA\n'`

In [11]: `match = re.match(r'^>(\S+)', contenido[0])
if match:
 identificador = match.group(1)`

In [12]: `print("Identificador:", identificador)`

Identificador: NM_014141.6

4.- ¿Cuál es el porcentaje de GC de la secuencia? (1 pto)

In [31]: `secuencia`

```
Out[31]: ['CTTCAAGAACCTACGGAGAGTCGGACTGCATCTCCGAGCGAGCTTTGGAGCGCCGCCGGAGG\n',  
'CGAAGGATGCAGGCCGCTCCGCGCCGGCTGCGGGCAGCGCTCTGCTGTGAGCTCAGCAGCTGCC\n',  
'TCTCAGAGCCTGGACGGCTCCCTCACGTCCAAAATGTATGAGCCACTTGTCTGGACTCCCCCA\n',  
'TGTGGCTTCAGCAGCTCCTCCATCTGGTAGCTATTCTCCGGCTATGCCAAGATAAACAAAGAGA\n',  
'GGAGGTGCTGGGGATGGTCTCCATCAGACAGCGACCATTATCAATGGCTCAGGTTGACTTGGCAATC\n',  
'GGAAGCAGATCAGTGCATTGCAACCCAAGGAAGGTATAGCAGCTCAGATTGGGTGACCCAATACCGAT\n',  
'GCTCTACAGCGACACAGGGAGAACTGGAAACCCATCATCAAGATGGGAATATCTGGGCATTTCCCGA\n',  
'AACATTAACCTGACGGTGTGGTCCGGCACGAATTACAGCATCCGATTATTGCCGCTATGTGCGCATAG\n',  
'TGCCTCTGGATTGGAATGGAGAAAGGTCGATTGGACTCAGAATTGAAGTTATGGCTTCTTACTGGC\n',  
'TGATGTTATCAACTTGTGGCCATGTTGATTACCATATAGATTAGAAACAAGAAGATGAAACACTG\n',  
'AAAGATGTCATTGCCCTGAACTTAAGACGTCTGAAAGTGAAGGAGTAATCCTGCACGGAGAAGGACAGC\n',  
'AAGGAGATTACATTACCTTGGAACTGAAAAAAAGCCAAGCTGGCCTCAGTTAAACTTAGGAAGCAACCA\n',  
'GCTTGGCCCCATATATGCCACACATCAGTGTGACAGGAAGTTGCTGGATGACCCACTGGCACTCT\n',  
'GTGGTCATTGAGGCCAGGGCGGAGCATTAACCTCACTCTGGACAGGAGCATGCAGCACTCCGTACCA\n',  
'ATGGAGAGTTGACTACCTGGACTTGGACTATGAGATAACCTTGGAGGCATCCCTTCTGGCAAGCC\n',  
'CAGCTCAGCAGTAGAAAGAATTCAAAGGCTCATGGAAAGCATCAACTACAATGGCGTCAACATTACT\n',  
'GATCTTGGCAGAAGGAAGAAATTAGGCCCTCAAATGTGGGAATTGAGCTTTCTTGTGTGGAAACCC\n',  
'ATACGGTGCCTGTTTCAACGCTACAAGTTACCTGGAGGTGCCGGACGGCTAACCGGACCTGTT\n',  
'CTCAGTCAGTTCCAGTTAGGACATGGAAACCCAAATGGTCTCTGGTCTCAGTCACTTGCGGATAAT\n',  
'TTGGGCAATGTGGAGATTGACCTCACTGAAAGCAAAGTGGGTGTTCACATCAACATCACAGACCAAGA\n',  
'TGAGCCAAATCGATATTCCTCAGGTTCTGGGTGAATGATGGACAGTGGCACGAGGTCGCTCTAGC\n',  
'CAAGGAAAATTTGCTATTCTCACCATCGATGGAGATGAAGCATCAGCAGTTGAACTAATAGTCCCCTT\n',  
'CAAGTTAAAATGGCGAGAAGTACTTTTGAGGTTCTGAACCAGATGAATAACTCAAGTCAGTCACTTG\n',  
'TCCTTCAGCCTTCATTCAAGGATGCATCAGCTCATTCAAGTGGACGATCAACTGTAAATTATACGA\n',  
'AGTGGCACAAAGGAAGCCGGGAAGTTTGCAGATGTCAGCATTGACATGTGTCGATCATAGACAGATGT\n',  
'GTGCCCAATCACTGTGAGCATGGTGGAAAGTGCTCGCAACATGGACAGCTCAAATGCACTTGTGATG\n',  
'AGACAGGATAACAGTGGGCCACCTGCCACAACCTATCTACAGGCCCTCTGTGAAGCCTACAAACACCT\n',  
'AGGACAGACATCAAATTATTACTGGATAGATCCTGATGGCAGCGGACCTCTGGGCCCTGAAAGTTTAC\n',  
'TGCAACATGACAGAGGACAAGTGTGGACCATAGTGTCTCATGACTTGCAGATGCAGACGCCGTGGTCG\n',  
'GCTACAACCCAGAAAAACTCAGTGCACAGCTCGTTACAGGCCCTCATGGACAGATAAGTGCAT\n',  
'CACTGACAGTGCCGAGTACTGCAGCTAGTGTCTCTATTCTGCAAGATGTCAAGATTGTTAACACC\n',  
'CCAGATGGAAGCCCTTACACTTGGTGGTTGCAAGCCAACGAGAAGCACTACTCTGGGAGGCTCTG\n',  
'GGCCTGGAATCCAGAAATGTGCCCTGGCATCGAACGCAACTGCACAGATCCAAGTACTACTGTAAC\n',  
'CGACGCCGACTACAAGCAATGGAGGAAGGATGCTGGTTCTTATCATACAAAGATCACCTGCCAGTGA\n',  
'CAAGTGGTGGTTGGAGATACTGACCGTCAAGGCTCAGAAGCCAATTGAGCGTAGGTCTCTGCCAGT\n',  
'AAGGAGACAGGAATTATTGGAATGCCCTTCCAAACCCATCTTACCTGCACTTCTACTTT\n',  
'CCAAGGGAAACTAGCCTGACATTCTTCACTTCAAAACATTAACCCCTGGGGAGTGTGAACTT\n',  
'AATATGGGAAAGGAAGATTTCATCAAGCTGGAGCTGAAGTCTGCCACAGAAGTGTCTTTCTTGTGATG\n',  
'TGGGAAATGGGCCAGTAGAGATTGTAGTGGAGGTACCAACCCCTCTCAACGATGACCGAGTGGCAC\n',  
'CACTGCAGAGAGGAATGTCAAGCAGGCCAGCCTACAGGTGGACCGCTACCGCAGCAGATCCGCAAGG\n',  
'CCAACAGAAGGCCACACCCGCTGGAGCTCTACAGCCAGTTATTGTGGGTGGCTGGGGGCCAGCAGG\n',  
'GCTTCCCTGGCTGCATCCGCTCTGAGGATGAATGGGTGACACTTGACCTGGAGGAAAGAGCAAAGG\n',  
']
```

'CACATCTGGTTCATATCCGGATGCTCGGCCATTGCACCAAGCTATGGAACAAACTGTGAAAATGGAGGC\n',
'AAATGCCTAGAGAGATAACCACGGTTACTCCTGCGATTGCTTAATACTGCATATGATGGAACATTTCGA\n',
'ACAAAGATGTTGGTCATTTTGAAAGAAGGGATGTGGCTACGATATAACTTCAGGCCAGCAACAAA\n',
'TGCCAGAGACTCCAGCAGCAGAGTAGACAACGCTCCGACCAGCAGAACCTCCACCCGACCTGGCACAG\n',
'GAGGAGATCCGCTTCAGCTCAGCACCAAGGCAGCCCTGCATTCTCCTACATCAGCTCCTCACCA\n',
'CAGACTTCTGGCAGTCTCGTCAAACCCACTGGAAGCTTACAGATTGATACAACCTGGTGGCACCCG\n',
'AGAGCCATACAATATTGACGTAGACCACAGGAACATGGCCAATGGACAGCCCCACAGTGTCAACATCACC\n',
'CGCCACGAGAAGACCATCTTCTCAAGCTGATCATTACCTCTGTGAGTTACCATCTCCAAGTTCAT\n',
'CCGACACCCCTTCATTCTCCAAGTCGCTCTTCTGGGAAAAGTTAGAAACAGGGAAAATTGACCA\n',
'AGAGATTCAAAATACAACACCCCAGGATTCACTGGTTCGCTCCAGACTTCAGATGCCAACAGATGCC\n',
'CCTCTCAAGGCCCTTGAGGCAGACAAACGCTCGGTCACGCCACATCCAGGGCAGCTGGTGGAGT\n',
'CCAAC TGCGGGGCTCGCCGCTGACCCCTCCCCCATGTCGTCGCCACCGACCCCTGGCACCTGGATCA\n',
'CCTGGATTCAAGCCAGTGGATTTCATATAATCCAGGACAAGGCAAGCTATAAGAAATGGAGTCAAC\n',
'AGAAAATCAGGCTATCATTGGAGGCGTATTGCTGTGGTATTTCACCATCTGTGCAACCTGGTCTTCC\n',
'TGATCCGTACATGTTCCGCCACAAGGGCACCTACCATACCAACGAAGCAAGGGGGCGAGTCGGCAGA\n',
'GAGCGCGGACGCCCATCATGAAACAACGACCCCAACTTCACAGAGACCATTGATGAAAGCAAAAGGAA\n',
'TGGCTATTGAGGGGTGGCTACTTGGCTATGGGATAGGGAGGAGGAATTACTAGGGAGGAGAGAAAGG\n',
'GACAAAAGCACCTGCTTCATACTCTTGAGCACATCCTAAAATATCAGCACAAGTTGGGGAGGCAGGC\n',
'AATGGAATATAATGGAATATTCTTGAGACTGATCACAAAAAAACCTTTAATATTCTTATAGC\n',
'TGAGTTTCCCTCTGTATCAAAACAAAATAACAAAAATGTTAGAGTTAACGAAATGGTTGAA\n',
'TTTGTAGGTACTATCTGCTTATTGTGTGTTAGAGGTGTTAAAGACCCGTGGTAACAGGGCAA\n',
'GTTTCTACGTTTAAGAGCCTTAGAACGTGGTATTCTTGAGAAAAGCTAATGCACCTACA\n',
'GATGGCCCCAACATTCTCTTGTCTAGTCACCTTAATGGGCTTACAGAAACTAGTTCGT\n',
'GTTTATATACTATTCTTGTCTATAAGTCGAAAAGAAAGGGGCAAAGAGAACCTATTATTGC\n',
'CAGTTTTAAGCAGAGCTCAATCTATGCCAGCTCTGGCATCTGGGTTCTGACTGATACCAGCAGT\n',
'GAAGGAAGAGAGTGCATGGCACCTGGTGTAAACGACACAATCAGCACAACTGGAGAGAGGCATTAAAGA\n',
'ACCAGGGAAAGGTAGTTGATTTCATTGAATTCTACAAGCTAATATTGTTCCACGTATGAGTCTTAGA\n',
'CCAATAGCTGAACTATCAGCTGCAATACCATGGTGACCAGCTGTTACAAAAGATTCTGTTTAT\n',
'CTGAAACATACTGGATTATATATGTATAAGCGCCTCAATGGGAATTAGAGCCAGATGTTATGATTG\n',
'TTGCTTTTCTTTATAGTTAGTTAGCAAAATATGGATAATTCTAGTGAATGCATAAATTAGG\n',
'TTGCGTTCTTATTGCTTAAATCTGAGTTTCCACCCCTGTGACACAATCTAACAGACT\n',
'GTCCTGTAATGGACACAACACAATAAGTCAGTTATTCTGCTGTTACTCTGGATGATATGGAAAACA\n',
'CTGCCATATTAAATCAACTACTCCACGTGTTCCATCCAATCACACTGCTGTGATTCAAGGGATCTT\n',
'TCTTCTAAGACGGACACATTGAAACCTCAGGTTCATCACAAACCTGGTACCTGTTGCTTCCAGAGGATG\n',
'GAGAAGTGTAGTTAATCACACCTTAGTTAATCTGAAATCTGACCCAGTTATTAAACAATAAAC\n',
'CTCATTGATTATATTAAAAGTAATACACTTCTGTAAACAAATGGGGACAATGCATCCAAAATCTT\n',
'TTAAACAGATTACACAAAATTATTCCAGAAAGGCTACCATTTATCATCATTATATTCAAGCCTTTA\n',
'TACTTAATAAGCACTTCTAAAAAGTCTGAGATCCACCATCTGAGGAATTCAATATGTCACCTTTT\n',
'CCTTCTTGCCTGGAGAGGTTAAGAGGAGGTTGAAGGTATAGTCTATTGTTGATGGCCCGCT\n',
'GAATAAAATGGAAATTCTAGTTGTTAGAATTATGCATTCTTCAAGATTCTCAGTGTGCTTAACCTT\n',
'TTGGAGCACATCAGTTCTGGTAATGGAAAACATTACCTAGAGTTGCCAGTGGCACATTACACCAGTA\n',
'CAGAGCACATTCCAAAGGAGACATTGGACCAGTTAATTCCATACAAGTAACAGAACAAAAGGG\n',

'AATCCTGATGCCCTTTACCATTGCTGGTGAGCTAGGCAGTCAGTCATGGACACCCCTAATTTAAAAGG\n',
'TTTAATCATTCTTCTATAAAATACATTAAAATGGAAAAACTTAATATCACTAAATATCAGAACAT\n',
'GTAACATTACAATGACATATTGAAAGCAAAGGCTGTTTATTAGCCAAGATGATTACCTAGGAGT\n',
'TACTTTATGTATTGTTGAAAGCAAATTAAACATGATGTTTAGAAGTGGTCTGATTTTAAACCTGG\n',
'TTTACAGGTATTACTTCTGCACTTACCAAATAATGCCAGATGGAATTATTATTCTGCAATTCCCAT\n',
'GATAGCTCTGTTCTTATGCATTGTCACACTTCCCTTCCAAATGAGTAGAGAATTAAAGC\n',
'CACCCAAAACAGCTCTGCTACTAAATGTTCTACCTTCTCCCTCCTTCCACAAAAA\n',
'GGTAAAAATGAGATCCAATCCTCACCACAAATTCAACCTAGGACACTGGAATGACTGCAGGGATCA\n',
'GTGGTTCTCCATATCACCATCAATTAAAGACATATAGGACACTGTCCTCAAGAGGGTTACAATGTG\n',
'GCCATCAGACAGGAAACCAAACGGTGATAAGTATTAGTAACTAAGTGCACAAATAATGCTGGAAATC\n',
'TTGACCTCTCCTGGGATTATGGGTGTAACAAAATCCCTACATCTGTTATGAAGGCCATATTCAAGTAC\n',
'ATTTAAATGGTAAATAATCTGTTATGTGAAAGAAAAAGAATTAAGTCTTCTTCAACTCTCCTGG\n',
'ATAGCCTAGCACAGTGCAGCCTCCATAACCATGACATTCCGCCAAGCTCAGTGCCTAACCTGCTT\n',
'TGTCAATTACATCTCACAAATCTTGACATCTTACATTCAAGCAAGCACAAGTATGCT\n',
'GGTAGTAGCCTTTAAATAATGTATAGACAACAAACGACAAAAATAGACTGTTAAAGTTCA\n',
'GGGAAAGTTGGTGGCTGATTTAAAGTTGTCAGGAAACATCTCTGTATGAAGCAAATGCGATGTT\n',
'TGAAAAAAGCTAGGAGATGACTTGAATGAATGCAAGGTTAGTGAGATCCTAACGCTCTCAAATAGCATA\n',
'TCCCTAGAGCTCAAGAAAGCTGGCCAGGAGGTTGAAAAAGCTATTGTTGTTAAATTATTTCTGGC\n',
'CCTCTTAATATTTAAAATGTATTCCTTGCGCTTCAACCACCTGCTCAAAAAAAAGAGACTTGT\n',
'ACATGAAAGTTTCATTAAGAGCTGAAAACAAGAATTAGAGAGCCATTCTAGAAAATGCTCTACTGC\n',
'CCTGCATTGACAAACAAGCATCTTACTAACAAAGAGCAGGAATTAGGAGGCCAACGAGCTGAG\n',
'CATGAGCCAAGAGTCTGCTTAATGTTACTTTGAAAATCTGCTGAGCGGCCACCATATGCAGGCTGAG\n',
'AGCTGGGCACAGGCGAAGCCATTGGAAAGCACTTCAGGAACAAGCACACAGCTGTTGGACTTGAACATGCA\n',
'AGTGTTCAGGTTGTCAAGAAGCTTTCTTCTATGATGGAATGTTCTTCTATCCTACTTT\n',
'TTTCTCTTCTCCTCCTCACCACTTACCCCTGCTTACCGAGTAAACGTTTAATGGCCCGTTAT\n',
'GTCTCATGCCTCAAACAACACTGAATTGAAACCCCCATTTCCTCACCCCTGTTGAGCAAT\n',
'TTTCCCAAAAAAAGGGCAGCAATTATTAAATTGAATTCAAGTAAGCCAGCAGCAAGATAGGTCTAAATTG\n',
'CTAGTCCCAGTAGAACCAACCTGATCTAAACCACTGCGAAACAAACAGTAACAATGTCCTCAGCTGACTT\n',
'CAGCTAAGAACCAATGGCTCTACCCCGCCCCGCTTTTTGTTGTTTGTGTTTGAGACGG\n',
'AGCTTGCCTGTCCCCCAGGCTGGAGTCAGTGGCGCAATCTGGGCTCACTGCAACCTCCTCCTCCTC\n',
'CCACATTGAGGCGATTCTCCTGCCTCAGCCTCCAAAGTAGCTGGGATTACAGGCACCCGCATCACACCC\n',
'AGCTAATTTTTTTTTTGTATTAGTAGAAGCCAGGTTCACCATGTTGCCAGGGTGGCTCG\n',
'AACTCCTGACCTCAAGTGATCCGTCCACCTCGGCCCTGCAAATTGCTGGGATTACAGGTGTGAGCCACCG\n',
'TGCCGAGGCCAGCCCCATTAAATGATGTTGGTTAAGAGTGGACCATGAGAATTAGCTGACAGCAT\n',
'CCCCTTCTCTCCTGCCTGGGGACCCCTCCCTGTCAGCTCGAACCTTGTGAACTTGTCAACTTTGTC\n',
'CCGTATTAAAGATGGAGCTGTTTACCTACTTCATAAGACAGTTGCGAGGTGCCATTGATTCTGACTGC\n',
'AAAATACCTGAAACCCCTATATAAGACTGAAGTCACGGAGCCTAGTGAAGAGACTTACTTGTGGCTT\n',
'GTGGTTGAAAGTCACATCAAAGACAATGTGCCACGTTCAGGAATTGGAGACTTACTGGCATGGCTCT\n',
'ACAGCTGCTCAGTTATTAAATCATGCAGACTAACCTGTCACACTGGGAGATGCAACATAGCAAAGGACA\n',
'GAGAAATTAGAATTGGCTGAGAAAGCCCTAAATTCCACCTGAATGTAACCTACAGCTCCCTACCT\n',
'ACTCTCACACATGCCCTCAAACATGCTAGATTGGCTTACATAGGCCAACACAAATACAAACGTGACG\n',
'TGTTCATGTAGCCTAGTGGCTATATGCCTATTCTCCATGTACCCCTGCATGGTAGTGCTGCAAACTTAAA\n',

```
'GTACATTCTTCACAGCACTTTTCTAAGTGCATATAATGTCATTCAATGAAATGGGAAA\n',
'TCACGTTGAGAAGTTGGTCTGTCATCTCCATTGAGCAAAGACTGGCAGGAGATAATAAAATAATG\n',
'GGCACACATGTATTAATATACAGCACGCAAGTTACAGTAAAGTGTCTAGATAACAGCAACTGCTGCTTCGGAGAT\n',
'AGCTCTACCAAGACAGTCTGCAACCATTCCAAGTCTAGTTAACAGCAACTGCTGCTTCGGAGAT\n',
'GGCTGTGAAAATATGGAAGTTCCTCTCAAGTAGGCCAAGAACAGTTCTAGATTTACTAAGTTTATT\n',
'TGTCAGGTTTTAAATTTTCAGTGAGCGTGGTACTGCAGAGGTTAGTGCTGTGAAAGCTGGCTA\n',
'AATATTCTTCTGAAAGTCAAACAGGATTCCATCCCTGTGAAATAACACAAAATTCACTCTCTAAAA\n',
'GCAACAGCATGTAAACTAGAATGAAAGAAGGAAATTATGTACGTATGCCTAATATTCTTGTGAATGTCT\n',
'TTCATTAACTAAAATTATATTAGAAACCAGATTGATAAATAAAAATTCAAAGTAGTTTAATTATCCT\n',
'AAAA\n',
'\n']
```

```
In [80]: if isinstance(secuencia, list):
    secuencia = "".join(secuencia)
```

```
In [81]: secuencia = secuencia.replace("\n", "").replace(" ", "").upper()
```

```
In [82]: gc = secuencia.count("G") + secuencia.count("C")
```

```
In [83]: gc_total = secuencia.count("G") + secuencia.count("C")
```

```
In [84]: gc_porcentaje = (gc / len(secuencia)) * 100
```

```
In [85]: print(gc_porcentaje)%2
```

```
44.02369367463507
```

```
TypeError
```

```
Cell In[85], line 1
```

```
----> 1 print(gc_porcentaje)%2
```

```
Traceback (most recent call last)
```

```
TypeError: unsupported operand type(s) for %: 'NoneType' and 'int'
```

5.- Devuelve la secuencia de ARN que surgiría a partir de la siguiente secuencia. Recuerda que la secuencia de ARN resultante tiene que ser complementaria a la de ADN. (1 pto)

```
In [66]: secuencia
```

Out[66]:

```
'CTTCAAGAACCTACGGAGAGTCGGACTGCATCTCCGAGCGAGCTTTGGAGCGCCGCCGGAGG\n',
'CGAAGGATGCAGGCCGCTCCGGCGCCGGCTGCGGGCAGCGCTCTGCTGAGCTCAGCAGCTGCC\n',
'TCTCAGAGCCTGGACGGCTCCCTCACGTCCAAAATGTATGAGCCACTTGTCTGGACTCCCCCA\n',
'TGTGGCTTCAGCAGCTCCTCCATCTGGTAGCTATTCTCCGGCTATGCCAAGATAAACAAAGAGA\n',
'GGAGGTGCTGGGGATGGTCTCCATCAGACAGCGACCATTATCAATGGCTCAGGTTGACTTGGCAATC\n',
'GGAAGCAGATCAGTGCATTGCAACCCAAGGAAGGTATAGCAGCTCAGATTGGGTGACCCAATACCGGAT\n',
'GCTCTACAGCAGCACACAGGGAGAAACTGGAAACCCATCATCAAGATGGGAATATCTGGGCATTTCCCGGA\n',
'AACATTAACCTGACGGTGTGGTCCGGCACGAATTACAGCATCCGATTATTGCCGCTATGTGCGCATAG\n',
'TGCCTCTGGATTGGAATGGAGAAAGGTCGATTGGACTCAGAATTGAAGTTATGGCTTCTTACTGGC\n',
'TGATGTTATCAACTTGATGGCCATGTTGATTACCATATAGATTAGAAACAAGAAGATGAAACACTG\n',
'AAAGATGTCATTGCCCTGAACTTAAGACGTCTGAAAGTGAAGGAGTAATCCTGCACGGAGAACGGACAGC\n',
'AAGGAGATTACATTACCTGGAACTGAAAAAAAGCCAAGCTGGCCTCAGTTAAACTTAGGAAGCAACCA\n',
'GCTTGGCCCCATATATGCCACACATCAGTGTGACAGGAAGTTGCTGGATGACCCACTGGCACTCT\n',
'GTGGTCATTGAGGCCAGGGCGGAGCATTAACCTCACTCTGGACAGGAGCATGCAGCACTCCGTACCA\n',
'ATGGAGAGTTTGAECTACCTGGACTTGGACTATGAGATAACCTTGGAGGCATCCCTTCTGGCAAGCC\n',
'CAGCTCAGCAGTAGAAAGAATTCAAAGGCTCATGGAAAGCATCAACTACAATGGCGTCAACATTACT\n',
'GATCTTGGCAGAAGGAAGAAATTAGGCCCTCAAATGTGGGAATTGAGCTTTCTTGTGTGGAAACCCCT\n',
'ATACGGTGCCTGTTTCAACGCTACAAGTTACCTGGAGGTGCCGGACGGCTAACCGGACCTGTT\n',
'CTCAGTCAGTTCCAGTTAGGACATGGAAACCCAAATGGTCTCTGGTCTCAGTCACTTGCGGATAAT\n',
'TTGGGCAATGTGGAGATTGACCTCACTGAAAGCAAAGTGGGTGTTCACATCAACATCACAGACCAAGA\n',
'TGAGCCAAATCGATATTCCTCAGGTTCTGGGTGAATGATGGACAGTGGCACGAGGTTGCTTCCTAGC\n',
'CAAGGAAAATTTGCTATTCTCACCATCGATGGAGATGAAGCATCAGCAGTTGAACTAATAGTCCCCTT\n',
'CAAGTTAAAATGGCGAGAAGTACTTTTGAGGTTCTGAACCAGATGAATAACTCAAGTCAGTCACTTG\n',
'TCCTTCAGCCTCATTCAAGGATGCATGCAGCTCATTCAAGTGGACGATCAACTGTAAATTATACGA\n',
'AGTGGCACAAAGGAAGCCGGGAAGTTTGCAGATGTCAGCATTGACATGTTGCGATCATAGACAGATGT\n',
'GTGCCCAATCACTGTGAGCATGGTGGAAAGTGCTCGCAACATGGGACAGCTCAAATGCACTTGTGATG\n',
'AGACAGGATAACAGTGGGCCACCTGCCACAACCTATCTACAGGCCCTCTGTGAAGCCTACAAACACCT\n',
'AGGACAGACATCAAATTATTACTGGATAGATCCTGATGGCAGCGGACCTCTGGGGCCTCTGAAAGTTTAC\n',
'TGCAACATGACAGAGGACAAGTGTGGACCATAGTGTCTCATGACTTGCAGATGCAGACGCCGTGGTCG\n',
'GCTACAACCCAGAAAAACTCAGTGCACAGCTCGTTACAGGCCCTCATGGACAGATAAGTGCAT\n',
'CACTGACAGTGCCGAGTACTGCAGCTAGTGTCTCTATTCTGCAAGATGTCAAGATTGTTAACACC\n',
'CCAGATGGAAGCCCTTACACTTGGTGGTTGCAAGCCAACGAGAACGACTACTACTGGGAGGCTCTG\n',
'GGCCTGGAATCCAGAAATGTGCCCTGGCATCGAACGCAACTGCACAGATCCAAGTACTACTGTAAC\n',
'CGACGCCGACTACAAGCAATGGAGGAAGGATGCTGGTTCTTATCATACAAAGATCACCTGCCAGTGA\n',
'CAAGTGGTGGTTGGAGATACTGACCGTCAAGGCTCAGAAGCCAATTGAGCGTAGGTCTCTGCCTGCC\n',
'AAGGAGACAGGAATTATTGGAATGCCCTCTTCCAAACCCATCTCACCTGCACTTCTACTTT\n',
'CCAAGGGAAACTAGCCTGACATTCTTCTACTTCAAAACATTAACCCCTGGGGAGTGTGAACTT\n',
'AATATGGGAAAGGAAGATTTCATCAAGCTGGAGCTGAAGTCTGCCACAGAAGTGTCTTTCTT\n',
'TGGGAAATGGGCCAGTAGAGATTGAGGTACCAACCCCTCTCAACGATGACCAGTGGCACCGGGT\n',
'CACTGCAGAGAGGAATGTCAAGCAGGCCAGCCTACAGGTGGACGGCTACCGCAGCAGATCCGCAAGGCC\n',
'CCAACAGAAGGCCACACCCGCTGGAGCTCTACAGCCAGTTATTGTGGGTGGCTGGGGGCCAGCAGG\n',
'GCTTCCCTGGCTGCATCCGCTCTTGAGGATGAATGGGTGACACTTGACCTGGAGGAAAGAGCAAAGGT\n',
```

'CACATCTGGTTCATATCCGGATGCTCGGCCATTGCACCAAGCTATGGAACAAACTGTGAAAATGGAGGC\n',
'AAATGCCTAGAGAGATAACCACGGTTACTCCTGCGATTGCTTAATACTGCATATGATGGAACATTTCGA\n',
'ACAAAGATGTTGGTCATTTTGAAAGAAGGGATGTGGCTACGATATAACTTCAGGCCAGCAACAAA\n',
'TGCCAGAGACTCCAGCAGCAGAGTAGACAACGCTCCGACCAGCAGAACCTCCACCCGACCTGGCACAG\n',
'GAGGAGATCCGCTTCAGCTCAGCACCAAGGCAGCCCTGCATTCTCCTACATCAGCTCCTCACCA\n',
'CAGACTTCTGGCAGTCTCGTCAAACCCACTGGAAGCTTACAGATTGATACAACCTGGTGGCACCCG\n',
'AGAGCCATACAATATTGACGTAGACCACAGGAACATGGCCAATGGACAGCCCCACAGTGTCAACATCACC\n',
'CGCCACGAGAAGACCATCTTCTCAAGCTGATCATTACCTCTGTGAGTTACCATCTCCAAGTTCAT\n',
'CCGACACCCCTTCATTCTCCAAGTCGCTCTTCTGGGAAAAGTTAGAAACAGGGAAAATTGACCA\n',
'AGAGATTCAAAATACAACACCCCAGGATTCACTGGTTCGCTCCAGACTTCACCCAGATGCC\n',
'CCTCTCAAGGCCCTTGAGGCAGACAAACGCCTCGGTCACGCCACATCCAGGGCAGCTGGTGGAGT\n',
'CCAAC TGCGGGGCTCGCCGCTGACCCCTCCCCCATGTCGTCGCCACCGACCCCTGGCACCTGGATCA\n',
'CCTGGATTCAAGCCAGTGGATTCCATATAATCCAGGACAAGGCAAGCTATAAGAAATGGAGTCAAC\n',
'AGAAAATCGGCTATCATTGGAGGCGTATTGCTGTGGTATTTCACCATCTGTGCAACCTGGTCTTCC\n',
'TGATCCGTACATGTTCCGCCACAAGGGCACCTACCATACCAACGAAGCAAGGGGGCGAGTCGGCAGA\n',
'GAGCGCGGACGCCCATCATGAAACAACGACCCCAACTTCACAGAGACCATTGATGAAAGCAAAAGGAA\n',
'TGGCTATTGAGGGGTGGCTACTTGGCTATGGGATAGGGAGGAGGAATTACTAGGGAGGAGAGAAAGG\n',
'GACAAAAGCACCTGCTTCATACTCTTGAGCACATCCTAAAATATCAGCACAAGTTGGGGAGGCAGGC\n',
'AATGGAATATAATGGAATATTCTTGAGACTGATCACAAAAAAACCTTTTAATATTCTTATAGC\n',
'TGAGTTTCCCTTCTGTATCAAAACAAAATAACAAAAATGTTAGAGTTAACGAAATGGTTGAA\n',
'TTTGTAGGTACTATCTGCTTATTGTGTGTTAGAGGTGTTAAAGACCCGTGGTAACAGGGCAA\n',
'GTTTCTACGTTTAAGAGCCTTAGAACGTGGTATTCTTGAGAAAAGCTAATGCACCTACA\n',
'GATGGCCCCAACATTCTCTTGTCTAGTCACCTTAATGGGCTTACAGAAACTAGTTCGT\n',
'GTTTATATACTATTCTTGTCTATAAGTCGAAAAGAAAGGGGCAAAGAGAACCTATTATTGC\n',
'CAGTTTTAAGCAGAGCTCAATCTATGCCAGCTCTGGCATCTGGGTTCTGACTGATACCAGCAGT\n',
'GAAGGAAGAGAGTGCATGGCACCTGGTGTAAACGACACAATCAGCACAACTGGAGAGAGGCATTAAAGA\n',
'ACCAGGGAAAGGTAGTTGATTTCATTGAATTCTACAAGCTAATATTGTTCCACGTATGAGTCTTAGA\n',
'CCAATAGCTGAACTATCAGCTGCAATACCATGGTGACCAGCTGTTACAAAAGATTCTGTTTAT\n',
'CTGAAACATACTGGATTATATATGTATAAGCGCCTCAATGGGAATTAGAGCCAGATGTTATGATTG\n',
'TTGCTTTTCTTTATAGTTAGTTAGCAAAATATGGATAATTCTAGTGAATGCATAAATTAGG\n',
'TTGCGTTCTTATTGCTTAAATCTGGTAGTTTCCACCCCTGTGACACAATCTAACAGACT\n',
'GTCCTGTAATGGACACAACACAATAAGTCAGTTATTCTGCTGTTACTCTGGATGATATGGAAAACA\n',
'CTGCCATATTAAATCAACTACTCCACGTGTTCCATCCAATCACACTGCTGTGATTCAAGGGATCTT\n',
'TCTTCTAAGACGGACACATTGAACCTCAGGTTCATCACAAACCTGGTACCTGTTGCTTCCAGAGGATG\n',
'GAGAAGTGTAGTTAATCACACCTTTAGTTAATCTGAAATCTGACCCAGTTATTAAACAATAAAC\n',
'CTCATTGATTATATTAAAAGTAATACACTTCTGTAAACAAATGGGGACAATGCATCCAAAATCTT\n',
'TTAAACAGATTACACAAAATTATTCCAGAAAGGCTACCATTTATCATCATTATATTCAAGCCTTTA\n',
'TACTTAATAAGCACTTCTAAAAAGTCTTGAGATCCACCATCTGAGGAATTCAATATGTCACCTTTT\n',
'CCTTCTTGCCTGGAGAGGTTAAGAGGAGGTTGAAGGTATAGTCTATTGTTGATGGCCCGCT\n',
'GAATAAAATGGAAATTCTAGTTGTTAGAATTATGCATTCTTCAAGATTCTCAGTGTGCTTAACCTT\n',
'TTGGAGCACATCAGTTCTGGTAATGGAAAACATTACCTAGAGTTGCCAGTGGCACATTACACCAGTA\n',
'CAGAGCACATTCCAAAGGAGACATTGGACCAGTTAATTCCATACAAGTAACAGAACAAAAGGG\n',

'AATCCTGATGCCCTTTACCATTGCTGGTGAGCTAGGCAGTCAGTCATGGACACCCCTAATTTAAAAGG\n',
'TTTAATCATTCTTCTATAAAATACATTAAAATGGAAAAACTTAATATCACTAAATATCAGAACAT\n',
'GTAACATTACAATGACATATTGAAAGCAAAGGCTTTTATTAGCCAAGATGATTACCTAGGAGT\n',
'TACTTTATGTATTGTTGAAAGCAAATTAAACATGATGTTTAGAAGTGTCTGATTTTAAACCTGG\n',
'TTTACAGGTATTACTTCTGCACTTACCAAATAATGCCAGATGGAATTATTATTCTGCAATTCCCAT\n',
'GATAGCTCTGTTCTTATGCATTGTCACACTTCCCTTCCAAATGAGTAGAGAATTAAAGC\n',
'CACCCAAAACAGCTCTGCTACTAAATGTTCTACCTTCTCCCTCCTTCCACAAAAA\n',
'GGTAAAAATGAGATCCAATCCTCACCACAAATTCAACCTAGGACACTGGAATGACTGCAGGGATCA\n',
'GTGGTTCTCCATATCACCATCAATTAAAGACATATAGGACACTGTCCTCAAGAGGGTTACAATGTG\n',
'GCCATCAGACAGGAAACCAAACGGTGGATAAGTATTAGTAACTAAGTGCACAAATAATGCTGGAAATC\n',
'TTGACCTCTCCTGGGATTATGGGTGTAACAAAATCCCTACATCTGTTATGAAGGCCATATTCAAGTAC\n',
'ATTTAAATGGTAAATAATCTGTTATGTGAAAGAAAAAGAATTAAAGTCTTCTTCAACTCTCCTGG\n',
'ATAGCCTAGCACAGTGCAGCCTCCATAACCATGACATTCCGCCAAGCTCAGTGCCTAACCTGCTT\n',
'TGTCAATTACATCTCACAAATCTTGACATCTTACATTCAAGCAAGCACAAGTATGCT\n',
'GGTAGTAGCCTTTAAATAATGTATAGACAACAAACGACAAAAATAGACTGTTAAAGTTCA\n',
'GGGAAAGTTGGTGGCTGATTTAAAGTTGTCAGGAAACATCTCTGTATGAAGCAAATGCGATGTT\n',
'TGAAAAAAGCTAGGAGATGACTTGAATGAATGCAAGGTTAGTGAGATCCTAACGCTCTCAAATAGCATA\n',
'TCCCTAGAGCTCAAGAAAGCTGGCCAGGAGGTTGAAAAAGCTATTGTTGTTAAATTATTTCTGGC\n',
'CCTCTTAATATTTAAAATGTATTCCTTGCGCTTCAACCACCTGCTCAAAAAAAAGAGACTTGT\n',
'ACATGAAAGTTTCATTAAGAGCTGAAAACAAGAATTAGAGAGCCATTCTAGAAAATGCTCTACTGC\n',
'CCTGCATTGACAAACAAGCATCTTACTAACAAAGAGCAGGAATTAGGAGGACAAGAAAAAGCATTGG\n',
'CATGAGCCAAGAGTCTGCTTAATGTTACTTTGAAAATCTGCTGAGCGGCCACCATATGCAGGCTGAG\n',
'AGCTGGGCACAGGCGAAGCCATTGGAAGCACTTCAGGAACAAGCACACAGCTGTTGGACTTGAACATGCA\n',
'AGTGTTCAGGTTGTCAAGAAGCTTTCTTCTATGATGGAATGTTCTTCTATCCTACTTT\n',
'TTTCTCTTCTCCTCCTCACCACTTACCTGCTTACGCACTAAACGTTTAATGGCCCGTTTAT\n',
'GTCTCATGCCTCAAACAACACTGAATTGAAACCCCCCATTTTCTTCAACCACCTGTTGAGCAAT\n',
'TTTCCCAAAAAAAGGGCAGCAATTAAATTGAATTCAAGTAAGCCAGCAGCAAGATAGGTCTAAATTG\n',
'CTAGTCCCAGTAGAACCAACCTGATCTAAACCACTGCGAAACAAACAGTAACAATGTCCTCAGCTGACTT\n',
'CAGCTAAGAACCAATGGCTCTACCCCCGCCCGCTTTTGTGTTTTGTTGAGACGG\n',
'AGCTTGCCTGTCCCCCAGGCTGGAGTCAGTGGCGCAATCTGGGCTCACTGCAACCTCCTCCTCCTC\n',
'CCACATTGAGGCGATTCTCCTGCCTCAGCCTCCAAAGTAGCTGGATTACAGGCACCCGCATCACACCC\n',
'AGCTAATTTTTTTTTTGTATTAGTAGAAGCCAGGTTCACCATGTTGCCAGGGTGGCTCG\n',
'AACTCCTGACCTCAAGTGATCCGTCCACCTCGGCCCTGCAAATTGCTGGGATTACAGGTGTGAGCCACCG\n',
'TGCCGAGGCCAGCCCCATTTTAAATGATGTTGGTTAAGAGTGGACCATGAGAATTAGCTGACAGCAT\n',
'CCCCTTCTCTCCTGCCTGGGGACCCCTCCCTGTCAGCTCGAACCTTGTGAACTTCTGAACTTTGTC\n',
'CCGTATTAAAGATGGAGCTGTTTACCTACTTCATAAGACAGTTGCGAGGTGCCATTGATTCTGACTGC\n',
'AAAATACCTGAAACCCCTATATAAGACTGAAGTCACGGAGCCTAGTGAAGAGACTTACTTGTGGCTT\n',
'GTGGTTGAAAGTCACATCAAAGACAAATGTGCCACGTTCAGGAATTGGAGACTTACTGGCATGGCTCT\n',
'ACAGCTGCTCAGTTATTAAATCATGCAGACTAACCTGTCACACTGGGAGATGCAACATAGCAAAAGGACA\n',
'GAGAAATTAGAATTGGCTATGCAAGAAAGCCCTAAATTCCACCTGAATGAACTTACAGCTCCCTACCT\n',
'ACTCTCACACATGCCCTCAAACATGCTAGATTGGCTTACATAGGCCAACACAAATACAAACGTGACG\n',
'TGTTCATGTAGCCTAGTGGCTATATGCCTATTCTCCATGTACCCCTGCATGGTAGTGCTGCAAACTTAAA\n',

```
'GTACATTCTTCACAGCACTTTTCTAAGTCATATAATGTCATTCAATGAAATGGGAAA\n',
'TCACGTTGAGAAGTTGGTCTGTCATCTCCATTGAGCAAAGACTGGCAGGAGATAATAAAATAATG\n',
'GGCACACATGTATTAATATACAGCACGCATTACAAGTTATTTCCAGATAAAATTGTCTATAAGAAC\n',
'AGCTCTACCAAGACAGTCTGCACCATTCCAAGTCTCAGTTAATTCAGCAACTGCTGCTTCGGAGAT\n',
'GGCTGTGAAAATATGGAAGTCCCTCTCAAGTAGGCCAAGAACAGTTCTAGATTTACTAAGTTTATT\n',
'TGTCAGGTTTTAAATTTTCAGTGAGCGTGGTAGTGCTGTGAAAGCTGGCTA\n',
'AATATTCTTCTGAAAGTCAAACAGGATTCCATCCCTGTGAAATAACACAAAATTCACTCTCTAAAA\n',
'GCAACAGCATGTAAACTAGAATGAAAGAAGGAAATTATGTACGTATGCCTAATATTCTTGTGAATGTCT\n',
'TTCATTAACTAAAATTATATTAGAAACCAGATTGATAAATAAAAATTCAAAGTAGTTTAATTATCCT\n',
'AAAA\n',
'\n']
```

In [72]: complemento = {"A":"U", "T":"A", "G":"C", "C":"G"}

In [75]: if isinstance(secuencia, list):
 secuencia = "".join(secuencia)

In [76]: secuencia = secuencia.replace("\n", "").replace(" ", "").upper()

In [77]: arn = "".join(complemento[base] for base in secuencia)

In [78]: print("ARN complementario (100 primeros nt):", arn[:100], "...")

ARN complementario (100 primeros nt): GAAGUUUCUUGGAUGCCUCAGCCUGACGUAGAGGCGUCGCUCGAGAACCU CGCGCGGCCUC CGCUUCCUACGUCCGCCGAGGCGCGCGCG ...

6.- Realiza una función que realice una digestión con las siguientes enzimas de restricción en nuestra secuencia. La función tiene que devolver la siguiente información: (2 pto)

1. Enzimas que tienen sitios de corte en la secuencia y posición (o posiciones de corte)
2. Enzimas que no tienen sitios de corte (si las hay)
3. Resultado de la digestión: los fragmentos resultantes de nuestra secuencia
4. Indicar número y tamaño de cada fragmento

In [14]: enzR = {
 'HindIII' : 'A*AGCTT',
 'BamHI' : 'G*GATCC',
 'AluI' : 'AG*CT',
 'Sau3AI' : '*GATC',
 'EcoRI' : 'G*AATTC'

}

```
In [15]: def encontrar_cortes(secuencia, enzimas):
    secuencia = secuencia.upper()
    enzimas_con_corte = []
    enzimas_sin_corte = []
    cortes = []

    for enzima, sitio in enzimas.items():
        sitio_limpio = sitio.replace('*', '')
        pos_corte_rel = sitio.index('*')

        posiciones = []
        start = 0
        while True:
            idx = secuencia.find(sitio_limpio, start)
            if idx == -1:
                break
            posiciones.append(idx + pos_corte_rel)
            start = idx + 1

        if posiciones:
            enzimas_con_corte[enzima] = posiciones
            cortes.extend(posiciones)
        else:
            enzimas_sin_corte.append(enzima)

    return enzimas_con_corte, enzimas_sin_corte, sorted(set(cortes))
```

```
In [16]: def construir_fragmentos(secuencia, cortes):
    fragmentos = []
    inicio = 0
    for c in cortes:
        fragmentos.append(secuencia[inicio:c])
        inicio = c
    fragmentos.append(secuencia[inicio:])
    tamanos = [len(f) for f in fragmentos]
    return fragmentos, tamanos
```

```
In [17]: def imprimir_reporte(enzimas_con_corte, enzimas_sin_corte, fragmentos, tamanos):
    print("◆ Enzimas que cortan:")
```

```
for e, pos in enzimas_con_corte.items():
    print(f" - {e}: posiciones {pos}")
print("\n◆ Enzimas que no cortan:", enzimas_sin_corte)
print("\n◆ Fragmentos resultantes:")
for i, frag in enumerate(fragmentos, 1):
    print(f" Fragmento {i} ({len(frag)} bp): {frag}")
print("\n◆ Resumen:")
print(f" Número total de fragmentos: {len(fragmentos)}")
print(f" Tamaños: {tamanos}")
```

```
In [18]: def digerir_adn(secuencia, enzimas):
    enzimas_con_corte, enzimas_sin_corte, cortes = encontrar_cortes(secuencia, enzimas)
    fragmentos, tamanos = construir_fragmentos(secuencia, cortes)
    imprimir_reporte(enzimas_con_corte, enzimas_sin_corte, fragmentos, tamanos)
    return enzimas_con_corte, enzimas_sin_corte, fragmentos, tamanos
```

```
In [19]: seq = "AAGCTTGGATTCCGATCGGATCCAGCTTAGCTGAATTC"
digerir_adn(seq, enzR)
```

- ◆ Enzimas que cortan:
 - HindIII: posiciones [1]
 - BamHI: posiciones [19]
 - AluI: posiciones [3, 26, 31]
 - Sau3AI: posiciones [14, 19]
 - EcoRI: posiciones [8, 34]

- ◆ Enzimas que no cortan: []

- ◆ Fragmentos resultantes:
 - Fragmento 1 (1 bp): A
 - Fragmento 2 (2 bp): AG
 - Fragmento 3 (5 bp): CTTGG
 - Fragmento 4 (6 bp): AATTCC
 - Fragmento 5 (5 bp): GATCG
 - Fragmento 6 (7 bp): GATCCAG
 - Fragmento 7 (5 bp): CTTAG
 - Fragmento 8 (3 bp): CTG
 - Fragmento 9 (5 bp): AATTC

- ◆ Resumen:
 - Número total de fragmentos: 9
 - Tamaños: [1, 2, 5, 6, 5, 7, 5, 3, 5]

```
Out[19]: ({'HindIII': [1],  
           'BamHI': [19],  
           'AluI': [3, 26, 31],  
           'Sau3AI': [14, 19],  
           'EcoRI': [8, 34]},  
          [],  
          ['A', 'AG', 'CTTGG', 'AATTCC', 'GATCG', 'GATCCAG', 'CTTAG', 'CTG', 'AATTC'],  
          [1, 2, 5, 6, 5, 7, 5, 3, 5])
```

In []:

7.- Realiza una función que mediante la introducción de una secuencia de ADN devuelva lo siguiente: (3 pto) Todos los ORFs posibles que puedan surgir de la secuencia. Selecciona los 6 ORFs más largos e indica si corresponden a la strand positiva o negativa. Devuelve como mejor predicción, el más largo de todos. Calcula el porcentaje de aminoácidos aromáticos de esa proteína elegida y su longitud relativa a la secuencia original de ADN.

pip install biopython

```
In [31]: from Bio.Seq import Seq  
from Bio.SeqUtils.ProtParam import ProteinAnalysis
```

```
In [32]: def encontrar_orfs(secuencia):  
    secuencia = secuencia.upper()  
    adn = Seq(secuencia)  
    orfs = []  
  
    # Buscar ORFs en las 6 posibles lecturas  
    for strand, nuc in [(+1, adn), (-1, adn.reverse_complement())]:  
        for frame in range(3):  
            trans = nuc[frame:].translate(to_stop=False)  
            trans_str = str(trans)  
  
            for prot in trans_str.split('*'): # dividir por STOP  
                if len(prot) > 0:  
                    orfs.append({  
                        'proteina': prot,  
                        'longitud': len(prot),  
                        'frame': frame + 1,  
                        'strand': '+' if strand == 1 else '-'  
                    })
```

```
    return orfs
```

```
In [33]: def mostrar_orfs(orfs):
    orfs_ordenados = sorted(orfs, key=lambda x: x['longitud'], reverse=True)
    print("◆ Se encontraron", len(orfs_ordenados), "ORFs posibles.")
    print("◆ Los 6 ORFs más largos son:")
    for i, orf in enumerate(orfs_ordenados[:6], 1):
        print(f" {i}. {orf['longitud']} aa | Frame: {orf['frame']} | Strand: {orf['strand']}")
    return orfs_ordenados
```

```
In [34]: def analizar_mejor_proteina(orfs, secuencia):
    mejor = sorted(orfs, key=lambda x: x['longitud'], reverse=True)[0]

    print("\n◆ Mejor predicción (más largo):")
    print(f" Longitud: {mejor['longitud']} aa | Frame: {mejor['frame']} | Strand: {mejor['strand']}")

    analisis = ProteinAnalysis(mejor['proteina'])
    arom = analisis.aromaticity() * 100
    print(f"◆ Porcentaje de aminoácidos aromáticos: {arom:.2f}%")

    rel = (mejor['longitud'] * 3 / len(secuencia)) * 100
    print(f"◆ Longitud relativa de la proteína: {rel:.2f}% del ADN original")

    return mejor, arom, rel
```

```
In [35]: def analizar_orfs(secuencia):
    orfs = encontrar_orfs(secuencia)
    orfs_ordenados = mostrar_orfs(orfs)
    mejor, arom, rel = analizar_mejor_proteina(orfs_ordenados, secuencia)
    return mejor, orfs_ordenados[:6], arom, rel
```

```
In [36]: seq = "ATGCCATTGTAATGGCCGCTGAAAGGGTGCCCGATAG"
analizar_orfs(seq)
```

◆ Se encontraron 8 ORFs posibles.

◆ Los 6 ORFs más largos son:

1. 13 aa | Frame: 1 | Strand: -
2. 12 aa | Frame: 3 | Strand: +
3. 12 aa | Frame: 2 | Strand: -
4. 12 aa | Frame: 3 | Strand: -
5. 8 aa | Frame: 2 | Strand: +
6. 7 aa | Frame: 1 | Strand: +

◆ Mejor predicción (más largo):

Longitud: 13 aa | Frame: 1 | Strand: -

◆ Porcentaje de aminoácidos aromáticos: 7.69%

◆ Longitud relativa de la proteína: 100.00% del ADN original

/opt/anaconda3/lib/python3.13/site-packages/Bio/Seq.py:2877: BiopythonWarning: Partial codon, len(sequence) not a multiple of three. Explicitly trim the sequence or add trailing N before translation. This may become an error in future.

warnings.warn(

```
Out[36]: ({'proteina': 'LSGTLSAAHYNGH', 'longitud': 13, 'frame': 1, 'strand': '-'},
           {'proteina': 'LSGTLSAAHYNGH', 'longitud': 13, 'frame': 1, 'strand': '-'},
           {'proteina': 'GHCNGPLKGCPI', 'longitud': 12, 'frame': 3, 'strand': '+'},
           {'proteina': 'YRAPFQRPIITMA', 'longitud': 12, 'frame': 2, 'strand': '-'},
           {'proteina': 'IGHPFSGPLQWP', 'longitud': 12, 'frame': 3, 'strand': '-'},
           {'proteina': 'WAAERVPD', 'longitud': 8, 'frame': 2, 'strand': '+'},
           {'proteina': 'MAIVMGR', 'longitud': 7, 'frame': 1, 'strand': '+'}],
           7.6923076923076925,
           100.0)
```