

Assingment 5

2.1. Sequenz des Human T-lymphotropic virus 1

Die ersten 120 Basen des Genoms:

```
ggctcgcatc tctccttcac gcgcccgcg ccttacctga ggccgccatc cacgccggtt
gagtcgcgtt ctgccgcctc ccgctgtgg tgcctcctga actacgtccg ccgtctaggt
```

2.2. Sequenz (die ersten 120 Basen) des ersten Gens des Eupatorium vein clearing virus (NC_010738):

```
ATGAGCTCAAGCCATGAGAATGAAGCTGACGCTTCTCAACACATATCTTCTAACGAAGAATACATCTTCT
CTAACGAGAATGAAAATGGATTTATTGGTACAATGACTATTCATCAGGACCAATTAAGAAAAGATAAATCA
ACTACAGCTCTCGCTGAAAAGTGATTCAATCTTTAGTCGAAGCATCCTCGACAGATTATCTCGCAAAAAAC
```

(Die ersten 120 Basen der Sequenz an sich:

```
tggtatcaga gccatattta gatcaatcct cggttgactg aacaatgagc tcaagccatg
agaatgaagc tgacgttct caacacatat cttctaacga agaatacatc ttcttaacg)
```

3.1. Die ersten 30 Aminosäuren (AS) des 5'3' Frame 1 des **ersten genes**:

MGQIFSRASPIRPPRGLAAHHLNQLQA

- Ob die Suche nach der Aminosäuresequenz sinnvoller als die nach der Genomsequenz ist, hängt natürlich auch davon ab was genau gesucht wird. Generell kann aber gesagt werden, dass aufgrund des degenerierten Codes (Triplets aus 4 Basen codieren für 20 AS, dadurch codieren mehrere Triplets für eine AS) und des vorhandenen Leserasters die Aminosäuresequenz auf den direkten Blick wesentlich aufschlussreicher ist (*Basensequenzen müssen sich nicht zwangsweise gleichen um für dieselbe AS-Sequenz zu codieren*). Außerdem lassen sich durch die Aminosäuren erste Rückschlüsse beziehungsweise Vermutungen bezüglich Lage des Peptids im Protein, eventuelle Motive und mögliche Modifikationen treffen.
- Zum einen müssen alle möglichen Leseraster kontrolliert werden (daher drei Frames) und zum anderen ist nicht klar ob die forward oder die reverse Sequenz gegeben war, daher verdoppelt sich die Anzahl der Frames erneut (5'3'/3'5').

3.2. Die ersten 30 Aminosäuren (AS) des 5'3' Frame 1 des **ersten genes**:

MSSSHENEADASQHISSNEEYIFSNENENG

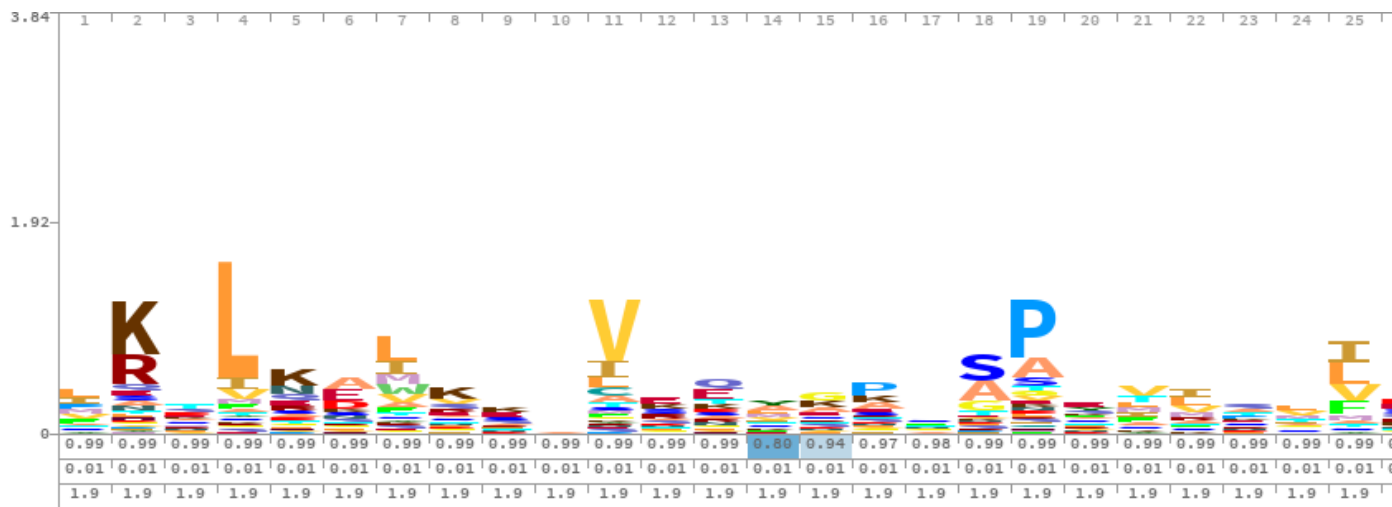
4.1.

Accession Logo (PF00607.19) von (Id.: Gag_p24):



<http://pfam.xfam.org/family/PF00607.19#tabview=tab4>

Den ersten Teil des Logos noch einmal vergrößert:



Das Programm gibt mir an, dass das gefundene Motiv Ähnlichkeiten zu meiner eingegebenen Sequenz (5'3' Frame 1) zwischen der 147 und der 344 AS hat.

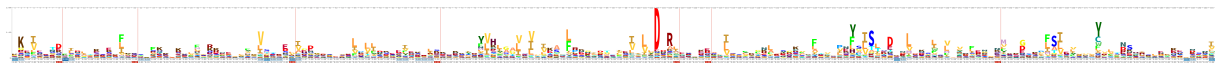
Vergleiche ich die Anfänge dieser Sequenzen miteinander, so fällt auf, dass sich insbesondere die konservierten AS sich decken (K 148, L 150, A 152, K 154, V 157...).

```

MGQIFSRAS PIPRPPRGLA AHHWLNFLQA AYRLEPGSS YDFHQLKKFL KIALETPVWI
CPINYSLLAS LLPKGYGRV NEILHILIQT QAIQPSRPAP PPPSSSTHDP PDSDPQIPPP
YVEPTAPQVL PVMHPHGAPP NHRPWQMKDL QAIKQEVSA APGSPQFMQT IRLAVQQFDP
TAKDLQDLLQ YLCSSLVASL HHQQLDSLIS EAETRGITGY NPLAGPLRVQ ANNPQQQGLR
REYQQLWLAA FAALPGSAKD PSWASILQGL EEPYHAFVER LNIALDNGLP EGTPKDPILR
SLAYSANKE CQKLLQARGH TNSPLGDMIR ACQAWTPKDK TKVLVVQPKK PPPNQPCFRC
GKAGHWSRDC TQPRPPGPC PLCQDPHTWK RDCPRLKPTI PEPEPEEDAL LLDLPADIPH
  
```

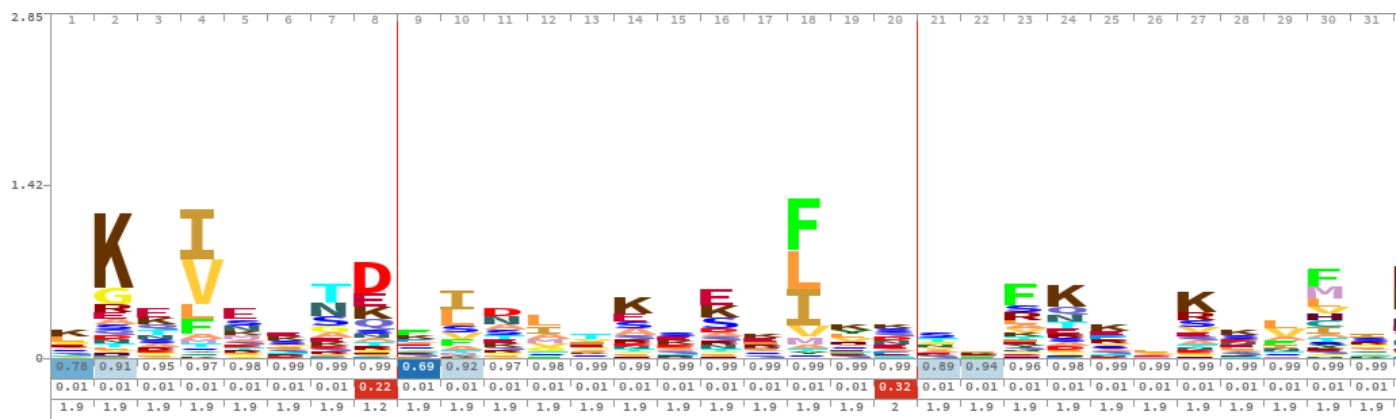
4.2.

Accession Logo (PF01107.17) von (Id.:MP):



http://pfam.xfam.org/family/PF01107/logo_image

Den ersten Teil des Logos noch einmal vergrößert:



Das Programm gibt mir an, dass das gefundene Motiv Ähnlichkeiten zu meiner eingegebenen Sequenz (5'3' Frame 1) zwischen der 42 und der 235 AS hat.

Vergleiche ich die Anfänge dieser Sequenzen miteinander, so sind zunächst kaum Ähnlichkeiten in den „wahrscheinlicheren“/konservierteren AS zu erkennen. Vermutlich decken sich die Sequenzen in einem anderen Bereich stark. In der nachfolgenden Sequenz habe ich die 42. AS der bei HMMSCAN eingegebenen Sequenz fett markiert.

```

ATGAGCTCAAGCCATGAGAATGAAGCTGACGCTTCTCAACACATATCTTCTAACGAAGAATACATCTTCT
CTAACGAGAATGAAAATGGATTTATTGGTACAATGACTATTCATCAGGACCAATTAAGAAAGATAAATCA
ACTACAGCTCTCGCTGAAAAGTGATTCAATCTTTAGTCGAAGCATCCTCGACAGATTATCTCGCAAAAAC
GAGATCATCTATTGTGTTAACACACAAGAAAAATCCATTGACATTAAAGATGTTGAAGGAACAGTCTATT
TACCGTTAGTAACTAAACAGGAAATAGATAACAAGCTAAGTAAGATAAAATCTTCTATAGCAGTAAAAAT

```

5. siehe 2.2., 3.2. und 4.2. für die erste Gensequenz von