

## Human T-cell leukemia virus type I

### Aufgabe 2

[<https://www.ncbi.nlm.nih.gov/nuccore/9626453>]

*Die erste 100 Basenpaare der kodierende Sequenz.*

```
a tgggccaaat cttttccgt agcgctagcc  
ctattccgcg gccgccccgg gggctggccg ctcatcactg gcttaacttc ctccaggcgg  
catatcgcc
```

### Aufgabe 3

*Die erste 1000 Basenpaare des kodierenden Sequenz.*

```
a  
tgggccaaat cttttccgt agcgctagcc ctattccgcg gccgccccgg gggctggccg ctcatcactg  
gcttaacttc ctccaggcgg catatcgcc agaaccgggt cctccagtt acgattcca ccagttaaaa  
aaatttetta aaatagcttt agaaacaccg gtctggatct gcccattaa ctactccctc ctageccagc  
tactcccaaa aggatacccc ggccgggtga atgaaatttt acacatactc atccaaacct aagcccagat  
cccgteccgc cccgcgcgc cgcgcgcgtc atctccacc cagaccccc cggattctga cccacaaatc  
ccccctccct atgttgagcc tacagcccc caagtcttc cagtcatgca cccacatggt gccctccca  
accaccgccc atggcaaatg aaagacctac aggccattaa gcaagaagtc tccaagcgg ccctggaag  
ccccagttt atgcagacca tccggcttgc ggtgcagcag ttgaccca ctgcaaaga cctccaagac  
ctctgcagt acctttgctc ctccctgtg gttccctcc atcaccagca gctagatagc cttatatg  
aggccgaaac tcgagggtatt acaggttata accccttagc cgtgtccctc cgtgtccaag ccaacaatcc  
acaacaacaa ggattaaggc gagaatacca gcaactctgg ctgcgcgcct tcgcgcgcct gccagggagt  
gccaaagacc ctctctgggc ctctatctc caaggcctgg aggagcctta ccagccttc gtagaacgcc  
tcaacatagc tcttgacaat gggctgccag aaggcaegcc caagacccc attttacgtt ccttagccta  
ctctaatgca aacaaagaat gccaaaaatt actacaggcc cgagggcaca ctaatagccc tctaggagat  
atgttgccggg cttgtcagg
```

*Ersten 30 Aminosäuren von 1.5'3' Frame*

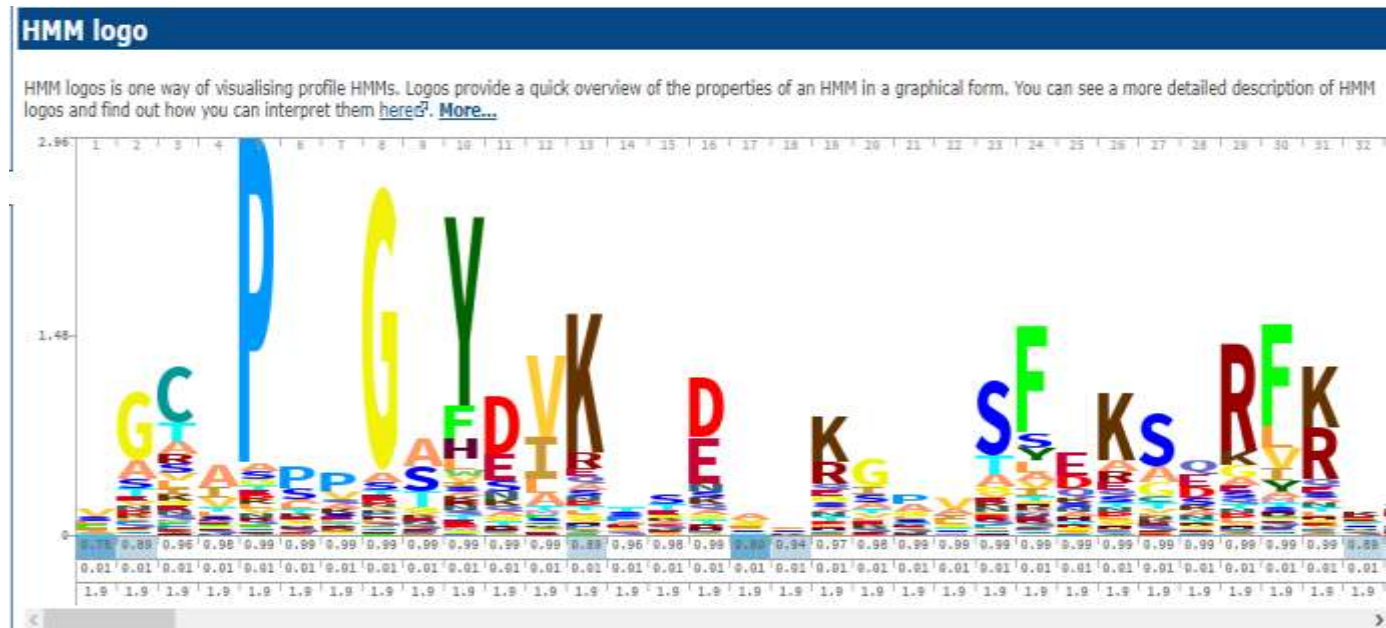
[<https://web.expasy.org/cgi-bin/translate/dna2aa.cgi>]

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

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



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

Im vergleich mit

MGQIFSRASPIRPPRGLAAHHWLNFLQA

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

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

## Aufgabe 5

### 5.2 Hepatitis C virus genotype 1,(NC 004102.1)

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

Die ersten 100 bp

```
atgagcacg aatcctaaac ctcaaagaaa aaccaaactg aacaccaacc gtcgccacac ggacgtcaag  
ttcccgggtg gcggtcagat cgttggtgga g
```

### 5.3 Die ersten 1000 Baasenpaare

```
atgagcacg aatcctaaac|ctcaaagaaa aaccaaactg aacaccaacc gtcgcccaca ggacgtcaag
ttccccgggtg|gcggtcagat cgttggtgga gtttacttgt tgccgcgcag gggccctaga ttgggtgtgc
gcgcgacgag gaagacttcc gageggtcgc aacctcgagg tagacgtcag cctatcccca|aggcacgtcg
gcccgagggc aggacctggg ctacgcccgg gtacccttgg ccctctatg|gcaatgaggg ttgcgggtgg
gcgggatggc tctgtctcc ccgtggctct cggcctagct|ggggccccac agacccccgg cgtaggctcg
gcaatttggg taaggtcacc gataccctta|ctgcggtt cgcgcacctc atggggtaca taccgtctgt
cggcgccct cttggaggcg|ctgccagggc cctggcgcat ggcgtccggg ttctggaaga cggcgtgaac
tatgcaacag|ggaaccttcc tgggtgctct ttctctatct tcttcttggc cctgctctct tgcctgactg
tgcccgttc agcctacca gtgcgaatt ccteggggt ttaccatgtc accaatgatt|gcctaactc
gagtatttg tacgaggcgg ccgatccat cctgcacact cgggggtgtg|tcccttgcgt tcgcagggtt
aacgcctcga ggtgttgggt ggcggtgacc cccacggtgg|ccaccaggga cggcaaacctc
cccacaacgc agcttcgacg tcatatcgat ctgcttgcg|ggagcgccac cctctgctcg gccctctacg
tgggggacct gtgcgggtct gtcttcttg|ttgtcaact gttaccttc tctccaggc gccactggac
gacgcaagac tgcaattgtt|ctatctatcc cggccatata acgggtcacc gcatggcatg ggatatgatg
atgaactggt|ccctacggc agcgttgggtg g
```

### Die ersten 30 Aminosäuren von 1.5'3' Frame

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

MSTNPKPQRKTKRNTNRRPQDVKFPGGGQI

### 5.4

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

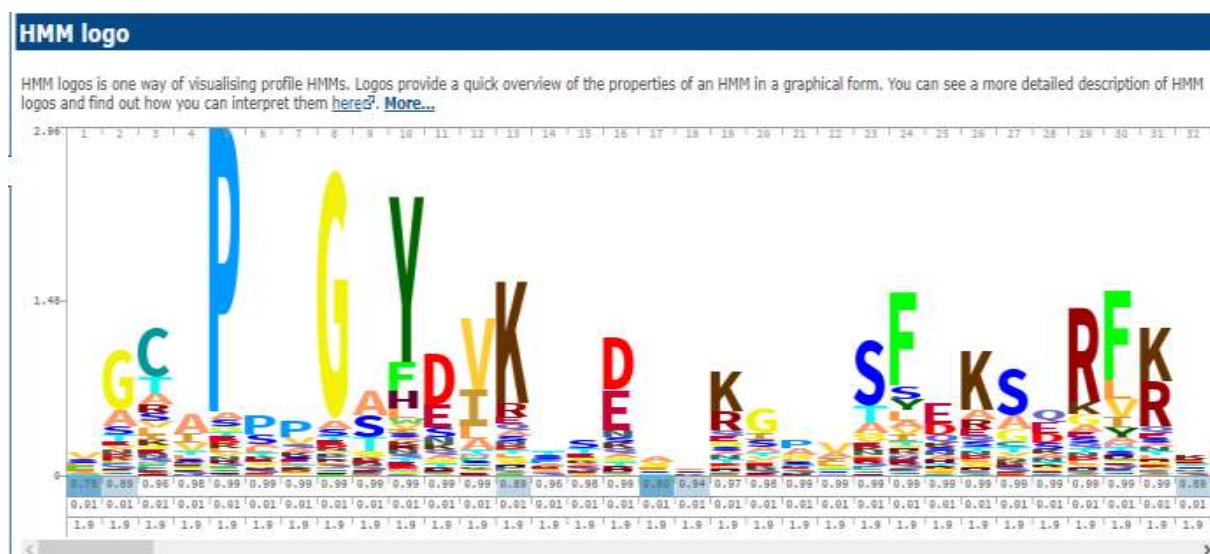


Abb.2. HMM Logo des gefundenen Profils aus Pfam Datenbank

Aminosäuresequenz aus Aufgabe 5.3 im Vergleich zum HMM-Logo:

-gleiche Aminosuren: grün

- Sonstige Aminosuren, die hohe Wahrscheinlichkeit beseitzen: schwarz

- nicht erkennbare Wahrscheinlichkeit für bestimmte Aminosäure: -

[-GCAPP~~P~~GAYDVK--D--KG--~~S~~F-KS-RF]

Es gibt also an Stelle 5, 7 und 24 eine gleiche Aminosäure.

Ansonsten gibt es 19 Aminosäuren, bei denen eine hohe Wahrscheinlichekeit zur eindeutigen Zuordnung gegeben ist.

---

Diese Aufgabe wurde zusammen mit Kushal Panta siehe

[<https://github.com/14livingdead/livingdead/tree/master/assignment%205>]

und Olga Ruda bearbeitet.