

Problem 1

A DNA sequence is a string made up of the letters A, T, G, and C. To find the complement of a DNA sequence, As are replaced by Ts, Ts by As, Gs by Cs, and Cs by Gs. For example, the complement of AATTGCCGT is TTAACGGCA.

In [25]:

```
def dna_short(old):
    trans_table = str.maketrans('tagc','atcg') #지정한 문자를 특정 문자로 변환
    new = old.translate(trans_table)
    return new

def dna_long(old):
    new = "" #if case로 전체를 분리하여 변환
    for _ in range(0, len(old)):
        if(old[_] == 't'):
            new += 'a'
        elif(old[_] == 'a'):
            new += 't'
        elif(old[_] == 'g'):
            new += 'c'
        elif(old[_] == 'c'):
            new += 'g'
        else:
            new += old[_]
    return new

# Test Case
p53="""1 ttcccatcaa gccctagggc tctcgtggc tgcgtggagt tgtagtctga acgtttctat
61 cttggcgaga agcgctacg ctccccctac cgagtccgc ggtaattctt aaagcacctg
121 caccgcccc ccgcgcctg cagaggcgc agcaggtctt gcacctcttc tgcattcat
181 tctccaggtc tcagacctgt ctccctcatt caaaaaatat ttattatcga gctcttactt
241 gctaccacgc actgatatag gcactcagga atacaacaat gaataagata gtagaaaaat
301 tctatatcct cataaggctt acgtttccat gtactgaaag caatgaacaa ataaatctta
361 tcagagtgat aagggttgtg Waaggagatta aataagatgg tgtgatataa agtatctggg
421 agaaaaagtt aggtgtgat attacggaaa gccttcctaa aaaatgacat ttttaactgat
481 gagaagaaag gatccagctg agagcaaacg caaaagcttt ctctcttcca cccttcatat
541 ttgacacaa gcaggattcc tccaaaatga tttccacca tttcgccctc acagctctgg
601 cttgcagaat tttccacccc aaaatgttag tatctacggc accaggtcgg cgagaatcct
661 gactctgcac cctctcccc aactccattt ctttgcttc ctccggcagg cggattactt
721 gcccttactt gtcatggcga ctgtccagct ttgtgccagg agcctcgagc gggttgatgg
781 gattggggtt ttccccctcc atgtgctcaa gactggcgct aaaagttttg agcttctcaa
"""

print(dna_short(p53))
print(dna_long(p53))
```

```
1 aagggtagtt cgggatcccg aggagcaccg acgacctca acatcagact tgcgaagata
61 gaaccgctct tcgcggatgc gagggggatg gctcagggcg ccattaagaa tttcgtggac
121 gtggcggggg ggcgcgcgac gtctccgcgc tcgtccagaa cgtggagaag acgtagagta
181 agaggtccga agtctggaca gagggagtaa gttttttata aataatagct cgagaatgaa
241 cgatgggtcg tgactatata cgtgagtcct tatgttgta cttattctat catcttttta
301 agatatagga gtattccgaa tgcaaaggta catgactttc gttacttggt tatttagaat
361 agtctcacta ttcccaacac Wttcctctaatt ttattctacc acactatatt tcatagacct
421 tcttttgcaa tcccacacta taatgccttt cggaaggatt ttttactgta aaattgacta
481 ctcttctttc ctaggctgac tctcgtttgc gttttcgaaa gaaggaaggt gggaagtata
541 aactgtgtta cgtcctaagg aggttttact aaaggtggtt aagacgggag tgcgagacc
601 gaacgtctta aaaggtgggg ttttacaatc atagatgccg tggccagcc gctcttagga
661 ctgagacgtg ggaggagggg ttgaggtaaa ggaaacgaag gaggccgtcc gcctaataa
721 cgggaatgaa cagtaccgct gacaggtcga aacacggtcc tcggagcgct cccaactacc
781 ctaaccccaa aaggggaggg tacacgagtt ctgaccgcga ttttcaaaac tcgaagagtt
```

```
1 aagggtagtt cgggatcccg aggagcaccg acgacctca acatcagact tgcgaagata
61 gaaccgctct tcgcggatgc gagggggatg gctcagggcg ccattaagaa tttcgtggac
121 gtggcggggg ggcgcgcgac gtctccgcgc tcgtccagaa cgtggagaag acgtagagta
181 agaggtccga agtctggaca gagggagtaa gttttttata aataatagct cgagaatgaa
241 cgatgggtcg tgactatata cgtgagtcct tatgttgta cttattctat catcttttta
301 agatatagga gtattccgaa tgcaaaggta catgactttc gttacttggt tatttagaat
361 agtctcacta ttcccaacac Wttcctctaatt ttattctacc acactatatt tcatagacct
421 tcttttgcaa tcccacacta taatgccttt cggaaggatt ttttactgta aaattgacta
```

```
481 ctcttctttc ctaggtcgac tctcgtttgc gttttcgaaa gaaggaaggt gggaagtata
541 aactgtgtta cgtcctaagg aggtttttact aaaggtggtt aagacgggag tgtcgagacc
601 gaacgtctta aaaggtgggg ttttacaatc atagatgccg tggccagcc gctcttagga
661 ctgagacgtg ggaggagggg ttgaggtaaa ggaaacgaag gaggccgtcc gcctaataaa
721 cgggaatgaa cagtaccgct gacaggtcga aacacggtcc tcggagcgtc cccaactacc
781 ctaaccccaa aaggggaggg tacacgagtt ctgaccgcga ttttcaaaac tcgaagagtt
```