

Human T-cell leukemia virus type I

Aufgabe 2

[<https://www.ncbi.nlm.nih.gov/nuccore/22129792>]

Die ersten 100 Basenpaare der kodierende Sequenz.

```
atgggccaatcttttccgtagcgtagccctattccgcggccgccccgggggtggccgctcatcact  
g gcttaacttctccaggcggcatatcgcc
```

Aufgabe 3

Erste 1000 Basenpaare der kodierenden Sequenz.

```
a  
tgggccaat cttttccgtagcgtagcc ctattccgcg gccgccccgg  
gggctggccg ctcatcactg gcttaacttc ctccaggcgg catatcgcc  
agaaccgggt ccctccagtt acgatttcca ccagttaaaa aaatttctta  
aaatagcttt agaaacaccg gtctggatct gccccattaa ctactccctc  
ctagccagcc tactcccaa aggatacccc ggccgggtga atgaaatttt  
acacatactc atccaaaccc aagcccagat cccgtcccgc cccgcgccgc  
cgccgccgtc atcctccacc caccgacccc cggattctga cccacaaatc  
ccccctccct atgttgagcc tacagccccc caagtccttc cagtcatgca  
cccacatggg gccctccca accaccgcc atggcaaag aaagacctac  
aggccattaa gcaagaagtc tcccaagcgg cccctggaag cccccagttt  
atgcagacca tccggcttgc ggtgcagcag tttgaccca ctgccaaaga  
cctccaagac ctctgcagt acctttgctc ctccctcgtg gttccctcc  
atcaccagca gctagatagc cttatatcag aggcgaaac tcgaggtatt  
acagggtata accccttagc cggccccctc cgtgtccaag ccaacaatcc  
acaacaacaa ggattaaggc gagaatacca gcaactctgg ctgcgcgcct  
tcgcgcgcct gccaggaggt gccaaagacc ctctctgggc ctctatcctc  
caaggcctgg aggagcctta ccacgccttc gtagaacgcc tcaacatagc  
tcttgacaat gggctgccag aaggcacgcc caaagacccc attttacgtt  
ccttagccta ctctaagca aacaaagaat gccaaaaatt actacaggcc  
cgagggcaca ctaatagccc tctaggagat atgttgccgggcttgtcaggc
```

Ersten 30 Aminosäuren von 1.5'3' Frame

[<https://web.expasy.org/cgi-bin/translate/dna2aa.cgi>]

MGQIFSRASPIPRPPRGLAAHHWLNFLQA

- Bei Aminosäuresequenzen wird ersichtlich, wo sich die kodierenden Sequenzen/ Exons befinden. Zudem kann man erkennen, wo das Exon beginnt- bei einer Aminosäure ist dies eher zu sehen. Es wird sichtbar, ob es verschiedene Reading-Frames gibt.
- Einerseits gibt es drei Reading-Frames, jeweils von 5' -> 3' Richtung. Es ist jedoch vorzuziehen, dass auch die Sequenz von 3' -> 5' zu übersetzen.

Aufgabe 4

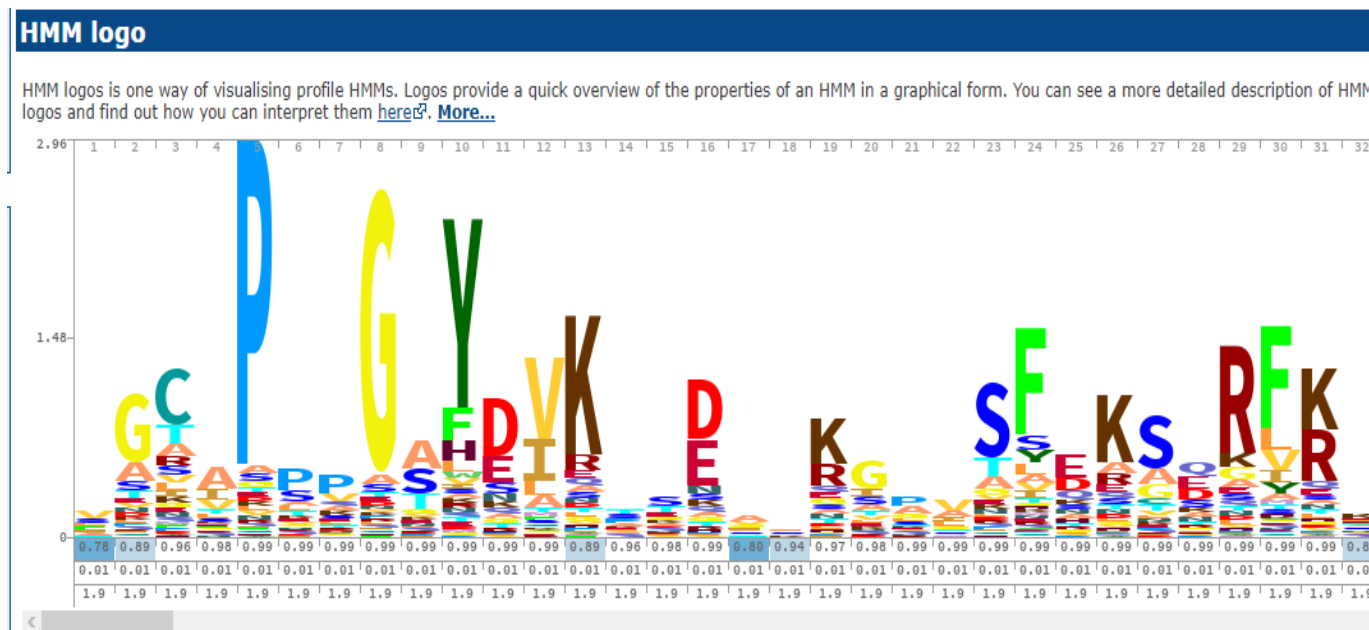


Abb.1. HMM Logo des gefundenen Profils aus Pfam Datenbank

<http://pfam.xfam.org/family/PF15905.4#tabview=tab4>

Im vergleich mit

MGQIFSRASPIRPPRGLAAHHWLNFLQA

-GCAPPPGAYDVK--D--KG--SFEKS-RF, es gibt nur es gleiche Aminosäure an 2.und 9. Stelle.

(„-“ = Die Wahrscheinlichkeit der vorkommende Aminosäure an dieser Stellen könnte nicht bestimmen werden). Die anderen vorkommenden Aminosäuren besitzen höhere Wahrscheinlichkeit und sind erkennbar.

Aufgabe 5

Sulfolobales Mexican fusellovirus 1, complete genome

NCBI Reference Sequence: NC_020882.1

<https://www.ncbi.nlm.nih.gov/nuccore/472438225>

5.2

Erste 100 Basenpaare der kodierenden Sequenz:

```
at gagcaaggta aaatacaccc ccgtggaggg gtcctatttc ctcctctaca aaaagggaaa  
ggaatacgtg caagtgaaac taggagaaat gtacttcc
```

5.3

Kodierende Sequenz: 360 Basen

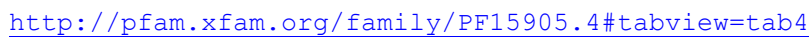
```
at gagcaaggta aaatacaccc ccgtggaggg gtcctatttc ctcctctaca aaaagggaaa  
ggaatacgtg caagtgaaac taggagaaat gtacttccgc ctaggcaaaa gtgagttacg  
cctcctaggg ttattgagga tgtaggcga ggaagtaaac aaacgagacc ctaggctaac  
ggtaatgcaa accagtacga gagattataa cgaaagacta aggaggcttg ccagtaaggg  
gcttatagtc accagaaggg agggaaacca atactacttg aggctatcaa ggcttggtca  
attgctcata gaggcaattg cgcccaagta ctatgatccg gaggtagaac agacatga
```

5`-3` Frame1 (Forward) Übersetzte Basensequenz in die Aminosäuresequenz mit ExSPaY:

<https://web.expasy.org/cgi-bin/translate/dna2aa.cgi>

```
MSKVKYTPVEGSYFLLYKKGKEYVQVKLGEMYFRLGKSELRLGLLRMLGEEVNKRDPRLTVMQTSTRDY  
NERLRRLASKGLIVTRREGNQYYLRLSRLGQLLIEAIAPKYDPEVEQT-
```

Das Suchergebnis nach HMM-Profilen aus der Pfam Datenbank:



MSKVKYTPVEGSYFLLYKKGKEYVQVKLGE

(„-“ = an dieser Stelle kann die Wahrscheinlichkeit der vorkommenden Aminosäure nicht eindeutig bestimmt werden)
Bei allen anderen Stellen ist die Wahrscheinlichkeit der vorkommenden Aminosäure im Vergleich zu den anderen Aminosäuren am höchsten und deswegen bestimmbar und aussagekräftig.