group_15_package

```
library(biocd)
```

Github link

Description

The package contains functions that can convert a DNA base sequence to corresponding RNA sequence, and finally the corresponding amino acid chain. Lastly, a barplot is created that shows the distribution of the each amino acid. In this way, a DNA sequence of any length can be converted to its corresponding amino acid sequence quickly and easily, and visualized.

The package consists of the following five functions:

Function 1: create_DNA()

Creates a DNA sequence of a given length. The input is the wanted length of the DNA sequence (number of bases). The output is a string of letters of the four bases in DNA; T, A, G, and C.

Example:

```
create_DNA(10)
#> [1] "TCGAAATCGT"
```

Function 2: DNA_to_RNA()

Converts a given DNA string consisting of the letters T, A, G, and C, to the corresponding RNA sequence. This is done by substituting all instances of the letter "T" with "U". The input is a DNA string. The output is a string of letters that are the bases of the RNA sequence.

Example:

```
DNA_to_RNA("AGCTTCGTAG")
#> [1] "AGCUUCGUAG"
```

Function 3: codons_extraction()

The function receives the string of RNA sequence as an input and splits it into several shorter strings composed by the characters representing each codon. The function outputs an array consisting of all the "substrings". If the input string is not divisible by three, then the last bases which do not fit into the triplets are discarded.

Example:

```
codons_extraction("AGCUUCGUAG")
#> [1] "AGC" "UUC" "GUA"
```

Function 4: aa_translation()

Function 4 takes an array of codon triplets as inputs, and translates each one of these into the corresponding amino acid (represented by a letter). The output consists of the amino acids written as a concatenation of the characters into a string.

Example:

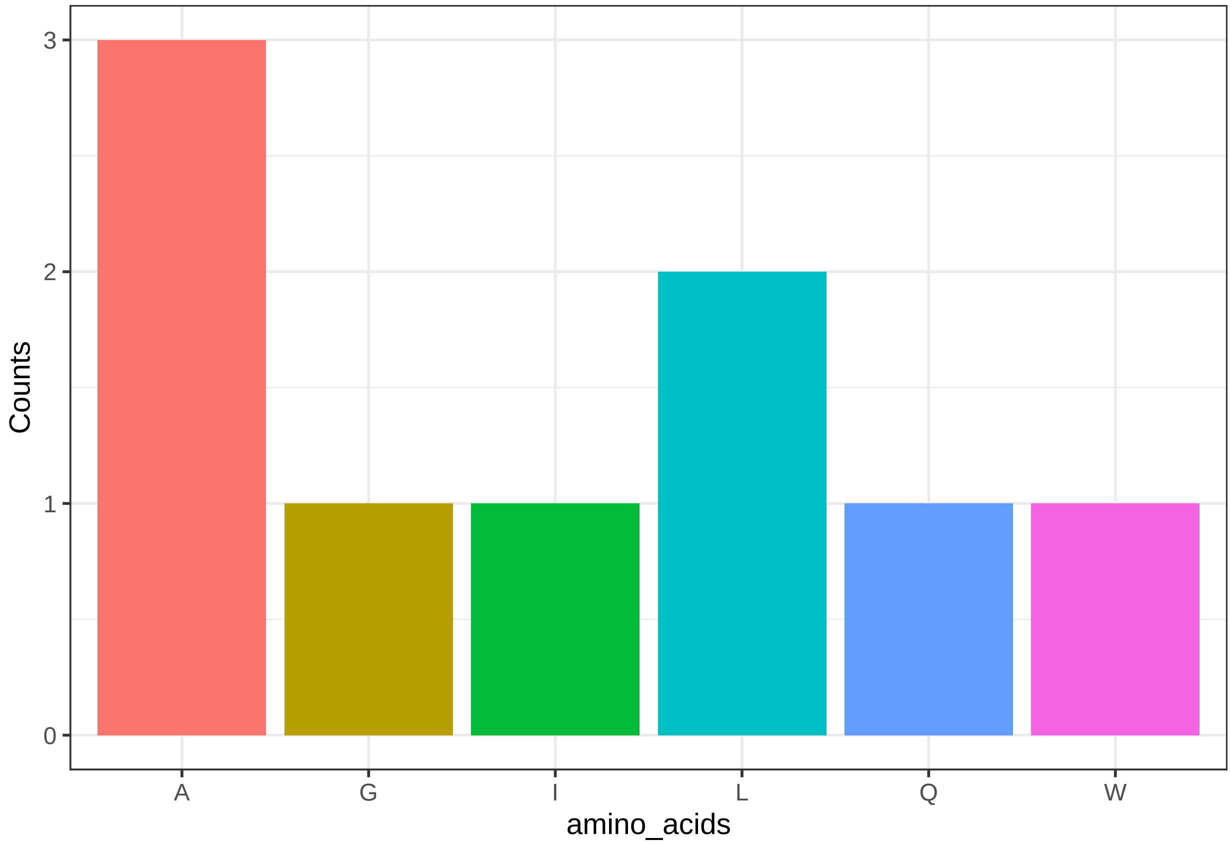
```
aa_translation(c("UCA", "GAG", "CUG", "AUG"))
#> [1] "SELM"
```

Function 5: aa_distribution_plot()

The function aa_distribution_plot takes plots the distribution of amino acids in a protein sequence. The function first identifies the unique letters in the string. These are used as a pattern to count the number of occurences for each of them. The count for each of the amino acids are then plotted as a bar chart.

Example:

```
aa_distribution_plot("GLLIAQWAA")
```



Pipeline of the functions in the package

Demonstrates how all the functions work together

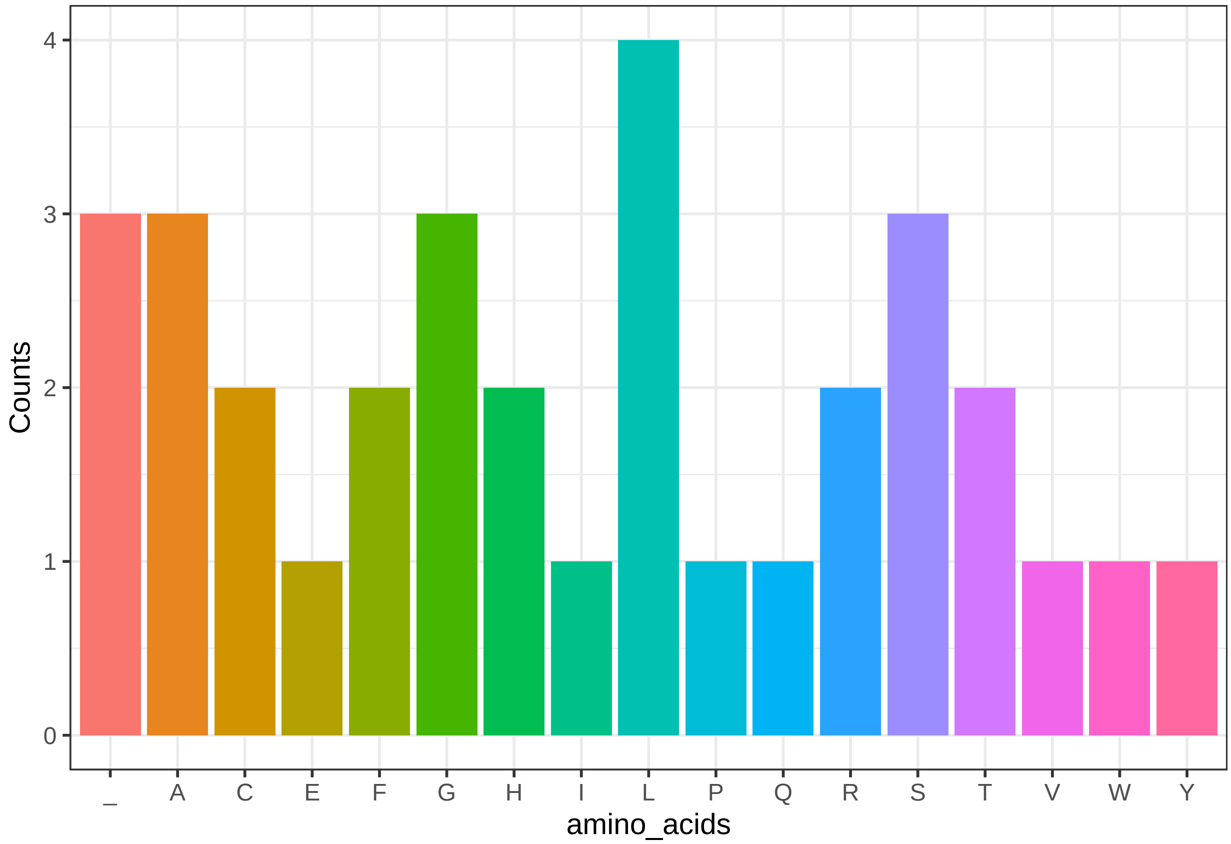
```
DNA <- create_DNA(100)

RNA <- DNA_to_RNA(DNA)

codons <- codons_extraction(RNA)

aa_seq <- aa_translation(codons)

aa_distribution_plot(aa_seq)
```


Counts
0
1
2
3
4
_
A
C
E
F
G
H
I
L
P
Q
R
S
T
V
W
Y
amino_acids