

## Aufgabe 2:

die ersten 100 Basenpaare der Genomsequenz von Human T-lymphotrophic virus 1

ATGGGCCAAATCTTTTCCCGTAGCGCTAGCCCTATTCCGCGGCCGCCCGGGGGCTGGCCGCTCATCACT  
GGCTTAACTTCCTCCAGGCGGCATATCGCC

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

## Aufgabe 3:

Die ersten 1000 Basenpaare

ATGGGCCAAATCTTTTCCCGTAGCGCTAGCCCTATTCCGCGGCCGCCCGGGGGCTGGCCGCTCATCACT  
GGCTTAACTTCCTCCAGGCGGCATATCGCCTAGAACCCGGTCCCTCCAGTTACGATTTCCACCAGTTAAA  
AAAATTTCTTAAATAGCTTTAGAAACACCGGTCTGGATCTGCCCCATTAATACTCCCTCCTAGCCAGC  
CTACTCCCAAAGGATACCCCGGCCGGGTGAATGAAATTTACACATACTCATCCAAACCAAGCCCAGA  
TCCCGTCCCGCCCCGCGCCGCCGCCGCGTCATCCTCCACCCACGACCCCCCGGATTCTGACCCACAAAT  
CCCCCTCCCTATGTTGAGCCTACAGCCCCCAAGTCCTTCCAGTCATGCACCCACATGGTGCCCCTCCC  
AACCACGCCCATGGCAAATGAAAGACCTACAGGCCATTAAGCAAGAAGTCTCCCAAGCGGCCCTGGAA  
GCCCCAGTTTATGCAGACCATCCGGCTTGCGGTGCAGCAGTTTGACCCCACTGCCAAAGACCTCCAAGA  
CCTCCTGCAGTACCTTTGCTCCTCCCTCGTGGCTTCCCTCCATCACCAGCAGCTAGATAGCCTTATATCA  
GAGGCCGAAACTCGAGGTATTACAGGTTATAACCCCTTAGCCGGTCCCCTCCGTGTCCAAGCCAACAATC  
CACAACAACAAGGATTAAGGCGAGAATACCAGCAACTCTGGCTCGCCGCCCTTCGCCGCCCTGCCAGGGAG  
TGCCAAAGACCCTTCTGGGCCTCTATCCTCCAAGGCCTGGAGGAGCCTTACCACGCCTTCGTAGAACGC  
CTCAACATAGCTCTTGACAATGGGCTGCCAGAAGGCACGCCCAAAGACCCATTTTACGTTCTTAGCCT  
ACTCTAATGCAAACAAAGAATGCCAAAATTACTACAGGCCCGAGGGCACACTAATAGCCCTCTAGGAGA  
TATGTTGCGGGCTTGTCAGG

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

Die ersten 30 Aminosäuren des 5'3' Frame 1:

MGQIFSRASPIRPPRGLAAHHLNQLQA

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

- In Aminosäuresequenzen ist es deutlich zu erkennen, wo Methionin entspricht dem Startpunkt für Translation ist. Außerdem sind die Introns und Exons auch in jeder Frame leichter zu erkennen als wenn man die Genomsequenz analysiert.
- Für eine DNA-Sequenz gibt es sechs Reading-Frames: drei von der 5-3' Richtung und drei von der entgegengesetzten Richtung 3-5', da eine Aminosäure durch ein Basentriplett codiert ist.

#### Aufgabe 4:

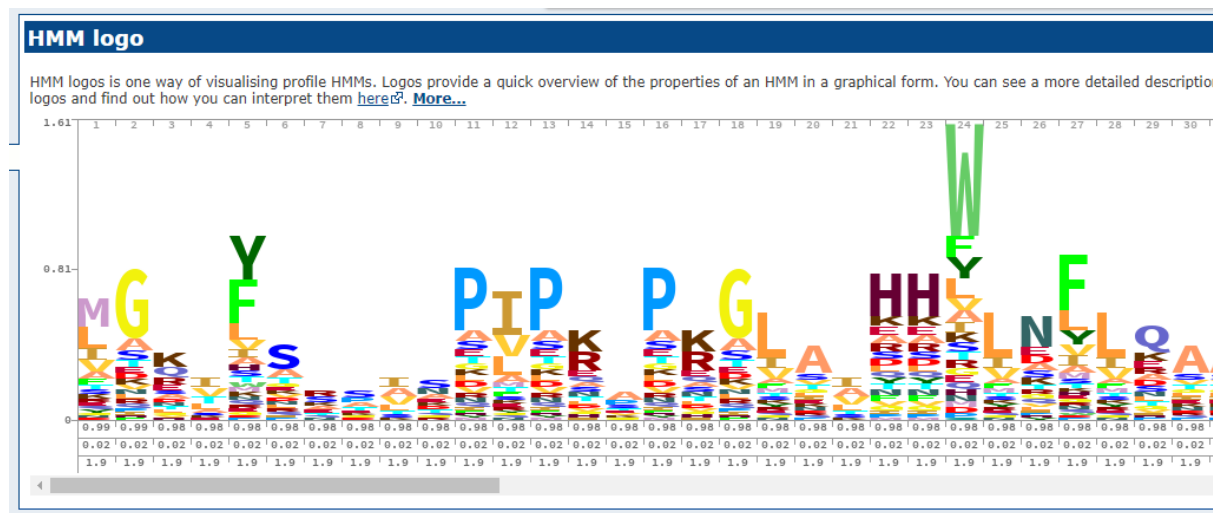


Abbildung 1: HMM-Logo der ersten 30 Aminosäuresequenz (siehe Aufgabe 3) aus Pfam Datenbank. Hier sind die Profile für die ersten 30 Aminosäuren. (<http://pfam.xfam.org/family/PF02228.15>)

Profil HMM: M/L-G-?-?-Y/F-S-?-?-?-P-I-P-K/R-?-P-K/R-G-L-A-I/A/V-H-H-W-L-N-F-L-Q-A

Such-Seq: M-G-Q-I-F-S-R-S-A-S-P-I-P-R-P-P-R-G-L-A-A-H-H-W-L-N-F-L-Q-A

Vergleich: Das Symbol '?' zeigt ein undeutliches Ergebnis aus dem Profil für die gehörige Basenstelle. Es sind insgesamt 7 Stellen mit unklarer Aussage. An anderen Stellen zeigen die Profile eine relativ gute Einschätzung/Wahrscheinlichkeit für das Vorkommen der Aminosäure wie in der Suchsequenz, die aus der Datenbank erhalten ist. Zwar treten bei einigen Stellen eine Möglichkeit für 2 oder mehrere AS auf (z.B. bei der 5. Stelle, 21. Stelle, usw.), von denen aber eine mit der Sequenz übereinstimmt.

#### Aufgabe 5: Genomsequenz von Human immunodeficiency virus 1 (NC\_001802)

<https://www.ncbi.nlm.nih.gov/nucleotide/9629357>

100 Basenpaare

ATGGGTGCGAGAGCGTCAGTATTAAGCGGGGAGAATTAGATCGATGGGAAAAAATTCGGTTAAGGCCAG  
GGGGAAGAAAAATATAAATTAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAA

1000 Basenpaare

ATGGGTGCGAGAGCGTCAGTATTAAGCGGGGAGAATTAGATCGATGGGAAAAAATTCGGTTAAGGCCAG  
GGGGAAGAAAAATATAAATTAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAA  
TCCTGGCCTGTTAGAAACATCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACA  
GGATCAGAAGAAGCTTAGATCATTATATAATACAGTAGCAACCCCTCTATTGTGTGCATCAAAGGATAGAGA  
TAAAAGACACCAAGGAAGCTTTAGACAAGATAGAGGAAGAGCAAAAAGTAAGAAAAAGCACAGCA  
AGCAGCAGCTGACACAGGACACAGCAATCAGGTCAGCCAAAATTACCTATAGTGCAGAACATCCAGGGG  
CAAATGGTACATCAGGCCATATCACCTAGAAGCTTTAAATGCATGGGTAAAAGTAGTAGAAGAGAAGGCTT  
TCAGCCAGAAAGTATACCATGTTTTAGCATTATCAGAAGGAGCCACCCCAAGATTTAAACACCAT  
GCTAAACACAGTGGGGGACATCAAGCAGCCATGCAATGTTAAAGAGACCATCAATGAGGAAGCTGCA  
GAATGGGATAGAGTCAATCCAGTGCATGCAGGCGCTATTGCACCAGGCCAGATGAGAGAACCAAGGGGAA  
GTGACATAGCAGGAAGTACTAGTACCCTTCAGGAACAAATAGGATGGATGACAAATAATCCACCTATCCC  
AGTAGGAGAAATTTATAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAAATGTATAGCCCTACC  
AGCATTCTGGACATAAGACAAGGACCAAGGAACCCCTTTAGAGACTATGTAGACCGGTTCTATAAACTC

TAAGAGCCGAGCAAGCTTCACAGGAGGTAAAAAATTGGATGACAGAAACCTTGTGGTCCAAAATGCGAA  
CCCAGATTGTAAGACTATTT

30 Aminosäure aus der 5-3' Frame 1

MGARASVLSGGELDRWEKIRLRPGGKKKYK

Profil

--G-A-R/A-A/I-S-V/A-L-?-?-?-K/E-L-D-K/R-F/W/Y/L-E-K/R-I/V/L-R-L-R-P-G-G-K-K-K-Y-?

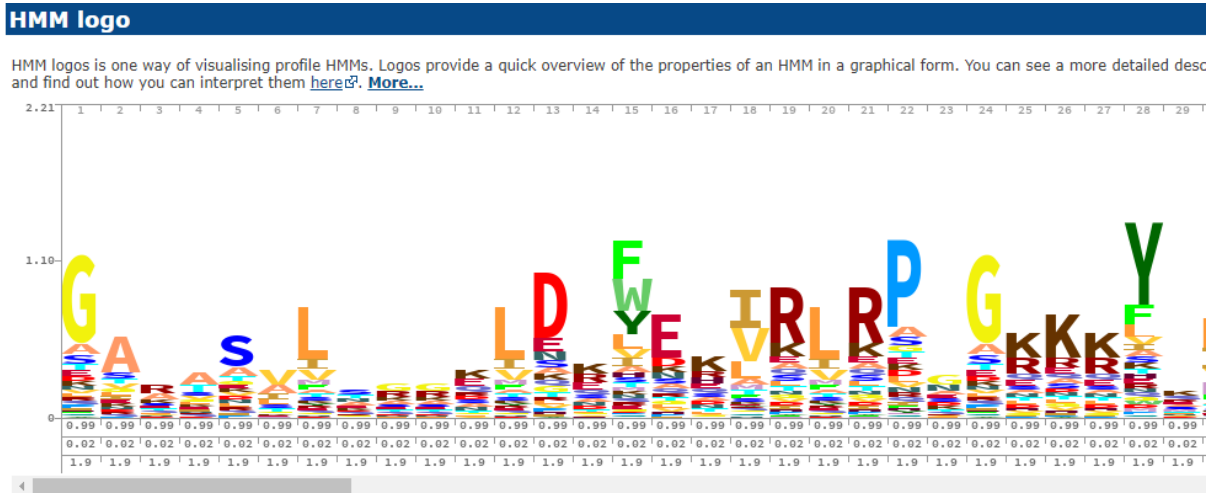


Abbildung 2: HMM-Logo des gefundenen Profil für die Aminosäuresequenz aus Pfam Datenbank.  
<http://pfam.xfam.org/family/PF00540>

Das Profil gibt eine relativ gute Aussage über die vorkommenden AS trotz der einigen Ungenauigkeiten oder mehreren AS-Möglichkeiten an einer bestimmten Stelle. Die erste AS Methionin, die dem Startpunkt entspricht, ist in dem HMM-Logo nicht gegeben.