

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 introdúcela 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. ¿Cual 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',  
'CTTCAAGAACCCTACGGAGAGTCGGACTGCATCTCCGCAGCGAGCTCTTGGAGCGCCGCCGGCCGGGAGG\n',  
'CGAAGGATGCAGGCGGCTCCGCGCGCCGGCTGCGGGGCAGCGCTCCTGCTGTGGATTGTCAGCAGCTGCC\n',  
'TCTGCAGAGCCTGGACGGCTCCCTCCACGTCCCAAAAATGTGATGAGCCACTTGTCTCTGGACTCCCCCA\n',  
'TGTGGCTTTTCAGCAGCTCCTCCTCCATCTCTGGTAGCTATTCTCCCGGCTATGCCAAGATAAACAAGAGA\n',  
'GGAGGTGCTGGGGGATGGTCTCCATCAGACAGCGACCATTATCAATGGCTTCAGGTTGACTTTGGCAATC\n',  
'GGAAGCAGATCAGTGCCATTGCAACCCAAGGAAGGTATAGCAGCTCAGATTGGGTGACCCAATACCGGAT\n',  
'GCTCTACAGCGACACAGGGAGAACTGGAAACCTATCATCAAGATGGGAATATCTGGGCATTTCCCGGA\n',  
'AACATTAACCTCTGACGGTGTGGTCCGGCACGAATTACAGCATCCGATTATTGCCCGCTATGTGCGCATAG\n',  
'TGCCTCTGGATTGGAATGGAGAAGGTCGCATTGGACTCAGAATTGAAGTTTATGGCTGTTCTTACTGGGC\n',  
'TGATGTTATCAACTTTGATGGCCATGTTGTATTACCATATAGATTCAGAAACAAGAAGATGAAAACACTG\n',  
'AAAGATGTCATTGCCTTGAACTTTAAGACGTCTGAAAGTGAAGGAGTAATCCTGCACGGAGAAGGACAGC\n',  
'AAGGAGATTACATTACCTTGAACTGAAAAAAGCCAAGCTGGTCTCAGTTTAACTTAGGAAGCAACCA\n',  
'GCTTGCCCCATATATGGCCACACATCAGTGATGACAGGAAGTTTGCTGGATGACCACCACTGGCACTCT\n',  
'GTGGTCATTGAGCGCCAGGGGCGGAGCATTAACTCACTCTGGACAGGAGCATGCAGCACTTCCGTACCA\n',  
'ATGGAGAGTTTGACTACCTGGACTTGGACTATGAGATAACCTTTGGAGGCATCCCTTTCTCTGGCAAGCC\n',  
'CAGCTCCAGCAGTAGAAAGAATTTCAAAGGCTGCATGGAAAGCATCAACTACAATGGCGTCAACATTACT\n',  
'GATCTTGCCAGAAGGAAGAAATTAGAGCCCTCAAATGTGGGAAATTTGAGCTTTTCTTGTGTGGAACCCT\n',  
'ATACGGTGCCTGTCTTTTTCAACGCTACAAGTTACCTGGAGGTGCCCGGACGGCTTAACCAGGACCTGTT\n',  
'CTCAGTCAGTTTCCAGTTTAGGACATGGAACCCAATGGTCTCCTGGTCTTCAGTCACTTTGCGGATAAT\n',  
'TTGGGCAATGTGGAGATTGACCTCACTGAAAGCAAAGTGGGTGTTACATCAACATCACACAGACCAAGA\n',  
'TGAGCCAAATCGATATTTCTCAGGTTCTGGGTTGAATGATGGACAGTGGCACGAGGTTTCGCTTCCTAGC\n',  
'CAAGGAAAAATTTTGCTATTCTCACCATCGATGGAGATGAAGCATCAGCAGTTCGAACATAATAGTCCCCTT\n',  
'CAAGTTAAAACTGGCGAGAAGTACTTTTTTGGAGGTTTTCTGAACCAGATGAATAACTCAAGTCACTCTG\n',  
'TCCTTCAGCCTTCATTCCAAGGATGCATGCAGCTCATTCAAGTGGACGATCAACTTGTAATTTATACGA\n',  
'AGTGGCACAAAGGAAGCCGGGAAGTTTCGCGAATGTCAGCATTGACATGTGTGCGATCATAGACAGATGT\n',  
'GTGCCCAATCACTGTGAGCATGGTGGAAAGTGCTCGAAACATGGGACAGCTTCAAATGCACTTGTGATG\n',  
'AGACAGGATACAGTGGGGCCACCTGCCACAACCTCTATCTACGAGCCTTCTGTGAAGCCTACAAACACCT\n',  
'AGGACAGACATCAAATTATTACTGGATAGATCCTGATGGCAGCGGACCTCTGGGGCCTCTGAAAGTTTAC\n',  
'TGCAACATGACAGAGGACAAAGTGTGGACCATAGTGTCTCATGACTTGAGATGCAGACGCCTGTGGTCG\n',  
'GCTACAACCCAGAAAAATACTCAGTGACACAGCTCGTTTACAGCGCCTCCATGGACCAGATAAGTGCCAT\n',  
'CACTGACAGTGCCGAGTACTGCGAGCAGTATGTCTCCTATTTCTGCAAGATGTCAAGATTGTTGAACACC\n',  
'CCAGATGGAAGCCCTTACACTTGGTGGGTTGGCAAAGCCAACGAGAAGCACTACTACTGGGGAGGCTCTG\n',  
'GGCCTGGAATCCAGAAATGTGCCTGCGGCATCGAACGCAACTGCACAGATCCCAAGTACTACTGTAAGT\n',  
'CGACGCGGACTACAAGCAATGGAGGAAGGATGCTGGTTTCTTATCATACAAAGATCACCTGCCAGTGAGC\n',  
'CAAGTGGTGGTTGGAGATACTGACCGTCAAGGCTCAGAAGCCAAATTGAGCGTAGGTCTCTGCGCTGCC\n',  
'AAGGAGACAGGAATTATTGGAATGCCGCCTCTTTCCCAAAACCATCCTCCTACCTGCACTTCTTACTTT\n',  
'CCAAGGGGAACTAGCGCTGACATTTCTTTCTACTTCAAAACATTAACCCCTGGGGAGTGTTTCTTGAA\n',  
'AATATGGGAAAGGAAGATTTTCATCAAGCTGGAGCTGAAGTCTGCCACAGAAGTGCTTTTTCTTTGATG\n',  
'TGGGAAATGGGCCAGTAGAGATTGTAGTGAGGTACCAACCCCTCTCAACGATGACCAGTGGCACCGGGT\n',  
'CACTGCAGAGAGGAATGTCAAGCAGGCCAGCCTACAGGTGGACCGGCTACCGCAGCAGATCCGCAAGGCC\n',  
'CCAACAGAAGGCCACACCCGCCTGGAGCTCTACAGCCAGTTATTTGTGGGTGGTCTGGGGGCCAGCAGG\n']
```

'GCTTCCTGGGCTGCATCCGCTCCTTGAGGATGAATGGGGTGACACTTGACCTGGAGGAAAGAGCAAAGGT\n',
'CACATCTGGGTTTCATATCCGGATGCTCGGGCCATTGCACCAGCTATGGAACAACTGTGAAAATGGAGGC\n',
'AAATGCCTAGAGAGATACCACGGTTACTCCTGCGATTGCTCTAATACTGCATATGATGGAACATTTTGCA\n',
'ACAAAGATGTTGGTGCATTTTTTTGAAGAAGGGATGTGGCTACGATATAACTTTTCAGGCACCAGCAACAAA\n',
'TGCCAGAGACTCCAGCAGCAGAGTAGACAACGCTCCCGACCAGCAGAACTCCCACCCGGACCTGGCACAG\n',
'GAGGAGATCCGCTTCAGCTTCAGCACCACCAAGGCGCCCTGCATTCTCCTCTACATCAGCTCCTTCACCA\n',
'CAGACTTCTTGGCAGTCCTCGTCAAACCCACTGGAAGCTTACAGATTCGATACAACCTGGGTGGCACCCG\n',
'AGAGCCATACAATATTGACGTAGACCACAGGAACATGGCCAATGGACAGCCCCACAGTGTCAACATCACC\n',
'CGCCACGAGAAGACCATCTTTCTCAAGCTCGATCATTATCCTTCTGTGAGTTACCATCTGCCAAGTTCAT\n',
'CCGACACCCTCTTCAATTCTCCCAAGTCGCTCTTTCTGGGAAAAGTTATAGAAACAGGGAAAATTGACCA\n',
'AGAGATTCACAAATACAACACCCCAGGATTCACTGGTTGCCTCTCCAGAGTCCAGTTCAACCAGATCGCC\n',
'CCTCTCAAGGCCGCTTGAGGCAGACAAACGCTCGGCTCACGTCCACATCCAGGGCGAGCTGGTGGAGT\n',
'CCAAGTGCAGGGGCTCGCCGCTGACCCTCTCCCCCATGTGCTCCGCCACCGACCCCTGGCACCTGGATCA\n',
'CCTGGATTACAGCCAGTGCAGGATTTTCCATATAATCCAGGACAAGGCCAAGCTATAAGAAATGGAGTCAAC\n',
'AGAAACTCGGCTATCATTGGAGGCGTCATTGCTGTGGTGATTTTACCATCCTGTGCACCCTGGTCTTCC\n',
'TGATCCGGTACATGTTCCGCCACAAGGGCACCTACCATACCAACGAAGCAAAGGGGGCGGAGTGGCAGA\n',
'GAGCGCGGACGCCGCCATCATGAACAACGACCCCACTTACAGAGACCATTGATGAAAGCAAAAAGGAA\n',
'TGGCTCATTTGAGGGGTGGCTACTTGGCTATGGGATAGGGAGGAGGGAATTACTAGGGAGGAGAGAAAGG\n',
'GACAAAAGCACCCCTGCTTCATACTCTTGAGCACATCCTTAAAAATATCAGCACAAGTTGGGGGAGGCAGGC\n',
'AATGGAATATAATGGAATATTCTTGAGACTGATCACAAAAAAAAAACCTTTTTAATATTTCTTTATAGC\n',
'TGAGTTTTCCCTTCTGTATCAAAACAAAATAATACAAAAATGCTTTTAGAGTTTAAGCAATGGTTGAAA\n',
'TTTGTAGGTACTATCTGTCTTATTTTGTGTGTGTTAGAGGTGTTCTAAAGACCCGTGGTAACAGGGCAA\n',
'GTTTTCTACGTTTTTAAGAGCCCTTAGAACGTGGGTATTTTTTTCTTGAGAAAAGCTAATGCACCTACA\n',
'GATGGCCCCAACATTCTCTTCTTTGCTTCTAGTCAACCTTAATGGGCTGTTACAGAACTAGTTCGT\n',
'GTTTATATACTATTTCTTTGATGTCCTATAAGTCGGAAAAAGAAAGGGGCAAAGAGAACCTATTATTTGC\n',
'CAGTTTTTAAGCAGAGCTCAATCTATGCCAGCTCTCTGGCATCTGGGGTTCCTGACTGATACCAGCAGTT\n',
'GAAGGAAGAGAGTGCATGGCACCTGGTGTGTAACGACACAATCAGCACAACCTGGAGAGAGGCATTAAAGA\n',
'ACCAGGGAAGGTAGTTTGATTTTTCATTGAATTCTACAAGCTAATATTGTTCCACGTATGTAGTCTTAGA\n',
'CCAATAGCTGTAACCTATCAGCTGCAATACCATGGTGACCAGCTGTTACAAAAGATTTTTTCTGTTTTAT\n',
'CTGAAACATACTGGATTTATATATGTATAAGCGCCTCAATGGGGAATTAGAGCCAGATGTTATGATTTGT\n',
'TTGCTCTTTTTCTTTTATAGTTTAGTTATAGCAAAAATATGGATAATTTCTAGTGAATGCATAAATTAGG\n',
'TTGCGTTTCTTATTTTGCTTTAAATCTCTGGTAGTTTTTCCACCCCTGTGACACAATCCTAATAGACAGT\n',
'GTCCTGTAAATGGACACAACACAATAAAGTCAAGTTATTATTGCTGTTACTCTGGATGATATGGAAAACA\n',
'CTGCCATATTTTAAATCACTACTCCACGTGTTTTTCCATCCAATCACACTGCTGTGATTGAGGGATCTT\n',
'TCTTCTAAGACGGACACATTTGAACCTCAGGTTTCATCACAAACCTGGTACCTGTTGCTTCCCAGAGGATG\n',
'GAGAAGTGTAGTTAATCACACCTCTTAGTTTAAATCTGAAATCTTGACCCAGTTATTTAACAAATAAATAC\n',
'CTCATTGATTATATTTAAAAGTAATACACTTCCTGTAAACAAATGGGGACAATGCATCCAAAAAATCTTT\n',
'TTAAACAGATTACACAAAAATTATTTCCAGAAAAGGCTACCATTTATCATCATTATATTTCAAGCCTCTTA\n',
'TACTTAATAAGCACTTTCTAAAAAGTCTTGAGATCCCACCATCTGAGGAATTCAATATGATCACTTTTT\n',
'CCTTCTTTGCCTGGGAGAGGTTAAGAGGAGTTTCGAAGGTATAGATGCTATTGTTCTGATGGCCCGGCT\n',
'GAATAAAATGGAAATTCTAGTTTGTTAGAATTATGCATTCTTTTTCAAGATTCTCAGTGTGCCTAACTTA\n',
'TTGAGACACATCAGTTTCTTGGGTAATGGAAAACATTACCTAGAGTTGCCAGTGGCACATTACACCAGTA\n',

'CAGAGCACATTCCAAAGGAGACATTGGACCAGTTAATTCCCATACAAGTCAAGGTAACAGAACAAAAGGG\n',
'AATCCTGATGCCCTTTTACCATTGCTGGTTGAGCTCAGGCACTGTCATGGACACCCTTAATTTTAAAAGG\n',
'TTTTAATCATTCTTCTATAAAATACATTTAAATGGAAAAATACTTAATATCACTAAATATCAGAACAAT\n',
'GTAACATTTACAAATGACATATTGAAAGCAAAGGCTGTTTTATTTAGCCAAGATGATTACCATTAGGAGT\n',
'TACTTTATGTATTGTTGAAAGCAAATTTTAAACATGATGTTTTAGAAGTGTCTGATTTTTTAAACCTGG\n',
'TTTACAGGTATTACTTCTGCACCTACCAAATAATGCCAGATGGAAATTTATTATTTCTTGCAATTCCTAT\n',
'GATAGCTCTGTTCTTTATGCATTGTCTCAACACTTTCCTTTTTTCCCAAATGAGTAGAGAATTAAGC\n',
'CACCCAAAACAGCTTCTGCTACTAAAATGTTCTCATCCTTCTCCTCCTCTCCTTTTCTGCCACAAAA\n',
'GGTGAAAAATGAGATCCAATCCTCTCACCAAAATTTCAAACCTAGGACACTGGAATGACTGCAGGGATCA\n',
'GTGGTTCTCCCATATCACCATCAATTAAGACATATAGGACACTGTCTTCCTTCAAGAGGGTTACAATGTG\n',
'GCCATCAGACAGGAAACCAAACGGTGGATAAAGTATTAAGTAACTAAGTGCCAAATAAATGCTGGAAATC\n',
'TTGACCTCTCCTTGGGATTATGGGTGTAACAAAAATCCCTACATCTGTTTATGAAGGCCATATTCAGTAC\n',
'ATTTTAAATGGTAAATAATCTGTTTATGTGAAGAAAAAGAATTAAGTCTTCTTCCAACCTCTCCTTGG\n',
'ATAGCCTAGCACAGTGCAGCCTCCATAACCATGACATTCCTCGCCCAAGCTCTCAGTGCCTAATCCTGCTT\n',
'TGTCATTACATCTCACAAAATCTTGACATCTTACATTCGAATACATTATCAAGCAAGCACAAAGTATGCT\n',
'GGTAGTAGCCTCTTTAAATAATATGTATAGACAACAACAACGACAAAAAATAGACTGTTTTAAAGTTTCA\n',
'GGGAAAGTTGGTGGCTGATTTAAAGTTGTGCAGGAAACATCTTCTGTGTATGAAGCAAATGTCGATGTTT\n',
'TGAAAAAAGCTAGGAGATGACTTTGAATGAATGCAAGGTTAGTGAGATCCTAAGCTCTCAAAATAGCATA\n',
'TTCCCTAGAGCTCAAGAAAGCTGGTCCAGGAGGTTGAAAAAGCTATTTTGTTGTTAAATTATTTTCTGGC\n',
'CCTTCTTAATATTTAAAAATGTATTTCCCTTGTGGCTTTCAACCACCTGCTCAAAAAAAGAGACTTGTT\n',
'ACATGAAAGTTTTTATTAAAGAGCTGAAAACAAGAATTTAGAGAGCCATTCTAGAAAATGTCCTACTGC\n',
'CCTGCATTTTGACAAACAAGCATCCTTTACTAACAAGAGCAGGAATTCAGAGGCACAAGAAAAAGCATTGG\n',
'CATGAGCCAAAGAGTCTGTCTTAATGTTACTTTTTGAAAATCTGCTGAGCGGCCACCATATGCAGGCTGAG\n',
'AGCTGGGCACAGGCGAAGCCATTGGAAGCACTTCAGGAACAAGCACACAGCTGTGGGACTTGAACATGCA\n',
'AGTGTTACAGGTTGTGTCAAGAAGCTTTTCTTCTCTATGATGGAATCTGTTCTTTTCTATCCTACTTT\n',
'TTTCTCTCTTCTCCTCCTCACCACATTATACCCTGCTCTTACGCAGTAAACGTTTTAATGGCCCGTTTAT\n',
'GTCTCATGCCTCCAAACAACACTGAATTTGAAACCCCCCATTTTTTCTTTTCAACCACCTGTTGAGCAAT\n',
'TTTCCCAAAAAAAGGGCAGCAATTATTAATTTGAATTCAAGTAAGCCAGCCAAAGATAGGTCCTAAATTG\n',
'CTAGTCCCAGTAGAACCACCTGATCCTAAACCAGTGCGAAACAAACAGTAACAATGTCCCAGCTGACTT\n',
'CAGCTAAGAACCAATGGCTCCTACCCCCGCCCGCTTTTTTTTGTGTTGTTTTGTTTTGTTTTGAGACGG\n',
'AGTCTTGCTCTGTCCCCAGGCTGGAGTGCACCTGGCGCAATCTCGGGCTCACTGCAACCTCCTCCTCCTC\n',
'CCACATTGAGGCGATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGCACCCGCCATCACACCC\n',
'AGCTAATTTTTTTTTTTTTTTTGTATTATTAGTAGAAGCCAGGTTTACCATGTTGGCCAGGGTGGTCTCG\n',
'AACTCCTGACCTCAAGTGATCCGTCCACCTCGGCCTTGCAAATTGCTGGGATTACAGGTGTGAGCCACCG\n',
'TGCCGAGCCAGCCCCATTTTTTAAATGATGTTTTGGTTAAGAGTGGACCATGAGAATTAGCTGACAGCAT\n',
'CCCCCTTCTCTCTCCTGCCTTGGTGGGACCCTCCCTGTGTGACCTTGGTCAAGTCTCGAACTTTTGTC\n',
'CCGTATTTAAGATGGAGCTGTTTTACCTACTTCATAAGACAGTTGCGAGGTGCCATTGATTCTTGACTGC\n',
'AAAATACCTTGAAACCTTATATAAAGACTGAAGTCAACGGAGCCTAGTGAAAGACTTACTTTGTGGCTT\n',
'GTGGTTGAAAGTCACATCAAAAGACAAATGTGGCCACGTTTCAGGAATTGGAGACTTACTGGCATGGCTCT\n',
'ACAGCTGCTCAGTTATTAATCATGCAGACTAACCTGTCAACACTGGGAGATGCAACATAGCAAAAGGACA\n',
'GAGAAATTAGAATTTTTTGTGCAGAAAGCCCTAAATCCCACCTGAATGTAACCTACAGCTCCCTTACCT\n',
'ACTCTCACACATGCCCTCAAACATGCTAGATTGGCTTATACATAGGCCAACACAAAATACAAACGTGACG\n',

```
'TGTTTCATGTAGCCTAGTGGCTATATGCCTATTCTCCATGTACCCTGCATGGTAGTGCTGCAAACCTTTAAA\n',  
'GTACATTTCTTTCACAGCAGTATTTTTTTTCATAAGTGGCATATAAATGTCATTCAATGAAATGGGGAAA\n',  
'TCACGTTGAGAAGTTGGTCTGTCATCTCCCATTGAGCAAAGACTGGCAGGAGATAATAAAAATAAATATG\n',  
'GGCACACATGTATTAATATACAGCACGCATTTACAAGTTTATTTTCCAGATAAAATTGTGCTATAAGAAC\n',  
'AGCTCTACCAAGACAGTCTGCACCATTTCCAAGTCTCAGTTAATTTACAGCAACTGCTGCTTTTCGGAGAT\n',  
'GGCTGTGAAAATATGGAAGTTCCTCTCAAGTAGGCCAAGAAACAGTTCTAGATTTTACTAAGTTTTATTT\n',  
'TGTCAGGTTTTTTAAATTTTTTTCAGTGAGCGTGGTGACTGCAGAGGTTAGTGCTGTGAAAAGCTGGGCTA\n',  
'AATATTCTTTCTGTAAAGTCAAACAGGATTCCATCCCCTGTGAAATAACACAAAATTTCACTCTCTAAAA\n',  
'GCAACAGCATGTAACTAGAATGAAAGAAGGAAATTATGTACGTATGCCTAATATTCTTTGTGAATGTCT\n',  
'TTCATTTAACTAAAATTATATTAGAAACCAGATTGATAAATAAAAAATTCAAAGTAGTTTTAATTATCCT\n',  
'AAAA\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 [61]: ['CTTCAAGAACCCTACGGAGAGTCGGACTGCATCTCCGCAGCGAGCTCTTGGAGCGCCGCCGGCCGGGAGG\n',
'CGAAGGATGCAGGCGGCTCCGCGCGCCGGCTGCGGGGCAGCGCTCCTGCTGTGGATTGTCAGCAGCTGCC\n',
'TCTGCAGAGCCTGGACGGCTCCCTCCACGTCCCAAAAAATGTGATGAGCCACTTGTCTCTGGACTCCCCCA\n',
'TGTGGCTTTCAGCAGCTCCTCCTCCATCTCTGGTAGCTATTCTCCCGGCTATGCCAAGATAAACAAGAGA\n',
'GGAGGTGCTGGGGGATGGTCTCCATCAGACAGCGACCATTATCAATGGCTTCAGGTTGACTTTGGCAATC\n',
'GGAAGCAGATCAGTGCCATTGCAACCCAAGGAAGGTATAGCAGCTCAGATTGGGTGACCCAATACCGGAT\n',
'GCTCTACAGCGACACAGGGAGAACTGGAAACCTATCATCAAGATGGGAATATCTGGGCATTTCCCGGA\n',
'AACATTAACCTCTGACGGTGTGGTCCGGCAGCAATTACAGCATCCGATTATTGCCCGCTATGTGCGCATAG\n',
'TGCCTCTGGATTGGAATGGAGAAGGTGCGATTGGACTCAGAATTGAAGTTTATGGCTGTTCTTACTGGGC\n',
'TGATGTTATCAACTTTGATGGCCATGTTGTATTACCATATAGATTCAGAAACAAGAAGATGAAAACACTG\n',
'AAAGATGTCATTGCCTTGAACTTTAAGACGTCTGAAAGTGAAGGAGTAATCCTGCACGGAGAAGGACAGC\n',
'AAGGAGATTACATTACCTTGAACTGAAAAAAGCCAAGCTGGTCTCAGTTTAACTTAGGAAGCAACCA\n',
'GCTTGCCCCATATATGGCCACACATCAGTGATGACAGGAAGTTTGCTGGATGACCACCACTGGCACTCT\n',
'GTGGTCATTGAGCGCCAGGGGCGGAGCATTAACTCACTCTGGACAGGAGCATGCAGCACTTCCGTACCA\n',
'ATGGAGAGTTTGACTACCTGGACTTGGACTATGAGATAACCTTTGGAGGCATCCCTTCTCTGGCAAGCC\n',
'CAGCTCCAGCAGTAGAAAGAATTTCAAAGGCTGCATGGAAAGCATCAACTACAATGGCGTCAACATTACT\n',
'GATCTTGCCAGAAGGAAGAAATTAGAGCCCTCAAATGTGGGAAATTTGAGCTTTTCTTGTGTGGAACCCT\n',
'ATACGGTGCCTGTCTTTTTCAACGCTACAAGTTACCTGGAGGTGCCCGGACGGCTTAACCAGGACCTGTT\n',
'CTCAGTCAGTTTCCAGTTTAGGACATGGAACCCAATGGTCTCCTGGTCTTCAGTCACTTTGCGGATAAT\n',
'TTGGGCAATGTGGAGATTGACCTCACTGAAAGCAAAGTGGGTGTTACATCAACATCACACAGACCAAGA\n',
'TGAGCCAAATCGATATTTCTCAGGTTCTGGGTTGAATGATGGACAGTGGCACGAGGTTGCTTCTTAGC\n',
'CAAGGAAAATTTTGCTATTCTCACCATCGATGGAGATGAAGCATCAGCAGTTCGAACTAATAGTCCCCTT\n',
'CAAGTTAAAACTGGCGAGAAGTACTTTTTTGGAGGTTTTCTGAACCAGATGAATAACTCAAGTCACTCTG\n',
'TCCTTCAGCCTTCATTCCAAGGATGCATGCAGCTCATTCAAGTGGACGATCAACTTGTAATTTATACGA\n',
'AGTGGCACAAAGGAAGCCGGGAAGTTTCGCGAATGTCAGCATTGACATGTGTGCGATCATAGACAGATGT\n',
'GTGCCCAATCACTGTGAGCATGGTGGAAAGTGCTCGCAAACATGGGACAGCTTCAAATGCACTTGTGATG\n',
'AGACAGGATACAGTGGGGCCACCTGCCACAACCTCTATCTACGAGCCTTCTGTGAAGCCTACAAACACCT\n',
'AGGACAGACATCAAATTATTACTGGATAGATCCTGATGGCAGCGGACCTCTGGGGCCTCTGAAAGTTTAC\n',
'TGCAACATGACAGAGGACAAAGTGTGGACCATAGTGTCTCATGACTTGAGATGCAGACGCCTGTGGTCTG\n',
'GCTACAACCCAGAAAAATACTCAGTGACACAGCTCGTTTACAGCGCCTCCATGGACCAGATAAGTGCCAT\n',
'CACTGACAGTGCCGAGTACTGCGAGCAGTATGTCTCCTATTTCTGCAAGATGTCAAGATTGTTGAACACC\n',
'CCAGATGGAAGCCCTTACACTTGGTGGGTTGGCAAAGCCAACGAGAAGCACTACTACTGGGGAGGCTCTG\n',
'GGCCTGGAATCCAGAAATGTGCCTGCGGCATCGAACGCAACTGCACAGATCCCAAGTACTACTGTAAGT\n',
'CGACGCGGACTACAAGCAATGGAGGAAGGATGCTGGTTTCTTATCATACAAAGATCACCTGCCAGTGAGC\n',
'CAAGTGGTGGTTGGAGATACTGACCGTCAAGGCTCAGAAGCCAAATTGAGCGTAGGTCTCTGCGCTGCC\n',
'AAGGAGACAGGAATTATTGGAATGCCGCCTCTTTCCAAACCCATCCTCCTACCTGCACTTCTCTACTTT\n',
'CCAAGGGGAACTAGCGCTGACATTTCTTTCTACTTCAAAACATTAACCCCTGGGGAGTGTTTCTTGAA\n',
'AATATGGGAAAGGAAGATTTATCAAGCTGGAGCTGAAGTCTGCCACAGAAGTGCTTTTTATTTGATG\n',
'TGGGAAATGGGCCAGTAGAGATTGTAGTGAGGTACCAACCCCTCTCAACGATGACCAGTGGCACCGGGT\n',
'CACTGCAGAGAGGAATGTCAAGCAGGCCAGCCTACAGGTGGACCGGCTACCGCAGCAGATCCGCAAGGCC\n',
'CCAACAGAAGGCCACACCCGCCTGGAGCTCTACAGCCAGTTATTTGTGGGTGGTCTGGGGGCCAGCAGG\n',
'GCTTCTGGGCTGCATCCGCTCCTTGAGGATGAATGGGGTGACACTTGACCTGGAGGAAAGAGCAAAGGT\n']

'CACATCTGGGTTTCATATCCGGATGCTCGGGCCATTGCACCAGCTATGGAACAACTGTGAAAATGGAGGC\n',
'AAATGCCTAGAGAGATACCACGGTTACTCCTGCGATTGCTCTAATACTGCATATGATGGAACATTTTGCA\n',
'ACAAAGATGTTGGTGCATTTTTTTGAAGAAGGGATGTGGCTACGATATAACTTTTCAGGCACCAGCAACAAA\n',
'TGCCAGAGACTCCAGCAGCAGAGTAGACAACGCTCCCGACCAGCAGAACTCCCACCCGGACCTGGCACAG\n',
'GAGGAGATCCGCTTCAGCTTCAGCACCAAGGCGCCCTGCATTCTCCTCTACATCAGCTCCTTCACCA\n',
'CAGACTTCTTGGCAGTCCTCGTCAAACCCACTGGAAGCTTACAGATTTCGATACAACCTGGGTGGCACCCG\n',
'AGAGCCATACAATATTGACGTAGACCACAGGAACATGGCCAATGGACAGCCCCACAGTGTCAACATCACC\n',
'CGCCACGAGAAGACCATCTTTCTCAAGCTCGATCATTATCCTTCTGTGAGTTACCATCTGCCAAGTTCAT\n',
'CCGACACCCTCTTCAATTCTCCCAAGTCGCTCTTTCTGGGAAAAGTTATAGAAACAGGGAAAATTGACCA\n',
'AGAGATTCACAAATACAACACCCCAGGATTCAGTGGTTGCCTCTCCAGAGTCCAGTTCAACCAGATCGCC\n',
'CCTCTCAAGGCCGCCTTGAGGCAGACAAACGCCTCGGCTCACGTCCACATCCAGGGCGAGCTGGTGGAGT\n',
'CCAAGTGCAGGCGCTCGCCGCTGACCCTCTCCCCCATGTCGTCCGCCACCGACCCCTGGCACCTGGATCA\n',
'CCTGGATTGAGCCAGTGCAGGATTTTCCATATAATCCAGGACAAGGCCAAGCTATAAGAAATGGAGTCAAC\n',
'AGAAACTCGGCTATCATTGGAGGCGTCATTGCTGTGGTGATTTTACCATCCTGTGCACCCTGGTCTTCC\n',
'TGATCCGGTACATGTTCCGCCACAAGGGCACCTACCATAACGAAGCAAAGGGGGCGGAGTGGGCAGA\n',
'GAGCGCGGACGCCGCCATCATGAACAACGACCCCAACTTCACAGAGACCATTGATGAAAGCAAAAAGGAA\n',
'TGGCTCATTTGAGGGGTGGCTACTTGGCTATGGGATAGGGAGGAGGGAATTACTAGGGAGGAGAGAAAGG\n',
'GACAAAAGCACCCCTGCTTCATACTCTTGAGCACATCCTTAAAATATCAGCACAAAGTTGGGGGAGGCAGGC\n',
'AATGGAATATAATGGAATATTCTTGAGACTGATCACAAAAAAAAAACCTTTTAAATATTTCTTTATAGC\n',
'TGAGTTTTCCCTTCTGTATCAAAACAAAATAATACAAAAATGCTTTTAGAGTTTAAGCAATGGTTGAAA\n',
'TTTGTAGGTACTATCTGTCTTATTTTGTGTGTGTTAGAGGTGTTCTAAAGACCCGTGGTAACAGGGCAA\n',
'GTTTTCTACGTTTTTAAAGGCCCTTAGAACGTGGGTATTTTTTTTCTTGAGAAAAGCTAATGCACCTACA\n',
'GATGGCCCCAACATTCTCTTCTTTTGCTTCTAGTCAACCTTAATGGGCTGTTACAGAACTAGTTCGT\n',
'GTTTATATACTATTTCTTTGATGTCCTATAAGTCGGAAAAGAAAGGGGCAAAGAGAACCTATTATTTGC\n',
'CAGTTTTTAAGCAGAGCTCAATCTATGCCAGCTCTCTGGCATCTGGGGTTCCTGACTGATACCAGCAGTT\n',
'GAAGGAAGAGAGTGCATGGCACCTGGTGTGTAACGACACAATCAGCACAACTGGAGAGAGGCATTAAAGA\n',
'ACCAGGGAAGGTAGTTTGATTTTTCATTGAATTCTACAAGCTAATATTGTTCCACGTATGTAGTCTTAGA\n',
'CCAATAGCTGTAAGTATCAGCTGCAATACCATGGTGACCAGCTGTTACAAAAGATTTTTCTGTTTTAT\n',
'CTGAAACATACTGGATTTATATATGTATAAGCGCCTCAATGGGGAATTAGAGCCAGATGTTATGATTTGT\n',
'TTGCTCTTTTTCTTTTATAGTTTAGTTATAGCAAAAATATGGATAATTTCTAGTGAATGCATAAATTAGG\n',
'TTGCGTTTCTTATTTTGCTTTAAATCTCTGGTAGTTTTTCCACCCCTGTGACACAATCCTAATAGACAGT\n',
'GTCCTGTAAATGGACACAACACAATAAAGTCAAGTTATTATTGCTGTTACTCTGGATGATATGGAAAACA\n',
'CTGCCATATTTTAAATCACTACTCCACGTGTTTTTCCATCCAATCACACTGCTGTGATTGAGGGATCTT\n',
'TCTTCTAAGACGGACACATTTGAACCTCAGGTTTCATCACAAACCTGGTACCTGTTGCTTCCCAGAGGATG\n',
'GAGAAGTGTAGTTAATCACACCTCTTAGTTTAAATCTGAAATCTTGACCCAGTTATTTAACAATAAATAC\n',
'CTCATTGATTATATTTAAAAGTAATACACTTCCTGTAAACAAATGGGGACAATGCATCCAAAAAATCTTT\n',
'TTAAACAGATTACACAAAAATTATTTCCAGAAAGGCTACCATTTATCATCATTATATTTCAAGCCTCTTA\n',
'TACTTAATAAGCACTTTCTAAAAAGTCTTGAGATCCCACCATTTCTGAGGAATTCAATATGATCACTTTTT\n',
'CCTTCTTTGCCTGGGAGAGGTTAAGAGGAGGTTTCGAAGGTATAGATGCTATTGTTCTGATGGCCCGGCT\n',
'GAATAAAATGGAAATCTAGTTTGTTAGAATTATGCATTCTTTTTCAAGATTCTCAGTGTGCCTAACTTA\n',
'TTGGAGCACATCAGTTTCTTGGGTAATGGAAAACATTACCTAGAGTTGCCAGTGGCACATTACACCAGTA\n',
'CAGAGCACATTCCAAAGGAGACATTGGACCAGTTAATTCCTATACAAGTCAAGGTAACAGAACAAAAGGG\n',

'AATCCTGATGCCCTTTTACCATTGCTGGTTGAGCTCAGGCACTGTCATGGACACCCTTAATTTTAAAAGG\n',
'TTTTAATCATTCTTCTATAAAATACATTTAAATGGAAAAATACTTAATATCACTAAATATCAGAACAAT\n',
'GTAACATTTACAAATGACATATTGAAAGCAAAGGCTGTTTTATTTAGCCAAGATGATTACCATTAGGAGT\n',
'TACTTTATGTATTGTTGAAAGCAAATTTTAAACATGATGTTTTAGAAGTGTCTGATTTTTAAACCTGG\n',
'TTTACAGGTATTACTTCTGCACCTACCAAATAATGCCAGATGGAAATTTATTATTTCTTGCAATTCCCAT\n',
'GATAGCTCTGTTCTTTATGCATTGTCTCAACACTTTCCCTTTTTTCCCAAAATGAGTAGAGAATTAAAGC\n',
'CACCCAAAACAGCTTCTGCTACTAAAATGTTCTCATCCTTCTCCTCCCTCTCCTTTTCTGCCACAAAA\n',
'GGTGAAAAATGAGATCCAATCCTCTCACCAAAATTTCAAACCTAGGACACTGGAATGACTGCAGGGATCA\n',
'GTGGTTCTCCCATATCACCATCAATTAAGACATATAGGACACTGTCTTCTTCAAGAGGGTTACAATGTG\n',
'GCCATCAGACAGGAAACCAAACGGTGGATAAAGTATTAAGTAACTAAGTGCCAAATAAATGCTGGAAATC\n',
'TTGACCTCTCCTTGGGATTATGGGTGTAACAAAAATCCCTACATCTGTTTATGAAGGCCATATTCAGTAC\n',
'ATTTTAAATGGTAAATAATCTGTTTATGTGAAGAAAAAGAATTAAGTCTTCTTCCAACCTCTCTCCTTGG\n',
'ATAGCCTAGCACAGTGCAGCCTCCATAACCATGACATTCCTCGCCCAAGCTCTCAGTGCCTAATCCTGCTT\n',
'TGTCATTACATCTCACAAAATCTTGACATCTTACATTCCAATACATTATCAAGCAAGCACAAGTATGCT\n',
'GGTAGTAGCCTCTTTAAATAATATGTATAGACAACAACAACGACAAAAAATAGACTGTTTTAAAGTTTCA\n',
'GGGAAAGTTGGTGGCTGATTTAAAGTTGTGCAGGAAACATCTTCTGTGTATGAAGCAAATGTCGATGTTT\n',
'TGAAAAAAGCTAGGAGATGACTTTGAATGAATGCAAGGTTAGTGAGATCCTAAGCTCTCAAAATAGCATA\n',
'TTCCCTAGAGCTCAAGAAAGCTGGTCCAGGAGTTGAAAAAGCTATTTTGTTGTTAAATTATTTTCTGGC\n',
'CCTTCTTAATATTTAAAAATGTATTTCCCTTGTGGCTTTCAACCACCTGCTCAAAAAAAGAGACTTGTT\n',
'ACATGAAAGTTTTATTAAAGAGCTGAAAAACAAGATTTAGAGAGCCATTCTAGAAAATGTCCTACTGC\n',
'CCTGCATTTGACAAACAAGCATCCTTTACTAACAAGAGCAGGAATTCAGAGGCACAAGAAAAAGCATTGG\n',
'CATGAGCCAAAGAGTCTGTCTTAATGTTACTTTTAAAAATCTGCTGAGCGGCCACCATATGCAGGCTGAG\n',
'AGCTGGGCACAGGCGAAGCCATTGGAAGCACTTCAGGAACAAGCACACAGCTGTGGGACTTGAACATGCA\n',
'AGTGTTCAAGTTGTGTCAAGAAGCTTTTCTTCTCTATGATGGAATCTGTTCTTTTCTATCCTACTTT\n',
'TTTCTCTCTTCTCTCCTCACCACATTATACCCTGCTCTTACGCAGTAAACGTTTTAATGGCCCGTTTAT\n',
'GTCTCATGCCTCCAAACAACACTGAATTTGAAACCCCCCATTTTTTCTTTTACCACCCTGTTGAGCAAT\n',
'TTTCCCAAAAAAAGGGCAGCAATTATTAATTTGAATTCAAGTAAGCCAGCCAAAGATAGGTCTAAATTG\n',
'CTAGTCCCAGTAGAACCACCTGATCCTAAACCACTGCGAAACAAACAGTAACAATGTCCCCAGCTGACTT\n',
'CAGCTAAGAACCAATGGCTCCTACCCCCGCCCGCTTTTTTTTGTGTTTTTGTGTTTTGAGACGG\n',
'AGTCTTGCTCTGTCCCCAGGCTGGAGTGCACCTGGCGCAATCTCGGGCTCACTGCAACCTCCTCCTCCTC\n',
'CCACATTGAGGCGATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGCACCCGCCATCACACCC\n',
'AGCTAATTTTTTTTTTTTTTTTGTATTATTAGTAGAAGCCAGGTTTACCATGTTGGCCAGGGTGGTCTCG\n',
'AACTCCTGACCTCAAGTGATCCGTCCACCTCGGCCTTGCAAATTGCTGGGATTACAGGTGTGAGCCACCG\n',
'TGCCGAGCCAGCCCCATTTTTTAAATGATGTTTTGGTTAAGAGTGGACCATGAGAATTAGCTGACAGCAT\n',
'CCCCTTCTCTCTCCTGCCTTGGTGGGACCTCCCTGTGTGACCTGGTCAAGTCTCGAACTTTTGTG\n',
'CCGTATTTAAGATGGAGCTGTTTTACCTACTTCATAAGACAGTTGCGAGGTGCCATTGATTCTTGACTGC\n',
'AAAATACCTTGAAACCTTATATAAAGACTGAAGTCAACGGAGCCTAGTGAAAGACTTACTTTGTGGCTT\n',
'GTGGTTGAAAGTCACATCAAAGACAAATGTGGCCACGTTTCAAGGAATTGGAGACTTACTGGCATGGCTCT\n',
'ACAGCTGCTCAGTTATTAATCATGCAGACTAACCTGTCAACACTGGGAGATGCAACATAGCAAAAGGACA\n',
'GAGAAATTAGAATTTTTTGTGCAGAAAGCCCTAAATTTCCACCTGAATGTAACCTTACAGCTCCCTTACCT\n',
'ACTCTCACACATGCCCTCAAACATGCTAGATTGGCTTATACATAGGCCAACACAAAATACAAACGTGACG\n',
'TGTTTCATGTAGCCTAGTGGCTATATGCCTATTCTCCATGTACCCTGCATGGTAGTGCTGCAAACCTTTAAA\n'


```
'GTACATTTCTTTCACAGCAGTATTTTTTTTCATAAGTGGCATATAAATGTCATTCAATGAAATGGGGAAA\n',  
'TCACGTTGAGAAGTTGGTCTGTCATCTCCCATTTGAGCAAAGACTGGCAGGAGATAATAAAAAATAATATG\n',  
'GGCACACATGTATTAATATACAGCACGCATTTACAAGTTTATTTTCCAGATAAAATTGTGCTATAAGAAC\n',  
'AGCTCTACCAAGACAGTCTGCACCATTTCCAAGTCTCAGTTAATTTACAGCAACTGCTGCTTTCGGAGAT\n',  
'GGCTGTGAAAATATGGAAGTTCCTCTCAAGTAGGCCAAGAAACAGTTCTAGATTTTACTAAGTTTTATTT\n',  
'TGTCAGGTTTTTTAAATTTTTTTCAGTGAGCGTGGTGACTGCAGAGGTTAGTGCTGTGAAAAGCTGGGCTA\n',  
'AATATTCTTTCTGTAAAGTCAAACAGGATTCCATCCCCTGTGAAATAACACAAAATTTCACTCTCTAAAA\n',  
'GCAACAGCATGTAACTAGAAATGAAAGAAGGAAATTATGTACGTATGCCTAATATTCTTTGTGAATGTCT\n',  
'TTCATTTAACTAAAATTATATTAGAAACCAGATTGATAAATAAAAAATTCAAAGTAGTTTTAATTATCCT\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]: 'CTTCAAGAACCCTACGGAGAGTCGGACTGCATCTCCGCAGCGAGCTCTTGAGCGCCGCCGGCCGGGAGG\nCGAAGGATGCAGGCGGCTCCGCGCGCCGGCTGC
GGGGCAGCGCTCCTGCTGTGGATTGTCAGCAGCTGCC\nnTCTGCAGAGCCTGGACGGCTCCCTCCACGTCCCAAAAATGTGATGAGCCACTTGTCTCTGGACTCCC
CCA\nnTGTGGCTTTCAGCAGCTCCTCCTCCATCTCTGGTAGCTATTCTCCCGCTATGCCAAGATAAACAAGAGA\nnGGAGGTGCTGGGGGATGGTCTCCATCAGA
CAGCGACCATTATCAATGGCTTCAGGTTGACTTTGGCAATC\nnGGAAGCAGATCAGTGCCATTGCAACCCAAGGAAGGTATAGCAGCTCAGATTGGGTGACCCAAT
ACCGGAT\nnGCTCTACAGCGACACAGGGAGAACTGGAAACCTATCATCAAGATGGGAATATCTGGGCATTTCCCGGA\nnAACATTAACCTCTGACGGTGTGGTCC
GGCACGAATTACAGCATCCGATTATTGCCCCGCTATGTGCGCATAG\nnTGCCCTCTGGATTGGAATGGAGAAGGTGCGATTGGACTCAGAATTGAAGTTTATGGCTGT
TCTTACTGGGC\nnTGATGTTATCAACTTTGATGGCCATGTTGTATTACCATATAGATTCAGAAACAAGAAGATGAAAACACTG\nnAAAGATGTCATTGCCTTGAAC
TTTAAGACGTCTGAAAGTGAAGGAGTAATCCTGCACGGAGAAGGACAGC\nnAAGGAGATTACATTACCTTGGAACTGAAAAAAGCCAAGCTGGTCCTCAGTTTTAA
CTTAGGAAGCAACCA\nnGCTTGGCCCCATATATGGCCACACATCAGTGATGACAGGAAGTTTGTGGATGACCACCACTGGCACTCT\nnGTGGTCATTGAGCGCCA
GGGGCGGAGCATTAACTCACTCTGGACAGGAGCATGCAGCACTTCGTACCA\nnATGGAGAGTTTGTACTACCTGGACTTGGACTATGAGATAACCTTTGGAGGCA
TCCCTTTCTCTGGCAAGCC\nnCAGCTCCAGCAGTAGAAAGAATTTCAAAGGCTGCATGGAAAGCATCAACTACAATGGCGTCAACATTACT\nnGATCTTGCCAGAA
GGAAGAAATTAGAGCCCTCAAATGTGGGAAATTTGAGCTTTTCTGTGTGGAACCT\nnATACGGTGCCTGTCTTTTTCAACGCTACAAGTTACCTGGAGGTGCCC
GGACGGCTTAACCAGGACCTGTT\nnCTCAGTCAGTTTCCAGTTTAGGACATGGAACCCCAATGGTCTCCTGGTCTTCAGTCACTTTGC GGATAAT\nnTTGGGCAAT
GTGGAGATTGACCTCACTGAAAGCAAAGTGGGTGTTACATCAACATCACACAGACCAAGA\nnTGAGCCAAATCGATATTTCTCAGGTTCTGGGTTGAATGATGG
ACAGTGGCACGAGGTTTCGTTCTTAGC\nnCAAGGAAAATTTTGTATTCTCACCATCGATGGAGATGAAGCATCAGCAGTTCGAACTAATAGTCCCCCTT\nnCAAGT
TAAAACTGGCGAGAAGTACTTTTTTGGAGTTTTCTGAACCAGATGAATAACTCAAGTCACTCTG\nnTCCTTCAGCCTTCATTCCAAGGATGCATGCAGCTCATTC
AAGTGGACGATCAACTTGTAATTTATACGA\nnAGTGGCACAAAGGAAGCCGGGAAGTTTCGCGAATGTCAGCATTGACATGTGTGCGATCATAGACAGATGT\nnG
TGCCCAATCACTGTGAGCATGGTGGAAAGTGCTCGAAACATGGGACAGCTTCAAATGCATTGTGATG\nnAGACAGGATACAGTGGGGCCACCTGCCACAACCTCT
ATCTACGAGCCTTCTGTGAAGCCTACAAACACCT\nnAGGACAGACATCAAATTATTACTGGATAGATCCTGATGGCAGCGGACCTCTGGGGCCTCTGAAAGTTTA
C\nnTGCAACATGACAGAGGACAAAGTGTGGACCATAGTGTCTCATGACTTGCAGATGCAGACGCCTGTGGTTCG\nnGCTACAACCCAGAAAAATACTCAGTGACACA
GCTCGTTTACAGCGCCTCCATGGACCAGATAAGTGCCAT\nnCACTGACAGTGCCGAGTACTGCGAGCAGTATGTCTCCTATTTCTGCAAGATGTCAAGATTGTTGA
ACACC\nnCCAGATGGAAGCCCTTACACTTGGTGGTTGGCAAAGCCAACGAGAAGCACTACTACTGGGGAGGCTCTG\nnGGCCTGGAATCCAGAAATGTGCCTGCG
GCATCGAACGCAACTGCACAGATCCCAAGTACTACTGTAACCTG\nnCGACGCGGACTACAAGCAATGGAGGAAGGATGCTGGTTTTCTATCATACAAAGATCACCTG
CCAGTGAGC\nnCAAGTGGTGGTTGGAGATACTGACCGTCAAGGCTCAGAAGCCAAATTGAGCGTAGGTCTCTGCGCTGCC\nnAAGGAGACAGGAATTATTGGAAT
GCCGCCTCTTTCCAAACCCATCCTCCTACCTGCACTTCTCTACTTT\nnCCAAGGGGAAACTAGCGCTGACATTTCTTTCTACTTCAAAACATTAACCCCTGGGG
AGTGTCTTTGAA\nnAATATGGGAAAGGAAGATTTATCAAGCTGGAGCTGAAGTCTGCCACAGAAGTGTCTTTTCTTTTATTTGATG\nnTGGGAAATGGGCCAGTAGA
GATTGTAGTGAGGTACCAACCCCTCTCAACGATGACCAGTGGCACCGGT\nnCACTGCAGAGAGGAATGTCAAGCAGGCCAGCTACAGGTGGACCGGCTACCGC
AGCAGATCCGCAAGGCC\nnCCAACAGAAGGCCACACCCGCTGGAGCTCTACAGCCAGTTATTTGTGGGTGGTGTGGGGGCCAGCAGG\nnGCTTCTGGGCTGCA
TCCGCTCCTTGAGGATGAATGGGGTGACACTTGACCTGGAGGAAAGAGCAAAGGT\nnCACATCTGGGTTTATATCCGGATGCTCGGGCCATTGCACCAGCTATGGA
ACAAACTGTGAAAATGGAGGC\nnAAATGCCTAGAGAGATACCACGGTTACTCCTGCGATTGCTCTAATACTGCATATGATGGAACATTTTGCA\nnACAAAGATGTT
GGTGCATTTTTTTGAAGAAGGGATGTGGCTACGATATAACTTTTCAAGCACCAAGCAACAAA\nnTGCCAGAGACTCCAGCAGCAGAGTAGACAACGCTCCCGACCAGCA
GAACTCCACCCGACCTGGCACAG\nnGAGGAGATCCGCTTCAAGCTTCAAGCACCAAGGCGCCCTGCATTCTCCTCTACATCAGCTCCTTCACCA\nnCAGACTT
CTTGGCAGTCTCTCGTCAAACCCACTGGAAGCTTACAGATTCGATACAACCTGGGTGGCACCCG\nnAGAGCCATACAATATTGACGTAGACCACAGGAACATGGCCA
ATGGACAGCCCCACAGTGTCAACATCACC\nnCGCCACGAGAAGACCATCTTTCTCAAGCTCGATCATTATCCTTCTGTGAGTTACCATCTGCCAAGTTCAT\nnCCG
ACACCTCTTCAATTCTCCCAAGTCGCTCTTTCTGGGAAAAGTTATAGAAACAGGGAAAATTGACCA\nnAGAGATTACAAATACAACACCCCAAGGATTCACTGGT
TGCCTCTCCAGAGTCCAGTTCAACCAGATCGCC\nnCCTCTCAAGGCCGCTTGAGGCAGACAAACGCCTCGGCTCACGTCCACATCCAGGGCGAGCTGGTGGAGT
\nCCAAGTGGGGGCTCGCCGCTGACCTCTCCCCATGTCGTCCGCCACCGACCCCTGGCACCTGGATCA\nnCCTGGATTCAAGCAGTGGGATTTTCCATATA
ATCCAGGACAAGGCCAAGCTATAAGAAATGGAGTCAAC\nnAGAACTCGGCTATCATTGGAGGCGTCATTGCTGTGGTGATTTTACCATCCTGTGCACCTGGTC
TTCC\nnTGATCCGGTACATGTTCCGCCACAAGGGCACCTACCATACCAACGAAGCAAAGGGGGCGGAGTCGGCAGA\nnGAGCGCGGACGCCGCCATCATGAACAAC
GACCCCAACTTACAGAGACCATTTGATGAAAGCAAAAAGGAA\nnTGCTCATTTTGGGGGTGGCTACTTGGCTATGGGATAGGGAGGAGGAATTACTAGGGAGGA
GAGAAAGG\nnGACAAAAGCACCTGCTTCATACTCTTGAGCACATCCTTAAAATATCAGCACAAAGTTGGGGGAGGCAGGC\nnAATGGAATATAATGGAATATTCTT
GAGACTGATCACAAAAAAAACCTTTTTAATATTCTTTATAGC\nnTGAGTTTTCCCTTCTGTATCAAAACAAAATAATACAAAAAATGCTTTTAGAGTTTAAAG

CAATGGTTGAAA\nTTTGTAGGTACTATCTGTCTTATTTTGTGTGTGTTTAGAGGTGTTCTAAAGACCCGTGGTAACAGGGCAA\nnGTTTTCTACGTTTTTAAGAG
CCCTTAGAACGTGGGTATTTTTTTTCTTGAGAAAAGCTAATGCACCTACA\nnGATGGCCCCAACATTCTCTCTTTTGTCTTAGTCAACCTTAATGGGCTGT
ACAGAACTAGTTCGT\nnGTTTATATACTATTTCTTTGATGTCTATAAGTCGAAAAAGAAAGGGGCAAAGAGAACCTATTATTTGC\nnCAGTTTTTAAGCAGAG
CTCAATCTATGCCAGCTCTCTGGCATCTGGGGTTCTGACTGATACCAGCAGTT\nnGAAGGAAGAGAGTGCATGGCACCTGGTGTGAACGACACAATCAGCACAA
CTGGAGAGAGGCATTAAAGA\nnACCAGGGAAGGTAGTTTGATTTTTTCATTGAATTCTACAAGCTAATATTGTTCCACGTATGTAGTCTTAGA\nnCCAATAGCTGTA
ACTATCAGCTGCAATACCATGGTGACCAGCTGTTACAAAAGATTTTTTCTGTTTTAT\nnCTGAAACATACTGGATTTATATATGTATAAGCGCTCAATGGGGAA
TTAGAGCCAGATGTTATGATTTGT\nnTTGCTCTTTTTCTTTATAGTTTAGTTATAGCAAAAATATGGATAATTTCTAGTGAATGCATAAATTAGG\nnTTGCGTTT
CTTATTTTGTCTTAAATCTCTGGTAGTTTTTCCACCCCTGTGACACAATCCTAATAGACAGT\nnGTCCTGTAAATGGACACAACACAATAAAGTCAAGTTATTATT
GCTGTTACTCTGGATGATATGGAAAACA\nnCTGCCATATTTAAATCAACTACTCCACGTGTTTTTCCATCCAATCACACTGCTGTGATTACAGGGATCTT\nnTCTT
CTAAGACGGACACATTTGAACCTCAGGTTTCATCACAAACCTGGTACCTGTTGCTTCCCAGAGGATG\nnGAGAAGTGTAGTTAATCACACCTCTTAGTTTAATCTGA
AATCTTGACCCAGTTATTTAACAATAAATAC\nnCTCATTGATTATATTTAAAGTAATACACTTCTGTAAACAAATGGGGACAATGCATCCAAAAAATCTTT\nn
TTAAACAGATTACACAAAAATTATTTCCAGAAAGGCTACCATTTATCATCATTATATTTCAAGCCTCTTA\nnTACTTAATAAGCACTTTCTAAAAAGTCTTGAGAT
CCCACCATCTGAGGAATTCAATATGATCACTTTTT\nnCCCTTCTTTGCCTGGGAGAGGTTAAGAGGAGGTTTCTGAAGGTATAGATGCTATTGTTCTGATGGCCCGG
CT\nnGAATAAAATGGAAATTCTAGTTTGTTAGAATTATGCATTCTTTTTCAAGATTCTCAGTGTGCCTAACTTA\nnTTGGAGCACATCAGTTTCTTGGGTAATGGA
AAACATTACCTAGAGTTGCCAGTGGCACATTACACCAGTA\nnCAGAGCACATTCCAAAGGAGACATTGGACCAGTTAATCCCATACAAGTCAAGGTAACAGAACA
AAAGGG\nnAATCCTGATGCCCTTTTACCATTGCTGGTTGAGCTCAGGCACTGTGATGGACACCCTTAATTTTAAAGG\nnTTTAAATCATTCTCTATAAAATACA
TTTAAATGGAAAAATACTTAATATCACTAAATATCAGAACAAAT\nnGTAACATTTACAAATGACATATTGAAAGCAAAGGCTGTTTTATTTAGCCAAGATGATTAC
CATTAGGAGT\nnTACTTTATGTATTGTTGAAAGCAAATTTTAAACATGATGTTTTAGAAGTGTCTGATTTTTAAACCTGG\nnTTTACAGGTATTACTTCTGCAC
TTACCAAATAATGCCAGATGGAAATTTATTATTTCTTGCAATTTCCAT\nnGATAGCTCTGTTCTTTATGCATTGTCTCAACACTTTCCCTTTTTTCCCAAATGAG
TAGAGAATTAAGC\nnCACCACAAACAGCTTCTGCTACTAAAATGTTCTCATCCTTTCTCCTCCCTCTCCTTTTCTGCCACAAAA\nnGGTGAAAAATGAGATCCA
ATCCTCTCACCAAAATTTCAAACCTAGGACACTGGAATGACTGCAGGGATCA\nnGTGGTTCTCCCATATCACCATCAATTAAGACATATAGGACACTGTCTTCCTT
CAAGAGGGTTACAATGTG\nnGCCATCAGACAGGAAACCAAACGGTGGATAAAGTATTAAGTAAGTAACTAAGTGCCAAATAAATGCTGGAAATC\nnTTGACCTCTCCTTG
GGATTATGGGTGTAACAAAAATCCCTACATCTGTTTATGAAGGCCATATTCAGTAC\nnATTTTAAATGGTAAATAATCTGTTTATGTGAAGAAAAAGAATTAAGTC
TTTCTTCAAACCTCTCCTTGG\nnATAGCCTAGCACAGTGACGCTCCATAACCATGACATTTCCGCCCCAAGCTCTCAGTGCCTAATCCTGCTT\nnTGTCATTAC
ATCTCACAAAATCTTGACATCTTACATTCCAATACATTATCAAGCAAGCACAAAGTATGCT\nnGGTAGTAGCCTCTTTAAATAATATGTATAGACAACAACAACGAC
AAAAATAGACTGTTTTAAAGTTTCA\nnGGGAAAGTTGGTGGCTGATTTAAAGTTGTGCAGGAAACATCTTCTGTGTATGAAGCAATGTGATGTTT\nnTGAAAA
AAGCTAGGAGATGACTTTGAATGAATGCAAGGTTAGTGAGATCCTAAGCTCTCAAAATAGCATA\nnTTCCCTAGAGCTCAAGAAAGCTGGTCCAGGAGGTTGAAAA
AGCTATTTTGTGTAAATTATTTTCTGGC\nnCCTTCTTAATATTTAAAAATGTATTTCCCTTGTGGCTTTCAACCACCTGCTCAAAAAAGAGACTTGTT\nnAC
ATGAAAGTTTTATTAAAGAGCTGAAAAACAAGAATTAGAGAGCCATTCTAGAAAAATGTCCTACTGC\nnCCTGCATTTGACAAACAAGCATCCTTTACTAACAAG
AGCAGGAATTAGAGGCACAAGAAAAAGCATTGG\nnCATGAGCCAAAGAGTCTGTCTTAATGTTACTTTTGAAAATCTGCTGAGCGCCACCATATGCAGGCTGAG
\nAGCTGGGCACAGGCGAAGCCATTGGAAGCACTTCAGGAACAAGCACACAGCTGTGGGACTTGAACATGCA\nnAGTGTTCAAGTTGTGTCAAGAAGCTTTTCTTT
CCTTCTATGATGGAATCTGTTCTTTTCTATCCTACTTT\nnTTTCTCTCTTCTCCTCCTCACCACATTATACCCTGCTCTTACGCAGTAAACGTTTTAATGGCCCGT
TTAT\nnGTCTCATGCCTCCAAACAACACTGAATTTGAAACCCCCCATTTTTTCTTTTACCACCCTGTTGAGCAAT\nnTTTCCCCAAAAAAGGGCAGCAATTATTA
AATTGAATTCAAGTAAGCCAGCCAAAGATAGGTCCTAAATTG\nnCTAGTCCCAGTAGAACCACCTGATCCTAAACCAGTGCAGAAACAAACAGTAACAATGTCCCCA
GCTGACTT\nnCAGCTAAGAACCAATGGCTCTACCCCCGCCCCGCTTTTTTTTGTGTTTTTGTGTTTTTGTGTTTTGAGACGG\nnAGTCTTGCTCTGTCCCCAGGCTG
GAGTGCCTGGCGCAATCTCGGGCTCACTGCAACCTCCTCCTCTC\nnCACCATTGAGGCGATTCTCCTGCCTCAGCTCCCAAGTAGCTGGGATTACAGGCACCC
GCCATCACACCC\nnAGCTAATTTTTTTTTTTTTTTTGTATTATTAGTAGAAGCCAGGTTTACCATTGTTGGCCAGGGTGGTCTCG\nnAACTCCTGACCTCAAGTGAT
CCGTCCACCTCGGCCTTGCAAATTGCTGGGATTACAGGTGTGAGCCACCG\nnTGCCGAGCCAGCCCCATTTTTTAAATGATGTTTTGTTAAGAGTGGACCATGAG
AATTAGCTGACAGCAT\nnCCCCCTTCTCTCTCCCTGCCTTGGTGGGACCCCTCCCTGTGTGACCTTGGTCAAGTCTCGAACCTTTGTG\nnCCGTATTTAAGATGGA
GCTGTTTTACCTACTTCATAAGACAGTTGCGAGGTGCCATTGATTCTTGACTGC\nnAAAATACCTTGAAACCTTATATAAAGACTGAAGTCAACGGAGCCTAGTG
AAAGACTTACTTTGTGGCTT\nnGTGGTTGAAAGTCACATCAAAAGACAAATGTGGCCACGTTCAGGAATTGGAGACTTACTGGCATGGCTCT\nnACAGCTGCTCAG
TTATTAATCATGCAGACTAACCTGTCAACACTGGGAGATGCAACATAGCAAAAGGACA\nnGAGAAATTAGAATTTTTTGTGCAGAAAGCCCTAAATCCCACCTGA

```
ATGTAAC TTACAGCTCCCTTACCT\nACTCTCACACATGCCCTCAAACATGCTAGATTGGCTTATACATAGGCCAACACAAAATACAAACGTGACG\nnTGTTTCATG  
TAGCCTAGTGGCTATATGCCTATTCTCCATGTACCCTGCATGGTAGTGCTGCAAAC TTTAAA\nnGTACATTTCTTTCACAGCAGTATTTTTTTTCATAAGTGGCAT  
ATAAATGTCATTCAATGAAATGGGGAAA\nnTCACGTTGAGAAGTTGGTCTGTCATCTCCATTGAGCAAAGACTGGCAGGAGATAATAAAAAATAAATATG\nnGGCA  
CACATGTATTAATATACAGCACGCATTTACAAGTTTATTTTCCAGATAAAAATTGTGCTATAAGAAC\nnAGCTCTACCAAGACAGTCTGCACCATTTCCAAGTCTCA  
GTTAATTTACAGCAACTGCTGCTTTTCGGAGAT\nnGGCTGTGAAAATATGGAAGTTCCTCTCAAGTAGGCCAAGAAACAGTTCTAGATTTTACTAAGTTTTATTT\nn  
TGTCAGGTTTTTTTAAATTTTTTTCAGTGAGCGTGGTGACTGCAGAGGTTAGTGCTGTGAAAAGCTGGGCTA\nnAATATTCTTTCTGTAAGTCAAACAGGATTCCAT  
CCCCTGTGAAATAACACAAAATTTCACTCTCTAAAA\nnGCAACAGCATGTAACTAGAAATGAAAGAAGGAAATTATGTACGTATGCCTAATATTCTTTGTGAATGT  
CT\nnTTCATTTAACTAAAATTATATTAGAAACCAGATTGATAAATAAAAAATTCAAAGTAGTTTTAATTATCCT\nnAAAA\nn\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]: ['CTTCAAGAACCCTACGGAGAGTCGGACTGCATCTCCGCAGCGAGCTCTTGGAGCGCCGCCGGCCGGGAGG\n',
'CGAAGGATGCAGGCGGCTCCGCGCGCCGGCTGCGGGGCAGCGCTCCTGCTGTGGATTGTCAGCAGCTGCC\n',
'TCTGCAGAGCCTGGACGGCTCCCTCCACGTCCCAAAAATGTGATGAGCCACTTGTCTCTGGACTCCCCCA\n',
'TGTGGCTTTCAGCAGCTCCTCCTCCATCTCTGGTAGCTATTCTCCCGGCTATGCCAAGATAAACAAGAGA\n',
'GGAGGTGCTGGGGGATGGTCTCCATCAGACAGCGACCATTATCAATGGCTTCAGGTTGACTTTGGCAATC\n',
'GGAAGCAGATCAGTGCCATTGCAACCCAAGGAAGGTATAGCAGCTCAGATTGGGTGACCCAATACCGGAT\n',
'GCTCTACAGCGACACAGGGAGAACTGGAAACCTATCATCAAGATGGGAATATCTGGGCATTTCCCGGA\n',
'AACATTAACCTCTGACGGTGTGGTCCGGCAGCAATTACAGCATCCGATTATTGCCCGCTATGTGCGCATAG\n',
'TGCCTCTGGATTGGAATGGAGAAGGTGCGATTGGACTCAGAATTGAAGTTTATGGCTGTTCTTACTGGGC\n',
'TGATGTTATCAACTTTGATGGCCATGTTGTATTACCATATAGATTCAGAAACAAGAAGATGAAAACACTG\n',
'AAAGATGTCATTGCCTTGAACTTTAAGACGTCTGAAAGTGAAGGAGTAATCCTGCACGGAGAAGGACAGC\n',
'AAGGAGATTACATTACCTTGAACTGAAAAAAGCCAAGCTGGTCTCAGTTTAACTTAGGAAGCAACCA\n',
'GCTTGGCCCCATATATGGCCACACATCAGTGATGACAGGAAGTTTGCTGGATGACCACCACTGGCACTCT\n',
'GTGGTCATTGAGCGCCAGGGGCGGAGCATTAACTCACTCTGGACAGGAGCATGCAGCACTTCCGTACCA\n',
'ATGGAGAGTTTGACTACCTGGACTTGGACTATGAGATAACCTTTGGAGGCATCCCTTCTCTGGCAAGCC\n',
'CAGCTCCAGCAGTAGAAAGAATTTCAAAGGCTGCATGGAAAGCATCAACTACAATGGCGTCAACATTACT\n',
'GATCTTGCCAGAAGGAAGAAATTAGAGCCCTCAAATGTGGGAAATTTGAGCTTTTCTTGTGTGGAACCCT\n',
'ATACGGTGCCTGTCTTTTTCAACGCTACAAGTTACCTGGAGGTGCCCGGACGGCTTAACCAGGACCTGTT\n',
'CTCAGTCAGTTTCCAGTTTAGGACATGGAACCCAATGGTCTCCTGGTCTTCAGTCACTTTGCGGATAAT\n',
'TTGGGCAATGTGGAGATTGACCTCACTGAAAGCAAAGTGGGTGTTACATCAACATCACACAGACCAAGA\n',
'TGAGCCAAATCGATATTTCTCAGGTTCTGGGTTGAATGATGGACAGTGGCACGAGGTTGCTTCTTAGC\n',
'CAAGGAAAATTTTGCTATTCTCACCATCGATGGAGATGAAGCATCAGCAGTTCGAACTAATAGTCCCCTT\n',
'CAAGTTAAAACTGGCGAGAAGTACTTTTTTGGAGGTTTTCTGAACCAGATGAATAACTCAAGTCACTCTG\n',
'TCCTTCAGCCTTCATTCCAAGGATGCATGCAGCTCATTCAAGTGGACGATCAACTTGTAATTTATACGA\n',
'AGTGGCACAAAGGAAGCCGGGAAGTTTCGCGAATGTCAGCATTGACATGTGTGCGATCATAGACAGATGT\n',
'GTGCCCAATCACTGTGAGCATGGTGGAAAGTGCTCGCAAACATGGGACAGCTTCAAATGCACTTGTGATG\n',
'AGACAGGATACAGTGGGGCCACCTGCCACAACCTCTATCTACGAGCCTTCTGTGAAGCCTACAAACACCT\n',
'AGGACAGACATCAAATTATTACTGGATAGATCCTGATGGCAGCGGACCTCTGGGGCCTCTGAAAGTTTAC\n',
'TGCAACATGACAGAGGACAAAGTGTGGACCATAGTGTCTCATGACTTGAGATGCAGACGCCTGTGGTCTG\n',
'GCTACAACCCAGAAAAATACTCAGTGACACAGCTCGTTTACAGCGCCTCCATGGACCAGATAAGTGCCAT\n',
'CACTGACAGTGCCGAGTACTGCGAGCAGTATGTCTCCTATTTCTGCAAGATGTCAAGATTGTTGAACACC\n',
'CCAGATGGAAGCCCTTACACTTGGTGGGTTGGCAAAGCCAACGAGAAGCACTACTACTGGGGAGGCTCTG\n',
'GGCCTGGAATCCAGAAATGTGCCTGCGGCATCGAACGCAACTGCACAGATCCCAAGTACTACTGTAAGT\n',
'CGACGCGGACTACAAGCAATGGAGGAAGGATGCTGGTTTCTTATCATACAAAGATCACCTGCCAGTGAGC\n',
'CAAGTGGTGGTTGGAGATACTGACCGTCAAGGCTCAGAAGCCAAATTGAGCGTAGGTCTCTGCGCTGCC\n',
'AAGGAGACAGGAATTATTGGAATGCCGCCTCTTTCCAAACCCATCCTCCTACCTGCACTTCTCTACTTT\n',
'CCAAGGGGAACTAGCGCTGACATTTCTTTCTACTTCAAAACATTAACCCCTGGGGAGTGTTTCTTGAA\n',
'AATATGGGAAAGGAAGATTTATCAAGCTGGAGCTGAAGTCTGCCACAGAAGTGCTTTTTATTTGATG\n',
'TGGGAAATGGGCCAGTAGAGATTGTAGTGAGGTACCAACCCCTCTCAACGATGACCAGTGGCACCGGGT\n',
'CACTGCAGAGAGGAATGTCAAGCAGGCCAGCCTACAGGTGGACCGGCTACCGCAGCAGATCCGCAAGGCC\n',
'CCAACAGAAGGCCACACCCGCCTGGAGCTCTACAGCCAGTTATTTGTGGGTGGTCTGGGGGCCAGCAGG\n',
'GCTTCTGGGCTGCATCCGCTCCTTGAGGATGAATGGGGTGACACTTGACCTGGAGGAAAGAGCAAAGGT\n']
```

'CACATCTGGGTTTCATATCCGGATGCTCGGGCCATTGCACCAGCTATGGAACAACTGTGAAAATGGAGGC\n',
'AAATGCCTAGAGAGATACCACGGTTACTCCTGCGATTGCTCTAATACTGCATATGATGGAACATTTTGCA\n',
'ACAAAGATGTTGGTGCATTTTTTTGAAGAAGGGATGTGGCTACGATATAACTTTTCAGGCACCAGCAACAAA\n',
'TGCCAGAGACTCCAGCAGCAGAGTAGACAACGCTCCCGACCAGCAGAACTCCCACCCGGACCTGGCACAG\n',
'GAGGAGATCCGCTTCAGCTTCAGCACCAACCAAGGCGCCCTGCATTCTCCTCTACATCAGCTCCTTCACCA\n',
'CAGACTTCTTGGCAGTCCTCGTCAAACCCACTGGAAGCTTACAGATTTCGATACAACCTGGGTGGCACCCG\n',
'AGAGCCATACAATATTGACGTAGACCACAGGAACATGGCCAATGGACAGCCCCACAGTGTCAACATCACC\n',
'CGCCACGAGAAGACCATCTTTCTCAAGCTCGATCATTATCCTTCTGTGAGTTACCATCTGCCAAGTTCAT\n',
'CCGACACCCTCTTCAATTCTCCCAAGTCGCTCTTTCTGGGAAAAGTTATAGAAACAGGGAAAATTGACCA\n',
'AGAGATTCACAAATACAACACCCCAGGATTCACTGGTTGCCTCTCCAGAGTCCAGTTCAACCAGATCGCC\n',
'CCTCTCAAGGCCGCCTTGAGGCAGACAAACGCCTCGGCTCACGTCCACATCCAGGGCGAGCTGGTGGAGT\n',
'CCAAGTGCAGGCGCTCGCCGCTGACCCTCTCCCCCATGTCGTCCGCCACCGACCCCTGGCACCTGGATCA\n',
'CCTGGATTGAGCCAGTGCAGGATTTTCCATATAATCCAGGACAAGGCCAAGCTATAAGAAATGGAGTCAAC\n',
'AGAAACTCGGCTATCATTGGAGGCGTCATTGCTGTGGTGATTTTACCATCCTGTGCACCCTGGTCTTCC\n',
'TGATCCGGTACATGTTCCGCCACAAGGGCACCTACCATAACGAAGCAAAGGGGGCGGAGTGGGCAGA\n',
'GAGCGCGGACGCCGCCATCATGAACAACGACCCCAACTTCACAGAGACCATTGATGAAAGCAAAAAGGAA\n',
'TGGCTCATTTGAGGGGTGGCTACTTGGCTATGGGATAGGGAGGAGGGAATTACTAGGGAGGAGAGAAAGG\n',
'GACAAAAGCACCCCTGCTTCATACTCTTGAGCACATCCTTAAAATATCAGCACAAGTTGGGGGAGGCAGGC\n',
'AATGGAATATAATGGAATATTCTTGAGACTGATCACAAAAAAAAAACCTTTTAAATATTTCTTTATAGC\n',
'TGAGTTTTCCCTTCTGTATCAAAACAAAATAATACAAAAATGCTTTTAGAGTTTAAGCAATGGTTGAAA\n',
'TTTGTAGGTACTATCTGTCTTATTTTGTGTGTGTTAGAGGTGTTCTAAAGACCCGTGGTAACAGGGCAA\n',
'GTTTTCTACGTTTTTAAAGGCCCTTAGAACGTGGGTATTTTTTTTCTTGAGAAAAGCTAATGCACCTACA\n',
'GATGGCCCCAACATTCTCTTCTTTTGTCTTAGTCAACCTTAATGGGCTGTTACAGAACTAGTTCGT\n',
'GTTTATATACTATTTCTTTGATGTCCTATAAGTCGGAAAAGAAAGGGGCAAAGAGAACCTATTATTTGC\n',
'CAGTTTTTAAAGCAGAGCTCAATCTATGCCAGCTCTCTGGCATCTGGGGTTCCTGACTGATACCAGCAGTT\n',
'GAAGGAAGAGAGTGCATGGCACCTGGTGTGTAACGACACAATCAGCACAACCTGGAGAGAGGCATTAAAGA\n',
'ACCAGGGAAGGTAGTTTGATTTTTCATTGAATTCTACAAGCTAATATTGTTCCACGTATGTAGTCTTAGA\n',
'CCAATAGCTGTAACATATCAGCTGCAATACCATGGTGACCAGCTGTTACAAAAGATTTTTCTGTTTTAT\n',
'CTGAAACATACTGGATTTATATATGTATAAGCGCCTCAATGGGGAATTAGAGCCAGATGTTATGATTTGT\n',
'TTGCTCTTTTTCTTTTATAGTTTAGTTATAGCAAAAATATGGATAATTTCTAGTGAATGCATAAATTAGG\n',
'TTGCGTTTCTTATTTTGCTTTAAATCTCTGGTAGTTTTTCCACCCCTGTGACACAATCCTAATAGACAGT\n',
'GTCCTGTAAATGGACACAACACAATAAAGTCAAGTTATTATTGCTGTTACTCTGGATGATATGGAAAACA\n',
'CTGCCATATTTTAAATCACTACTCCACGTGTTTTTCCATCCAATCACACTGCTGTGATTGAGGGATCTT\n',
'TCTTCTAAGACGGACACATTTGAACCTCAGGTTTCATCACAAACCTGGTACCTGTTGCTTCCCAGAGGATG\n',
'GAGAAGTGTAGTTAATCACACCTCTTAGTTTAACTGAAATCTTGACCCAGTTATTTAACAATAAATAC\n',
'CTCATTGATTATATTTAAAAGTAATACACTTCCTGTAAACAAATGGGGACAATGCATCCAAAAAATCTTT\n',
'TTAAACAGATTACACAAAAATTATTTCCAGAAAGGCTACCATTTATCATCATTATATTTCAAGCCTCTTA\n',
'TACTTAATAAGCACTTTCTAAAAAGTCTTGAGATCCCACCATTTCTGAGGAATTCAATATGATCACTTTTT\n',
'CCTTCTTTCCTGGGAGAGGTTAAGAGGAGGTTTCGAAGGTATAGATGCTATTGTTCTGATGGCCCGGCT\n',
'GAATAAAATGGAAATTCAGTTTGTTAGAATTATGCATTCTTTTTCAAGATTCTCAGTGTGCCTAACTTA\n',
'TTGGAGCACATCAGTTTCTTGGGTAATGGAAAACATTACCTAGAGTTGCCAGTGGCACATTACACCAGTA\n',
'CAGAGCACATTCCAAAGGAGACATTGGACCAGTTAATTCCTATACAAGTCAAGGTAACAGAACAAAAGGG\n',

'AATCCTGATGCCCTTTTACCATTGCTGGTTGAGCTCAGGCACTGTCATGGACACCCTTAATTTTAAAAGG\n',
'TTTTAATCATTCTTCTATAAAATACATTTAAATGGAAAAATACTTAATATCACTAAATATCAGAACAAT\n',
'GTAACATTTACAAATGACATATTGAAAGCAAAGGCTGTTTTATTTAGCCAAGATGATTACCATTAGGAGT\n',
'TACTTTATGTATTGTTGAAAGCAAATTTTAAACATGATGTTTTAGAAGTGTCTGATTTTTAAACCTGG\n',
'TTTACAGGTATTACTTCTGCACCTACCAAATAATGCCAGATGGAAATTTATTATTTCTTGCAATTCCCAT\n',
'GATAGCTCTGTTCTTTATGCATTGTCTCAACACTTTCCCTTTTTTCCCAAAATGAGTAGAGAATTAAGC\n',
'CACCCAAAACAGCTTCTGCTACTAAAATGTTCTCATCCTTCTCCTCCCTCTCCTTTTCTGCCACAAAA\n',
'GGTGAAAAATGAGATCCAATCCTCTCACCAAAATTTCAAACCTAGGACACTGGAATGACTGCAGGGATCA\n',
'GTGGTTCTCCCATATCACCATCAATTAAGACATATAGGACACTGTCTTCTTCAAGAGGGTTACAATGTG\n',
'GCCATCAGACAGGAAACCAAACGGTGGATAAAGTATTAAGTAACTAAGTGCCAAATAAATGCTGGAAATC\n',
'TTGACCTCTCCTTGGGATTATGGGTGTAACAAAAATCCCTACATCTGTTTATGAAGGCCATATTCAGTAC\n',
'ATTTTAAATGGTAAATAATCTGTTTATGTGAAGAAAAAGAATTAAGTCTTCTTCCAACCTCTCTCCTTGG\n',
'ATAGCCTAGCACAGTGCAGCCTCCATAACCATGACATTCCTCGCCCAAGCTCTCAGTGCCTAATCCTGCTT\n',
'TGTCATTACATCTCACAAAATCTTGACATCTTACATTCCAATACATTATCAAGCAAGCACAAGTATGCT\n',
'GGTAGTAGCCTCTTTAAATAATATGTATAGACAACAACAACGACAAAAAATAGACTGTTTTAAAGTTTCA\n',
'GGGAAAGTTGGTGGCTGATTTAAAGTTGTGCAGGAAACATCTTCTGTGTATGAAGCAAATGTCGATGTTT\n',
'TGAAAAAAGCTAGGAGATGACTTTGAATGAATGCAAGGTTAGTGAGATCCTAAGCTCTCAAATAGCATA\n',
'TTCCCTAGAGCTCAAGAAAGCTGGTCCAGGAGTTGAAAAAGCTATTTTGTTGTTAAATTATTTTCTGGC\n',
'CCTTCTTAATATTTAAAAATGTATTTCCCTTGTGGCTTTCAACCACCTGCTCAAAAAAAGAGACTTGTT\n',
'ACATGAAAGTTTTATTAAAGAGCTGAAAAACAAGATTTAGAGAGCCATTCTAGAAAATGTCCTACTGC\n',
'CCTGCATTTGACAAACAAGCATCCTTTACTAACAAGAGCAGGAATTCAGAGGCACAAGAAAAAGCATTGG\n',
'CATGAGCCAAAGAGTCTGTCTTAATGTTACTTTTAAAAATCTGCTGAGCGGCCACCATATGCAGGCTGAG\n',
'AGCTGGGCACAGGCGAAGCCATTGGAAGCACTTCAGGAACAAGCACACAGCTGTGGGACTTGAACATGCA\n',
'AGTGTTCAGGTTGTGTCAAGAAGCTTTTCTTCTCTATGATGGAATCTGTTCTTTTCTATCCTACTTT\n',
'TTTCTCTCTTCTCCTCCTCACCACATTATACCCTGCTCTTACGCAGTAAACGTTTTAATGGCCCGTTTAT\n',
'GTCTCATGCCTCCAAACAACACTGAATTTGAAACCCCCCATTTTTTCTTTTACCACCCTGTTGAGCAAT\n',
'TTTCCCAAAAAAAGGGCAGCAATTATTAATTTGAATTCAAGTAAGCCAGCCAAAGATAGGTCTAAATTG\n',
'CTAGTCCCAGTAGAACCACCTGATCCTAAACCACTGCGAAACAAACAGTAACAATGTCCCCAGCTGACTT\n',
'CAGCTAAGAACCAATGGCTCCTACCCCCGCCCGCTTTTTTTTGTGTTTTTGTGTTTTGAGACGG\n',
'AGTCTTGCTCTGTCCCCAGGCTGGAGTGCACCTGGCGCAATCTCGGGCTCACTGCAACCTCCTCCTCCTC\n',
'CCACATTGAGGCGATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGCACCCGCCATCACACCC\n',
'AGCTAATTTTTTTTTTTTTTTTGTATTATTAGTAGAAGCCAGGTTTACCATGTTGGCCAGGGTGGTCTCG\n',
'AACTCCTGACCTCAAGTGATCCGTCCACCTCGGCCTTGCAAATTGCTGGGATTACAGGTGTGAGCCACCG\n',
'TGCCGAGCCAGCCCCATTTTTTAAATGATGTTTTGGTTAAGAGTGGACCATGAGAATTAGCTGACAGCAT\n',
'CCCCTTCTCTCTCCTGCCTTGGTGGGACCTCCCTGTGTGACCTGGTCAAGTCTCGAACTTTTGTG\n',
'CCGTATTTAAGATGGAGCTGTTTTACCTACTTCATAAGACAGTTGCGAGGTGCCATTGATTCTTGACTGC\n',
'AAAATACCTTGAAACCTTATATAAAGACTGAAGTCAACGGAGCCTAGTGAAAGACTTACTTTGTGGCTT\n',
'GTGGTTGAAAGTCACATCAAAGACAAATGTGGCCACGTTTCAAGGAATTGGAGACTTACTGGCATGGCTCT\n',
'ACAGCTGCTCAGTTATTAATCATGCAGACTAACCTGTCAACACTGGGAGATGCAACATAGCAAAAGGACA\n',
'GAGAAATTAGAATTTTTTGTGCAGAAAGCCCTAAATTTCCACCTGAATGTAACCTTACAGCTCCCTTACCT\n',
'ACTCTCACACATGCCCTCAAACATGCTAGATTGGCTTATACATAGGCCAACACAAAATACAAACGTGACG\n',
'TGTTTCATGTAGCCTAGTGGCTATATGCCTATTCTCCATGTACCCTGCATGGTAGTGCTGCAAACCTTTAAA\n'

```
'GTACATTTCTTTCACAGCAGTATTTTTTTTCATAAGTGGCATATAAATGTCATTCAATGAAATGGGGAAA\n',  
'TCACGTTGAGAAGTTGGTCTGTCATCTCCATTGAGCAAAGACTGGCAGGAGATAATAAAAATAAATATG\n',  
'GGCACACATGTATTAATATACAGCACGCATTTACAAGTTTATTTTCCAGATAAAATTGTGCTATAAGAAC\n',  
'AGCTCTACCAAGACAGTCTGCACCATTTCCAAGTCTCAGTTAATTTACAGCAACTGCTGCTTTCGGAGAT\n',  
'GGCTGTGAAAATATGGAAGTTCCTCTCAAGTAGGCCAAGAAACAGTTCTAGATTTTACTAAGTTTTATTT\n',  
'TGTCAGGTTTTTTAAATTTTTTTCAGTGAGCGTGGTGACTGCAGAGGTTAGTGCTGTGAAAAGCTGGGCTA\n',  
'AATATTCTTTCTGTAAAGTCAAACAGGATTCCATCCCCTGTGAAATAACACAAAATTTCACTCTCTAAAA\n',  
'GCAACAGCATGTAAACTAGAAATGAAAGAAGGAAATTATGTACGTATGCCTAATATTCTTTGTGAATGTCT\n',  
'TTCATTTAACTAAAATTATATTAGAAACCAGATTGATAAATAAAAAATTCAAAGTAGTTTTAATTATCCT\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

Traceback (most recent call last)

Cell In[85], line 1

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

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]: ['CTTCAAGAACCCTACCGAGAGTCCGACTGCATCTCCGACGACGACTCTTGAGAGCGCCGCCGCGGAGG\n',  
 'CGAAGGATGCAGGCGGCTCCGCGCGCCGGCTGCGGGGCAGCGCTCCTGCTGTGGATTGTCAGCAGCTGCC\n',  
 'TCTGCAGAGCCTGGACGGCTCCCTCCACGTCCCCAAAATGTGATGAGCCACTTGTCTCTGGACTCCCCCA\n',  
 'TGTGGCTTTTACGACGCTCCTCCTCATCTCTGGTAGCTATTCTCCCGGCTATGCCAAGATAAAACAAGAGA\n',  
 'GGAGGTGCTGGGGGATGGTCTCCATCAGACAGCGACCATTATCAATGGCTTCAGTTGACTTTGGCAATC\n',  
 'GGAAGCAGATCAGTGCCATTGCAACCCAAGGAAGGTATAGCAGCTCAGATTGGGTGACCCAATACCGGAT\n',  
 'GCTCTACAGCGACACAGGGAGAACTGGAACCCATCATCAAGATGGGAATATCTGGGCATTTCCCGGA\n',  
 'AACATTA ACTCTGACGGTGTGGTCCGGCACGAATTACAGCATCCGATTATTGCCCGCTATGTGCGCATAG\n',  
 'TGCCTCTGGATTGGAATGGAGAAGGTGCGATTGGACTCAGAATTGAAGTTTATGGCTGTTCTTACTGGGC\n',  
 'TGATGTTATCAACTTTGATGGCCATGTTGTATTACCATATAGATTAGAAAACAAGAAGATGAAAACACTG\n',  
 'AAAGATGTCATTGCCTTGAACTTTAAGACGTCTGAAAGTGAAGGAGTAATCCTGCACGGAGAAGGACAGC\n',  
 'AAGGAGATTACATTACCTTGAACTGAAAAAAGCCAAGCTGGTCTCAGTTTAAACTTAGGAAGCAACCA\n',  
 'GCTTGGCCCCATATATGGCCACACATCAGTGATGACAGGAAGTTTGCTGGATGACCACCACTGGCACTCT\n',  
 'GTGGTCATTGAGCGCCAGGGGCGGAGCATTAACCTCACTCTGGACAGGAGCATGCAGCACTTCCGTACCA\n',  
 'ATGGAGAGTTTGACTACCTGGACTTGACTATGAGATAACCTTTGGAGGCATCCCTTCTCTGGCAAGCC\n',  
 'CAGCTCCAGCAGTAGAAAGAATTTCAAAGGCTGCATGGAAAGCATCAACTACAATGGCGTCAACATTACT\n',  
 'GATCTTGCCAGAAGGAAGAAATTAGAGCCCTCAAATGTGGGAAATTTGAGCTTTTCTGTGTGGAACCT\n',  
 'ATACGGTGCCTGTCTTTTTCAACGCTACAAGTTACCTGGAGGTGCCCGGACGGCTTAACCAGGACCTGTT\n',  
 'CTCAGTCAGTTTCCAGTTTAGGACATGGAACCCCAATGGTCTCCTGGTCTTCAGTCACTTTGCGGATAAT\n',  
 'TTGGGCAATGTGGAGATTGACCTCACTGAAAGCAAAGTGGGTGTTACATCAACATCACACAGACCAAGA\n',  
 'TGAGCCAAATCGATATTTCTCAGGTTCTGGGTTGAATGATGGACAGTGGCACGAGGTTTCGCTTCTTAGC\n',  
 'CAAGGAAAATTTTGCTATTCTCACCATCGATGGAGATGAAGCATCAGCAGTTCGAACATAAGTCCCCTT\n',  
 'CAAGTTAAACTGGCGAGAAGTACTTTTTTGAGGTTTTCTGAACCAGATGAATAACTCAAGTCACTCTG\n',  
 'TCCTTCAGCCTTCATTCCAAGGATGCATGCAGCTCATTCAAGTGGACGATCAACTTGTAATTTTATACGA\n',  
 'AGTGGCACAAAGGAAGCCGGGAAGTTTCGCGAATGTCAGCATTGACATGTGTGCGATCATAGACAGATGT\n',  
 'GTGCCCAATCACTGTGAGCATGGTGAAAGTGCTCGCAAACATGGGACAGCTTCAAATGCACTTGTGATG\n',  
 'AGACAGGATACAGTGGGGCCACCTGCCACAACCTCTATCTACGAGCCTTCTGTGAAGCCTACAAACACCT\n',  
 'AGGACAGACATCAAATTATTACTGGATAGATCCTGATGGCAGCGGACCTCTGGGGCCTCTGAAAGTTTAC\n',  
 'TGCAACATGACAGAGGACAAAGTGTGGACCATAGTGTCTCATGACTTGCAGATGCAGACGCTGTGGTCG\n',  
 'GCTACAACCCAGAAAAATACTCAGTGACACAGCTCGTTTTACAGCGCTCCATGGACCAGATAAGTGCCAT\n',  
 'CACTGACAGTGCCGAGTACTGCGAGCAGTATGTCTCCTATTTCTGCAAGATGTCAAGATTGTTGAACACC\n',  
 'CCAGATGGAAGCCCTTACACTTGGTGGGTTGGCAAAGCCAACGAGAAGCACTACTACTGGGGAGGCTCTG\n',  
 'GGCCTGGAATCCAGAAATGTGCCTGCGGCATCGAACGCAACTGCACAGATCCCAAGTACTACTGTAAGT\n',  
 'CGACGCGGACTACAAGCAATGGAGGAAGGATGCTGGTTTTCTATCATACAAAGATCACCTGCCAGTGAGC\n',  
 'CAAGTGGTGGTTGGAGATACTGACCGTCAAGGCTCAGAAGCCAAATTGAGCGTAGGTCTCTGCGCTGCC\n',  
 'AAGGAGACAGGAATTATTGGAATGCCGCTCTTTCCCAAACCCATCCTCCTACCTGCACTTCTCTACTTT\n',  
 'CCAAGGGGAAACTAGCGCTGACATTTCTTTTACTTTCAAACATTAACCCCTGGGGAGTGTTTCTTGAA\n',  
 'AATATGGGAAAGGAAGATTTTCATCAAGCTGGAGCTGAAGTCTGCCACAGAAGTGTCTTTTCATTTGATG\n',  
 'TGGGAAATGGGCCAGTAGAGATTGTAGTGAGGTACCAACCCCTCTCAACGATGACCAGTGGCACCGGGT\n',  
 'CACTGCAGAGAGGAATGTCAAGCAGGCCAGCCTACAGGTGGACCGGTACCGCAGCAGATCCGCAAGGCC\n',  
 'CCAACAGAAGGCCACACCCGCTGGAGCTCTACAGCCAGTTATTTGTGGTGGTGTGGGGGCCAGCAGG\n',  
 'GCTTCTGGGCTGCATCCGCTCCTTGAGGATGAATGGGGTGACACTTGACCTGGAGGAAAGAGCAAAGGT\n']
```

'CACATCTGGGTTTCATATCCGGATGCTCGGGCCATTGCACCAGCTATGGAACAACTGTGAAAATGGAGGC\n',
'AAATGCCTAGAGAGATACCACGGTTACTCCTGCGATTGCTCTAATACTGCATATGATGGAACATTTTGCA\n',
'ACAAAGATGTTGGTGCATTTTTTTGAAGAAGGGATGTGGCTACGATATAACTTTTCAGGCACCAGCAACAAA\n',
'TGCCAGAGACTCCAGCAGCAGAGTAGACAACGCTCCCGACCAGCAGAACTCCCACCCGGACCTGGCACAG\n',
'GAGGAGATCCGCTTCAGCTTCAGCACCAACCAAGGCGCCCTGCATTCTCCTCTACATCAGCTCCTTCACCA\n',
'CAGACTTCTTGGCAGTCCTCGTCAAACCCACTGGAAGCTTACAGATTTCGATACAACCTGGGTGGCACCCG\n',
'AGAGCCATACAATATTGACGTAGACCACAGGAACATGGCCAATGGACAGCCCCACAGTGTCAACATCACC\n',
'CGCCACGAGAAGACCATCTTTCTCAAGCTCGATCATTATCCTTCTGTGAGTTACCATCTGCCAAGTTCAT\n',
'CCGACACCCTCTTCAATTCTCCCAAGTCGCTCTTTCTGGGAAAAGTTATAGAAACAGGGAAAATTGACCA\n',
'AGAGATTCACAAATACAACACCCCAGGATTCACTGGTTGCCTCTCCAGAGTCCAGTTCAACCAGATCGCC\n',
'CCTCTCAAGGCCGCCTTGAGGCAGACAAACGCCTCGGCTCACGTCCACATCCAGGGCGAGCTGGTGGAGT\n',
'CCAAGTGCAGGCGCTCGCCGCTGACCCTCTCCCCCATGTCGTCCGCCACCGACCCCTGGCACCTGGATCA\n',
'CCTGGATTGAGCCAGTGCAGGATTTTCCATATAATCCAGGACAAGGCCAAGCTATAAGAAATGGAGTCAAC\n',
'AGAAACTCGGCTATCATTGGAGGCGTCATTGCTGTGGTGATTTTACCATCCTGTGCACCCTGGTCTTCC\n',
'TGATCCGGTACATGTTCCGCCACAAGGGCACCTACCATAACGAAGCAAAGGGGGCGGAGTGGGCAGA\n',
'GAGCGCGGACGCCGCCATCATGAACAACGACCCCAACTTCACAGAGACCATTGATGAAAGCAAAAAGGAA\n',
'TGGCTCATTTGAGGGGTGGCTACTTGGCTATGGGATAGGGAGGAGGGAATTACTAGGGAGGAGAGAAAGG\n',
'GACAAAAGCACCCCTGCTTCATACTCTTGAGCACATCCTTAAAATATCAGCACAAAGTTGGGGGAGGCAGGC\n',
'AATGGAATATAATGGAATATTCTTGAGACTGATCACAAAAAATACCTTTTAAATATTTCTTTATAGC\n',
'TGAGTTTTCCCTTCTGTATCAAAACAAAATAATACAAAAATGCTTTTAGAGTTTAAGCAATGGTTGAAA\n',
'TTTGTAGGTACTATCTGTCTTATTTTGTGTGTGTTAGAGGTGTTCTAAAGACCCGTGGTAACAGGGCAA\n',
'GTTTTCTACGTTTTTAAAGGCCCTTAGAACGTGGGTATTTTTTTTCTTGAGAAAAGCTAATGCACCTACA\n',
'GATGGCCCCAACATTCTCTTCTTTTGTCTTAGTCAACCTTAATGGGCTGTTACAGAACTAGTTCGT\n',
'GTTTATATACTATTTCTTTGATGTCCTATAAGTCGGAAAAGAAAGGGGCAAAGAGAACCTATTATTTGC\n',
'CAGTTTTTAAGCAGAGCTCAATCTATGCCAGCTCTCTGGCATCTGGGGTTCCTGACTGATACCAGCAGTT\n',
'GAAGGAAGAGAGTGCATGGCACCTGGTGTGTAACGACACAATCAGCACAACTGGAGAGAGGCATTAAAGA\n',
'ACCAGGGAAGGTAGTTTGATTTTTCATTGAATTCTACAAGCTAATATTGTTCCACGTATGTAGTCTTAGA\n',
'CCAATAGCTGTAAGTATCAGCTGCAATACCATGGTGACCAGCTGTTACAAAAGATTTTTCTGTGTTTAT\n',
'CTGAAACATACTGGATTTATATATGTATAAGCGCCTCAATGGGGAATTAGAGCCAGATGTTATGATTTGT\n',
'TTGCTCTTTTTCTTTTATAGTTTAGTTATAGCAAAAATATGGATAATTTCTAGTGAATGCATAAATTAGG\n',
'TTGCGTTTCTTATTTTGCTTTAAATCTCTGGTAGTTTTTCCACCCCTGTGACACAATCCTAATAGACAGT\n',
'GTCCTGTAAATGGACACAACACAATAAAGTCAAGTTATTATTGCTGTTACTCTGGATGATATGGAAAACA\n',
'CTGCCATATTTTAAATCACTACTCCACGTGTTTTTCCATCCAATCACACTGCTGTGATTGAGGGATCTT\n',
'TCTTCTAAGACGGACACATTTGAACCTCAGGTTTCATCACAAACCTGGTACCTGTTGCTTCCCAGAGGATG\n',
'GAGAAGTGTAGTTAATCACACCTCTTAGTTTAAATCTGAAATCTTGACCCAGTTATTTAACAATAAATAC\n',
'CTCATTGATTATATTTAAAAGTAATACACTTCTGTAAACAAATGGGGACAATGCATCCAAAAATCTTT\n',
'TTAAACAGATTACACAAAAATTATTTCCAGAAAGGCTACCATTTATCATCATTATATTTCAAGCCTCTTA\n',
'TACTTAATAAGCACTTTCTAAAAAGTCTTGAGATCCCACCATTTCTGAGGAATTCAATATGATCACTTTTT\n',
'CCTTCTTTCCTGGGAGAGGTTAAGAGGAGGTTTCGAAGGTATAGATGCTATTGTTCTGATGGCCCGGCT\n',
'GAATAAAATGGAAATCTAGTTTGTTAGAATTATGCATTCTTTTTCAAGATTCTCAGTGTGCCTAACTTA\n',
'TTGGAGCACATCAGTTTCTTGGGTAATGGAAAACATTACCTAGAGTTGCCAGTGGCACATTACACCAGTA\n',
'CAGAGCACATTCCAAAGGAGACATTGGACCAGTTAATTCCTATACAAGTCAAGGTAACAGAACAAAAGGG\n',

'AATCCTGATGCCCTTTTACCATTGCTGGTTGAGCTCAGGCACTGTCATGGACACCCTTAATTTTAAAAGG\n',
'TTTTAATCATTCTTCTATAAAATACATTTAAATGGAAAAATACTTAATATCACTAAATATCAGAACAAT\n',
'GTAACATTTACAAATGACATATTGAAAGCAAAGGCTGTTTTATTTAGCCAAGATGATTACCATTAGGAGT\n',
'TACTTTATGTATTGTTGAAAGCAAATTTTAAACATGATGTTTTAGAAGTGTCTGATTTTTAAACCTGG\n',
'TTTACAGGTATTACTTCTGCACCTACCAAATAATGCCAGATGGAAATTTATTATTTCTTGCAATTCCCAT\n',
'GATAGCTCTGTTCTTTATGCATTGTCTCAACACTTTCCCTTTTTTCCCAAAATGAGTAGAGAATTAAGC\n',
'CACCCAAAACAGCTTCTGCTACTAAAATGTTCTCATCCTTCTCCTCCCTCTCCTTTTCTGCCACAAAA\n',
'GGTGAAAAATGAGATCCAATCCTCTCACCAAAATTTCAAACCTAGGACACTGGAATGACTGCAGGGATCA\n',
'GTGGTTCTCCCATATCACCATCAATTAAGACATATAGGACACTGTCTTCTTCAAGAGGGTTACAATGTG\n',
'GCCATCAGACAGGAAACCAAACGGTGGATAAAGTATTAAGTAACTAAGTGCCAAATAAATGCTGGAAATC\n',
'TTGACCTCTCCTTGGGATTATGGGTGTAACAAAAATCCCTACATCTGTTTATGAAGGCCATATTCAGTAC\n',
'ATTTTAAATGGTAAATAATCTGTTTATGTGAAGAAAAAGAATTAAGTCTTCTTCCAACCTCTCTCCTTGG\n',
'ATAGCCTAGCACAGTGCAGCCTCCATAACCATGACATTCCTCGCCCAAGCTCTCAGTGCCTAATCCTGCTT\n',
'TGTCATTACATCTCACAAAATCTTGACATCTTACATTCCAATACATTATCAAGCAAGCACAGTATGCT\n',
'GGTAGTAGCCTCTTTAAATAATATGTATAGACAACAACACGACAAAAAATAGACTGTTTTAAAGTTTCA\n',
'GGGAAAGTTGGTGGCTGATTTAAAGTTGTGCAGGAAACATCTTCTGTGTATGAAGCAAATGTCGATGTTT\n',
'TGAAAAAAGCTAGGAGATGACTTTGAATGAATGCAAGGTTAGTGAGATCCTAAGCTCTCAAAATAGCATA\n',
'TTCCCTAGAGCTCAAGAAAGCTGGTCCAGGAGTTGAAAAAGCTATTTTGTGTTAAATTATTTTCTGGC\n',
'CCTTCTTAATATTTAAAAATGTATTTCCCTTGTGGCTTTCAACCACCTGCTCAAAAAAAGAGACTTGTT\n',
'ACATGAAAGTTTTCATTAAGAGCTGAAAAACAAGATTTAGAGAGCCATTCTAGAAAATGTCCTACTGC\n',
'CCTGCATTTGACAAACAAGCATCCTTTACTAACAAGAGCAGGAATTCAGAGGCACAAGAAAAAGCATTGG\n',
'CATGAGCCAAAGAGTCTGTCTTAATGTTACTTTTAAAAATCTGCTGAGCGGCCACCATATGCAGGCTGAG\n',
'AGCTGGGCACAGGCGAAGCCATTGGAAGCACTTCAGGAACAAGCACACAGCTGTGGGACTTGAACATGCA\n',
'AGTGTTCAAGTTGTGTCAAGAAGCTTTTCTTCTCTATGATGGAATCTGTTCTTTTCTATCCTACTTT\n',
'TTTCTCTCTTCTCCTCCTCACCACATTATACCCTGCTCTTACGCAGTAAACGTTTTAATGGCCCGTTTAT\n',
'GTCTCATGCCTCCAAACAACACTGAATTTGAAACCCCCCATTTTTTCTTTTACCACCCTGTTGAGCAAT\n',
'TTTCCCAAAAAAAGGGCAGCAATTATTAATTTGAATTCAAGTAAGCCAGCCAAAGATAGGTCTAAATTG\n',
'CTAGTCCCAGTAGAACCACCTGATCCTAAACCACTGCGAAACAAACAGTAACAATGTCCCCAGCTGACTT\n',
'CAGCTAAGAACCAATGGCTCCTACCCCCGCCCGCTTTTTTTTGTGTTTTTGTGTTTTGAGACGG\n',
'AGTCTTGCTCTGTCCCCAGGCTGGAGTGCCTGGCGCAATCTCGGGCTCACTGCAACCTCCTCCTCTC\n',
'CCACATTGAGGCGATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGCACCCGCCATCACACCC\n',
'AGCTAATTTTTTTTTTTTTTTTGTATTATTAGTAGAAGCCAGGTTTACCATGTTGGCCAGGGTGGTCTCG\n',
'AACTCCTGACCTCAAGTGATCCGTCCACCTCGGCCTTGCAAATTGCTGGGATTACAGGTGTGAGCCACCG\n',
'TGCCGAGCCAGCCCCATTTTTTAAATGATGTTTTGGTTAAGAGTGGACCATGAGAATTAGCTGACAGCAT\n',
'CCCCTTCTCTCTCCTGCCTTGGTGGGACCTCCCTGTGTGACCTGGTCAAGTCTCGAACTTTTGTG\n',
'CCGTATTTAAGATGGAGCTGTTTTACCTACTTCATAAGACAGTTGCGAGGTGCCATTGATTCTTGACTGC\n',
'AAAATACCTTGAAACCTTATATAAAGACTGAAGTCAACGGAGCCTAGTGAAAGACTTACTTTGTGGCTT\n',
'GTGGTTGAAAGTCACATCAAAGACAAATGTGGCCACGTTTCAAGGAATTGGAGACTTACTGGCATGGCTCT\n',
'ACAGCTGCTCAGTTATTAATCATGCAGACTAACCTGTCAACACTGGGAGATGCAACATAGCAAAAGGACA\n',
'GAGAAATTAGAATTTTTTGTGCAGAAAGCCCTAAATTTCCACCTGAATGTAACCTTACAGCTCCCTTACCT\n',
'ACTCTCACACATGCCCTCAAACATGCTAGATTGGCTTATACATAGGCCAACACAAAATACAAACGTGACG\n',
'TGTTTCATGTAGCCTAGTGGCTATATGCCTATTCTCCATGTACCCTGCATGGTAGTGCTGCAAACCTTTAAA\n'

```
'GTACATTTCTTTCACAGCAGTATTTTTTTTCATAAGTGGCATATAAATGTCATTCAATGAAATGGGGAAA\n',
'TCACGTTGAGAAGTTGGTCTGTCATCTCCATTGAGCAAAGACTGGCAGGAGATAATAAAAATAAATATG\n',
'GGCACACATGTATTAATATACAGCACGCATTTACAAGTTTATTTTCCAGATAAAATTGTGCTATAAGAAC\n',
'AGCTCTACCAAGACAGTCTGCACCATTTCCAAGTCTCAGTTAATTTACAGCAACTGCTGCTTTCGGAGAT\n',
'GGCTGTGAAAATATGGAAGTTCCTCTCAAGTAGGCCAAGAAACAGTTCTAGATTTTACTAAGTTTTATTT\n',
'TGTCAGGTTTTTTAAATTTTTTTCAGTGAGCGTGGTGACTGCAGAGGTTAGTGCTGTGAAAAGCTGGGCTA\n',
'AATATTCTTTCTGTAAAGTCAAACAGGATTCCATCCCCTGTGAAATAACACAAAATTTCACTCTCTAAAA\n',
'GCAACAGCATGTAACTAGAAAGTAAAGAAAGGAAATTATGTACGTATGCCTAATATTCTTTGTGAATGTCT\n',
'TTCATTTAACTAAAATTATATTAGAAACCAGATTGATAAATAAAAAATTCAAAGTAGTTTTAATTATCCT\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): GAAGUUCUUGGGAUGCCUCUCAGCCUGACGUAGAGGCGUCGUCGAGAACCU CGCGCGGCCGCCUC
CGCUUCCUACGUCCGCCGAGGCGCGCGGCCG ...
```

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(sequencia, enzimas):
    sequencia = sequencia.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 = sequencia.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(sequencia, cortes):
    fragmentos = []
    inicio = 0
    for c in cortes:
        fragmentos.append(sequencia[inicio:c])
        inicio = c
    fragmentos.append(sequencia[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(sequencia, enzimas):
enzimas_con_corte, enzimas_sin_corte, cortes = encontrar_cortes(sequencia, enzimas)
fragmentos, tamanos = construir_fragmentos(sequencia, cortes)
imprimir_reporte(enzimas_con_corte, enzimas_sin_corte, fragmentos, tamanos)
return enzimas_con_corte, enzimas_sin_corte, fragmentos, tamanos
```

```
In [19]: seq = "AAGCTTGGGAATTCCGATCGGATCCAGCTTAGCTGAATTC"
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 = "ATGGCCATTGTAATGGGCCGCTGAAAGGGTGCCCGATAG"
    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': 'LSGTLAAHYNGH', 'longitud': 13, 'frame': 1, 'strand': '-'},
[{'proteina': 'LSGTLAAHYNGH', 'longitud': 13, 'frame': 1, 'strand': '-'},
{'proteina': 'GHCNGPLKGCPI', 'longitud': 12, 'frame': 3, 'strand': '+'},
{'proteina': 'YRAPFQRPITMA', '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)
```